Politecnico di Torino

PhD programme in Management, Production and Design

Consumer data effects on competition and market outcomes

Flavio Pino

2023

Politecnico di Torino

Department of Management and Production Engineering

PhD programme in Management, Production and Design

XXXV Cycle

S.S.D.: SECS-P/06

Consumer data effects on competition and market outcomes

Supervisor:
Prof. Carlo Cambini

Candidate:
Flavio Pino

Doctoral Examination Committee:

Prof. Paule Belleflamme, Referee, Université Catholique de Louvain CORE

Prof. Patrick Walbroeck, Referee, Télécom Paris

Prof. Laura abrardi, Politecnico di Torino

Prof. Luigi Buzzacchi, Politecnico di Torino

Prof. Clara Graziano, università di Udine

Coordinator: Prof. Arianna Alfieri

November 2019 - January 2023

Acknowledgements

I thank my supervisor Carlo Cambini for the constant support and guidance he gave me during these years. His feedback and mentorship have helped me shape my approach to research.

I am thankful to the referees, Paul Belleflamme and Patrick Walbroeck, for the valuable feedback they gave me on the first draft of this thesis. Their help allowed me to substantially improve the contents and exposition of this work.

I thank the CORE Department of Université Catholique de Louvain for welcoming me during my visiting period.

The research presented in this thesis has seen major improvements thanks to professors and colleagues from the DIGEP department, as well as the participants of conferences and seminars I have presented at during these years. Their feedback has helped me to refine my work. I am especially thankful to all the co-authors of this thesis: Laura Abrardi, Carlo Cambini and Raffaele Congiu for the second chapter; Laura Abrardi and Carlo Cambini for the third chapter; Federico Navarra and Luca Sandrini for the fourth.

Finally, I thank those who helped me go through these chaotic months. Cri, Edo and Mauro for always being there, either physically or otherwise; Sasha, for being the best housemate I could have asked for; Giammi, for being the safety net I can always rely on; the group of Corso Trento, for bearing with my constant complaining. I would have not made it out without you.

Contents

List of Tables

List of Figures

Declaration of Co-Authorship

The second chapter is a joint work with Laura Abrardi, Carlo Cambini and Raffaele Congiu (Politecnico di Torino).

The third chapter is a joint work with Laura Abrardi and Carlo Cambini (Politecnico di Torino).

The fourth chapter is a joint work with Federico Navarra (Università di Padova) and Luca Sandrini (Budapest University of Technology and Economics).

Introduction

The digital age has revolutionized the way that businesses operate. With the rise of digital markets, businesses can now collect and analyze an unprecedented amount of consumer data. This has enabled them to better understand their customers' wants and needs and tailor their products and services accordingly.

Consumer data can be obtained through various means. Some, usually with low accuracy, can be obtained through public repositories, such as the average wages in a city's zone. Other data are instead generated by the interaction between consumers and firms, and the latter can then use that information when it interacts with that consumer again (or consumers with similar tastes). However, most consumer data is brokered by big-tech players. Two similar yet distinct classes in this market are those of digital platforms and Data Brokers (DBs). Digital platforms, such as Google, Meta, and Amazon, collect information through the services they offer to consumers and then sell access to consumer segments through digital advertising. Instead, DBs such as Acxiom, BlueKai, Experian, and Teradata are less known to the general public, as they usually combine different sources of existing data instead of generating them through customer interaction.

Whether the increasing use of consumer data is benefiting consumers and markets as a whole is an open debate. The complexity lies in the many uses consumer data can have, together with privacy concerns linked to data misuse or abuse and the level of market concentration in the data brokerage sector. The aim of this thesis is to first assess the vast existing literature regarding the effects of consumer data, trying to extrapolate general insights that can be helpful in guiding policy interventions. Then, I shift my focus toward big-tech players and their ability to influence downstream markets.

In the first chapter, I provide a survey of the existing literature regarding consumer data and their effects. As this literature includes various strands, such as artificial intelligence and machine learning, I limit the scope of the analysis to those papers that model data as an explicit input to a decision problem. This choice thus excludes works regarding data-enabled technologies that do not explicitly model how the quantity or quality of data affect decisions. With the aim of extracting broader insights, I organize the existing literature depending on the data acquisition stage. Indeed, I find that the modelization choices of this stage often guide many of the market outcomes presented.

First, I focus on papers where data are exogenously available to firms or where firms can obtain them without strategic interactions (such as buying them at an exogenous cost). This class of models can aptly represent scenarios where the data used are publicly available. This modeling choice does not usually allow firms to fully internalize data externalities: as such, they present an overuse of data with respect to the social optimum, which leads to an increase in competition between firms. While this effect should benefit consumers, data overcollection can raise privacy concerns that should be taken into account when examining the effects of data.

Second, I focus on models where firms can obtain data by interacting with consumers one or multiple times. Firms must then take into account the effects that data collection has on consumers' future behavior in order to maximize the added value they gain from data. In this class of models, I find that the pro or anti-competitive effect of data is strongly linked to firm symmetry. Indeed, the data use over repeated periods can exacerbate a firm's starting advantage, increasing concentration in the market or even leading to market tipping. Moreover, firms can also trade or sell data among them to limit consumer interaction and thus reduce the compensation they should pay to consumers. This feature is especially relevant when consumer data are correlated, enabling firms to gain information on consumers they do not directly interact with.

Finally, I analyze models where data are obtained from intermediaries that can either directly interact with consumers or not to obtain such data. As intermediaries serve multiple firms, they have the ability to coordinate their sales in order to temper competition in the downstream market, as this allows them to extract higher rents from firms. In particular, intermediaries are able to internalize the effect that selling data to a firm has on its rivals and choose their data-selling strategy accordingly. Moreover, competing data intermediaries can strategically coordinate their actions to temper competition between them.

For each of these classes of models, I briefly describe the evolution of the theoretical literature and the features that drive the main findings. Moreover, whenever possible, I integrate the analysis with empirical works to better frame the insights of the theoretical literature.

In the second chapter, I build on the existing literature by focusing on the effects that a DB has on a downstream market where entry is endogenous. In particular, I focus on the case where consumer data sold by the DB allows firms to price discriminate consumers. Most of the literature on DBs has focused on downstream markets with a fixed number of firms, highlighting a pro-competitive effect of data. Indeed, price discrimination induces firms to lower their prices, benefiting consumers as a whole.

I study a monopolistic DB that has data regarding all consumers in the downstream market, which is modeled as a circular city where firms can enter by paying a fixed cost. This can be seen, for example, as the cost of opening a digital store through which a firm can serve consumers. Firms can buy data from the DB and then use them to price discriminate consumers, offering them tailored prices.

The analysis shows that the DB's equilibrium strategy depends on the selling mechanism he adopts. When the DB sells data through Take-It-Or-Leave-It offers, he sells data to all entering firms. In particular, he chooses to sell non-overlapping partitions if the downstream market horizontal differentiation is high and opts to sell the whole dataset to all entering firms otherwise. Instead, if the DB can change the offer he makes to a firm conditional on another firm's response, such as through auction mechanisms, the DB opts to only serve a subset of the entering firms to further temper downstream competition. Regardless of the selling mechanism, the data sale results in a reduction of the number of entering firms, which I refer to as *entry barrier effect*. The entry reduction results in an increase in market concentration, which in turn harms consumers with respect to the benchmark model with no data. The results thus highlight how the presence of a monopolistic DB, who can strategically use the data sale to influence downstream competition, can lead to harmful effects on consumers that should be taken into account by policymakers.

In the third chapter, I expand the previous model by adding competition in the DB market and by analyzing the effect of information accuracy on market outcomes. To do so, I model a vertically differentiated duopoly in the DB market, where DB_1 and DB_2 sell data partitions that grant firms a probability of identifying consumers to price discriminate them. In particular, DB_1's partitions have accuracy $\alpha \in [0, 1]$, and DB_2's partitions have accuracy $\beta\alpha, \beta \in [0, 1]$. α is thus a measure of information accuracy, while β is a proxy for the degree of DB competition.

In equilibrium, only DB_1 sells partitions to downstream firms, as his data are more accurate. However, DB_2 exerts competitive pressure on him, limiting the rents DB_1 can extract from downstream firms. While the data sale always decreases downstream entry, the effects on consumer surplus with respect to the benchmark are more nuanced. Indeed, I find that the degree of information accuracy determines the magnitude of the effect that the data sale has on consumer surplus. Instead, the degree of competition between DBs determines whether this effect will be positive or negative. In particular, consumer surplus is maximized (and higher than the benchmark model with no data) when both information and DB competition are perfect, and conversely, it is minimized when information is perfect, and the DB market is monopolistic.

However, this result critically depends on the amount of overlaps between the DBs datasets and the presence of synergies between them. If both DBs have enough exclusive data on specific consumer groups, then entering firms benefit from purchasing both of them. As the DBs can charge high prices for their datasets, firm entry in the downstream market is reduced, potentially leading to consumer harm. In particular, if data are *super-additive* (that is, the accuracy of the combined datasets is higher than the sum of the individual datasets' accuracies), consumer harm always emerges in equilibrium as the DBs can coordinate their prices to extract all available surplus from firms. The analysis thus shows that policymakers should aim to ensure a level playing field in the DB market, as it is the main determinant of whether the effect of data sales will benefit or harm consumers. Such an intervention should not only be based on the datasets' sizes but also on data that are exclusively accessible by a single DB.

Finally, in the fourth chapter, I focus on a different type of data intermediary, which is a digital platform. In particular, the chapter aims to analyze the effects that mandating data sharing can have on hybrid marketplaces such as Amazon. The term hybrid marketplace indicates a platform that hosts sellers but can also vertically integrate to compete with them. As data sharing is one of the main pillars of the upcoming Digital Markets Act, I focus on the intended and unintended effects this policy measure can have.

To do so, I model a continuum of downstream markets, each one containing a firm. I assume that all these markets belong to the same product group (such as "baby products" or "clothing and accessories"). Firms pay a per-transaction fee to the platform, consistent with Amazon's referral fees, which is unique for the given product group. Moreover, I only allow the platform to enter a market where a firm is already present, as empirical evidence has shown that platforms like Amazon usually enter markets after having observed their profitability.[1].

I study the effects of mandated data sharing when the downstream markets sell heterogeneous goods. In particular, I focus on the case where data allow firms to price discriminate consumers. The results show that mandating data sharing can indeed have unintended consequences. While total welfare increases, the platform can strategically use the per-transaction fee as a tool to temper competition with sellers. In particular, the platform can set a fee so high that downstream sellers opt to set monopolistic prices, even if the market is fully covered in equilibrium. This strategy allows the platform to effectively avoid competition with sellers, leading to higher profits. As sellers become more efficient in extracting surplus from consumers under mandated data sharing, I find that consumers are ultimately harmed by such a policy.

The results also highlight how the platform always achieves higher profits under data sharing, as data improves market efficiency, and the platform can appropriate most of the surplus through the fee. Then, a question naturally arises: if platforms unambiguously benefit from data sharing, why should a policymaker mandate it? I interpret this seemingly paradoxical result as the consequences of hidden costs I fail to model, such as investments in interoperability between the platform and sellers. Those costs could be non-negligible and entail competitive risks for the platform, as data sharing could stimulate entry by new platforms, exerting potential negative pressure on the incumbent. Indeed, the recitals of the DMA place market contestability among the important goals of the act. Further research is thus needed to better analyze these additional characteristics.

[1]see https://www.reuters.com/investigates/special-report/amazon-india-rigging/

CHAPTER 1

The microeconomics of consumer data*

Flavio Pino†

In recent years, academia, institutions, and policymakers have been focusing their attention on the impact of data in digital markets. The economic literature that explicitly models data and their collection as strategic variables is growing, but most studies focus on distinct settings with specific data uses. This chapter aims to organise this literature to extract general insights that hold across different models and assumptions. To do so, I identify three classes of models according to the way they model data collection. I find that each class is characterised by a specific impact of data on the market outcomes, regardless of the specific data use. First, when firms obtain data without strategic interactions, their use has a pro-competitive effect on the market. However, firms fail to fully internalise the data externalities, leading to data overuse and, in turn, privacy concerns. Second, when firms collect data from their interaction with consumers, data can facilitate market tipping, especially if firms are asymmetric in their starting positions. Third, when firms acquire data from data intermediaries, data are strategically sold to temper competition in the downstream market, allowing intermediaries to extract most of the surplus at the expense of firms and consumers. These general insights can facilitate future research and help policymakers to have a more general understanding of the competitive effects of data, depending on the situation at hand.

*A version of this chapter has been published in the *Journal of Industrial and Business Economics*. See Pino, F. The microeconomics of data – a survey. J. Ind. Bus. Econ. 49, 635–665 (2022). https://doi.org/10.1007/s40812-022-00220-6

†Politecnico di Torino, Department of Management, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy.

1. Introduction

The role of data in economics has seen a significant increase in importance in recent years, primarily due to the ever-increasing relevance of digital markets (European Commission, 2019). The use of data is widespread across every sector, thanks to their versatility; typical uses include improving products or services' quality and efficiency, personalization, matching, and discriminating between different consumers groups or individuals (Goldfarb and Tucker, 2019). Moreover, the rapid growth of online platforms has raised new challenges to competition and privacy authorities, who need to assess the potential outcomes of data-driven business models correctly.

Over the years, many authors have developed models to study the effects of data on various aspects, such as market structure, competition, welfare, and privacy. However, most of these works have only focused on a specific data effect in peculiar settings. This, in turn, has created a conundrum: while the impact of data is widely analyzed, the specificity of most studies makes it difficult to abstract more general insights. Moreover, since little is known about the data collection and sale processes (Montes et al., 2019), the theoretical models currently outnumber the empirical papers where the analyzed strategies can be observed in action. Thus, there can be a perceived detachment between the analyzed models and real-world situations.

Data have been described as 'the oil of the digital era'[1], and there is widespread consensus that their use significantly influences the economy. However, data have a considerable number of applications and reviewing how all of them impact market outcomes may result in ambiguous insights that would be of little help for policymakers or for suggesting future research developments. To set a boundary on the scope of this work, I focus my attention on models where data are explicitly modelled as an input to a decision problem. This chapter is thus positioned in the strand of literature commonly referred to as digital economics. This choice excludes most of the literature regarding artificial intelligence and machine learning, as that strand of literature often focuses on data-enabled technologies rather than on how different quantities (or qualities) of data affect such technologies. The reader can refer to Agrawal et al. (2019) for a broad analysis of artificial intelligence and economics and Abrardi, Cambini and Rondi (2022) for a comprehensive survey on artificial intelligence and machine learning.

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

Recent contributions have aimed to review the growing literature on digital economics by finding common characteristics that could help the authors abstract from the individual models. Goldfarb and Tucker (2019) organize the literature by identifying five types of cost reductions that stem from digital technologies and how they impact market outcomes. Bergemann and Bonatti (2019) instead focus on the characteristics of information products and their sale, as well as the interaction between firms and data intermediaries. The contribution of the present review is twofold. First, the literature on digital economics has widely expanded in the last years, bringing new approaches and insights that could help direct future research and policy action. This chapter aims to organize these recent additions and link them to previous research developments. Whenever available, I also present related evidence from empirical papers to better frame the insights described in the theoretical models. Second, I find that the assumptions regarding data collection are a strong driver for the models' market outcomes, regardless of the specific data use. Thus, I organize the literature depending on how data collection is modeled. This approach allows me to extract general insights that hold across different models and assumptions.

The chapter is organized into three sections, each dedicated to a class of models. First, I analyze the studies where firms collect data without strategically interacting with other actors. Examples include models where firms exogenously have data from the start or when firms can acquire data by paying a marginal cost. Second, I focus on papers where firms acquire data through single or repeated interactions with consumers. In this class, firms consider the trade-off between data collection and its effect on consumer behavior and the inter-temporal effects of data acquisition (Y. Chen, 1997; Fudenberg and Tirole, 1998). Third, I analyze models where firms can acquire data from strategic third parties, referred to as data intermediaries. These actors usually function as data collectors and aggregators, compounding different sources to better profile consumers (Commission, n.d.). As data intermediaries serve multiple firms, their selling strategies consider how selling data to a given firm impacts its rivals. Thus, data intermediaries have a higher degree of internalization of the overall data effects when compared to the first class of models.

When firms collect data without strategic interactions, data have a pro-competitive effect. The increase in competition is due to firms' over-collection and overuse of data, as they have a limited internalization of data externalities. On the other hand, data over-collection raises privacy concerns that policymakers should consider when accounting for the effects of data.

When firms obtain data from consumers, the effects of data strongly depend on firms' symmetry. Data acquisition and use in repeated periods can exacerbate a firm's starting advantage, potentially increasing concentration and even leading to market tipping. Moreover, firms can strategically trade or share data to limit their interaction with consumers and reduce the compensation they pay to consumers for data. This strategy is especially relevant when consumer data are correlated, as even small datasets can help firms infer information on consumers who did not disclose their data. Policymakers should thus pose particular attention to data sharing and data-driven mergers.

Finally, data acquisition through intermediaries results in various outcomes that policymakers should consider. Data intermediaries strategically sell their datasets to temper competition in the downstream markets to extract more profits at the expense of both firms and consumers. Moreover, the high concentration of the data intermediaries' industry (Commission, n.d.) grants them substantial market power, and competing data intermediaries can strategically coordinate their actions to temper competition between them. These insights suggest that further research is needed to assess better if and which policy interventions should be implemented to limit consumer harm in these scenarios.

The chapter is organized as follows. In Section 2, I focus on models where firms acquire data without strategic interactions with other actors. Section 3 describes works where firms acquire data from their interaction with consumers. Section 4 analyzes papers where firms acquire data from data intermediaries. In each section, I briefly describe the development of the theoretical models, highlighting the main differences and findings. I then abstract from the individual models to gather more general insights and policy implications for the entire class and better assess the effects of data arising from each data acquisition method. Finally, Section 5 concludes.

2. Data acquisition with no strategic interactions

I first analyze the strand of literature where firms acquire data without strategically interacting with other actors. Examples include models where data are exogenously available to firms or firms incur a marginal cost when acquiring data. For ease of exposition and to better highlight connections between works, I separately analyze models where data have different effects. In particular, I distinguish three cases: when data allows price discrimination, targeting ads, and more general effects like cost reduction or revenue increase.

2.1. Price discrimination

Price discrimination is a practice where firms can identify consumers, leading to targeted offers that extract surplus better. Data can thus be seen as a tool that allows customer identification, enabling price discrimination. Usually, price discrimination is a profitable strategy for firms: however, it can also lead to increased competition, dissipating profits.

Consider a duopoly in a spatial competition setting, where two symmetric firms exogenously have data on all consumers and can operate first-degree price discrimination (Thisse and Vives, 1988; Shaffer and Zhang, 1995; Bester and Petrakis, 1996; Taylor, 2003). On the one hand, firms can calibrate their targeted offers depending on consumers' willingness to pay: this allows them to extract higher profits from consumers and is commonly referred to in the literature as the surplus extraction effect (Liu and Serfes, 2004). On the other hand, since both firms can send targeted offers to all consumers, they anticipate their rival strategy and engage in price wars. Firms thus lower their tailored prices until they match the difference in consumers' willingness to pay between the two firms. This effect dissipates profits and is commonly referred to in the literature as intensified competition effect (Liu and Serfes, 2004). When firms are symmetric, the surplus extraction effect is lower than the intensified competition effect. Thus, firms realize lower profits under price discrimination than under mill pricing, while consumer surplus increases due to increased competition. However, firms' best response is always to commit to first-degree price discrimination: if the rival does not commit to it, a firm's profits increase with price discrimination, while if the rival commits to price discrimination, the firm limits its losses by also committing to it. In other words, firms face a prisoner's dilemma: while they would be better off by not engaging in price discrimination, it is their best response to the rival's strategy. These results suggested that privacy could harm consumers since it would temper competition between firms.

However, further analysis showed that the insight heavily depended on the assumption that firms could identify all consumers from the start. Introducing a cost of information acquisition (Y. Chen and Iyer, 2002; Shy and Stenbacka, 2013) leads firms to identify fewer consumers in the market, tempering the intensified competition effect and increasing profits while making consumers worse off. Similar results are obtained under different variations of the basic model with costly data acquisition. These variations include scenarios where information sharing is allowed (Shy and Stenbacka, 2016), where consumers are loyal to one firm (Anderson et al., 2019) or where consumers can actively hide from firms (Belleflamme and Vergote, 2016; Z. Chen et al., 2020). Moreover, Taylor and Wagman (2014) examine the effects of privacy by comparing the market outcomes of various fundamental models when firms can or cannot operate first-degree price discrimination. They find that first-degree price discrimination favors consumers in a multi-unit symmetric demand model. At the same time, it harms them in a Hotelling setting (Hotelling, 1929), a Salop setting (Vickrey, 1964; Salop, 1979), and a vertical differentiation setting (Tirole, 1988).

Market asymmetries can also influence the surplus extraction and intensified competition effects, leading to diverse outcomes. When two firms exogenously have data on all consumers and are vertically differentiated in a spatial competition setting, the high-quality firm can expand its market share, increasing profits (Shaffer and Zhang, 2002). Instead, if only one firm has data, semi-collusive behavior can arise through a first-mover advantage (Gu et al., 2019): the informed firm sets a high price, enabling his rival to undercut him and thus avoid a price war. Switching to a homogenous product market, asymmetries become crucial for profitability. If both firms have the same level of data accuracy, or only one firm is informed, firms end up in the Bertrand paradox (i.e., firms set prices equal to their marginal production costs and achieve zero profits). However, when both firms have imperfect tracking with different accuracies, they can achieve positive profits at equilibrium, making consumers worse off (Belleflamme et al., 2020).

Other authors have instead focused their attention on third-degree price discrimination. In these models, data enable firms to observe a consumer's type, often identified with the consumer's loyalty to one of the firms. Compared to first-degree price discrimination, third-degree price discrimination leads to a lower increase in competition between firms, as they do not engage in price wars over individual consumers. Moreover, third-degree price discrimination could allow firms only to identify consumer types, tempering even further competition. As a guiding example, consider a case where firms only identify their loyal consumers. In this scenario, firms can extract higher profits from them but do not try to poach their rivals' consumers as they do not identify them. Only identifying loyal consumers results in lower competition and higher firms' profits (Shaffer and Zhang, 2000; Iyer et al., 2005). Firms can also escape the prisoner's dilemma identified by Thisse and Vives (1988) if information quality is low enough. In this situation, firms' equilibrium strategy involves committing not to price discriminate, even if the information is free. However, this strategy is dominated once information quality rises (Liu and Serfes, 2004). Moreover, third-degree price discrimination can itself temper competition enough so that firms find it profitable to discriminate all consumer types, making them worse off (Armstrong and Zhou, 2010). An interesting analysis has been recently carried out by Bergemann and Bonatti (2015), who focus on third-degree price discrimination in a single-product monopoly setting. In particular, they analyze the welfare effects of market segmentation and demonstrate that market segmentation can achieve any combination of consumer surplus and producer surplus as long as (i) consumer surplus is nonnegative, (ii) producer surplus is greater or equal than his surplus under no segmentation and (iii) total surplus does not exceed the total value that consumers receive from the good.

Finally, introducing inaccuracies in the price discrimination process has ambiguous effects on industry profits and consumer surplus, depending on market specifics and the starting accuracy level (Y. Chen et al., 2001; R. Esteves, 2014; Mauring, 2021). In particular, information inaccuracy is crucial when competing firms have asymmetric market shares (Colombo et al., 2021). If the inaccuracy is high, firms are incentivized to deviate from the equilibrium, as not offering tailored prices avoids price wars. If instead inaccuracy is low, different equilibria arise depending on firms' starting differences in market shares, and the firm with the highest starting market share ends up with lower profits than its rival. Moreover, the effect of information inaccuracy on welfare also depends on market shares' asymmetry: an increase in accuracy benefits consumers when asymmetry is high and harms them when asymmetry is low. Similar results are found when considering allocating data property rights to firms or consumers (Hermalin and Katz, 2006); however, enabling consumers to control data sharing fully makes all consumers better off under a monopoly setting (S. N. Ali et al., 2020).

2.2. Ad targeting

The targetization of advertisement has seen a significant improvement in both scope and accuracy during the internet era, primarily due to the widespread availability of consumer data (Athey et al., 2013). Through consumer profiling, firms can reach precise targets, making specific consumers aware of the firm's presence in the market (Commission, n.d.). This strand of literature mainly focuses on a specific data effect: targeting can help firms reach consumers who otherwise would be unaware of their existence. Thus, unlike under price discrimination, data improves matching between firms and consumers and, in turn, the social value of advertising (Bergemann and Bonatti, 2011). However, some open questions remain regarding the benefits of data: do firms reach the social optimum when investing in ad targeting? Is there a threat of increasing market concentration?

A first result is given in a homogenous product market where advertising is costly (Roy, 2000): firms opt to create local monopolies when marginal ad costs are low, only targeting a share of the market and obtaining positive profits. This strategy allows them to avoid duplication costs (i.e., two firms sending an ad to the same consumer), which would increase competition and dissipate profits. While this strategy maximizes welfare, as the market is fully covered, all surplus is appropriated by firms due to market segmentation. Another valuable setting is competition between advertisers for ad slots, especially when one advertiser possesses data on users that visualize that specific slot. This information asymmetry allows the advertiser to identify valuable opportunities better. While the identification of peaches (i.e., high-valuation consumers) does not have a substantial impact on the bidding process, the ability to recognize lemons (i.e., low-valuation consumers) allows the informed firm to gain a considerable advantage over its competitors (Abraham et al., 2020).

The introduction of targeting inaccuracy, as well as richer settings, allowed for the further development of these insights. First, the link between an increase in targeting accuracy and a rise in firms' concentration shows to be robust even when considering firms competing over several markets (Bergemann and Bonatti, 2011) or when vertically separating firms between media firms, who show ads to consumers, and advertisers, who buy ad space (Rutt, 2012). Moreover, these additions allow studying the effects of firms' strategies regarding the social optimum. When targeting is poor (good), ad intensity is too high (low) with respect to the social optimum.

Other additions include limiting the effect of targeting through other means: the introduction of hiding technologies that allow consumers to block ads with a certain probability reduce consumer surplus due to consumers' inability to internalize the negative externalities they have on their peers (Johnson, 2013). In contrast, introducing a cap on available ad space can lead firms to prefer targeted advertising over general ads, increasing welfare (Athey and Gans, 2010). Another valuable addition is malvertising, i.e., malicious advertising such as spam or phishing. Suppose a firm offers a free service to consumers and can sell targeted ad slots to advertisers. If malicious advertisers buy the ad slots, consumers refrain from participating again in the service, as they experience a disutility. Depending on the probability of advertisers being malicious, the firm could be deterred from selling ad slots or be incentivized to screen advertisers to retain more consumers in future periods (Jullien et al., 2020).

Finally, a strand of literature recently focuses on mergers between digital platforms. These studies are timely due to recent developments in this sector, like the Facebook and Instagram merger in 2012. Concerning ad targeting, merged platforms are incentivized to limit the ad supply to maximize value, leading to a bottleneck (Prat and Valletti, 2022). Thus, this effect should be considered when analyzing platform mergers.

2.3. Cost reduction, revenue increase

Another strand of literature has analyzed the effects of data when they lead to cost reduction or revenue increase. In particular, the latter could be due to increased revenues extracted from consumers or data monetization. An example is provided by the search engine market, where accumulated search history improves search accuracy and thus decreases the cost of investing in quality. If one of the search engines has a more extensive search history log, this advantage can enable it to tip the market, reducing competition and welfare. These results also hold when considering a dynamic setting (Argenton and Pr A¼fer, 2012; Prüfer and Schottmüller, 2021). Mandatory sharing of search history would level up the field between firms to maintain competition, allowing a competitive oligopoly.

Data can also increase a product's quality premium through customization. Even when considering consumers who have a distaste for their data being used, the quality premium provided by data could be enough to attract new consumers in the market. Thus, data can have a positive welfare effect when they are quality-increasing (Campbell et al., 2015).

Finally, a more general approach is provided by setting up a model where firms directly supply utility to consumers, as in (Armstrong and Vickers, 2001). Through the competition in utilities, data effects can be identified as pro-competitive or anti-competitive, with minimal or no information required on market demand and competition intensity (De Corniere and Taylor, 2020). These results give insights regarding data collection policies: restricting data collection is only desirable when data are anti-competitive.

2.4. Discussion and existing evidence

In this section, I have reviewed models where data is already available to firms or where firms can source them at a marginal cost. This assumption often drives firms to overuse data due to the threat of facing rivals who could use data more aggressively (Thisse and Vives, 1988; Shaffer and Zhang, 1995; Bester and Petrakis, 1996; Taylor, 2003). The effects of data use strongly depend on firms' starting positions. While symmetric firms usually end up in a prisoner's dilemma, with lower industry profits and an increased consumer surplus, asymmetric firms can increase their advantage thanks to data: this can lead to dominant positions or even market tipping (Shaffer and Zhang, 2002; Gu et al., 2019; Belleflamme et al., 2020).

Regarding policy implications, the seminal works in this class of models often highlighted how welfare increased under data use due to more intense competition (Taylor and Wagman, 2014). However, introducing features that narrow the gap between the models and real-world situations changes this result. The presence of imperfect targeting (Y. Chen et al., 2001; R. Esteves, 2014; Mauring, 2021), loyal consumers (Anderson et al., 2019), limited ad capacity (Athey and Gans, 2010) or consumers' nuisance when their data are used (Jullien et al., 2020) lead to situations where consumers (or even the market as a whole) are worse off. Some papers have focused on data ownership to address these trade-offs. While results strongly depend on the models' assumptions, all papers where data have multi-faced effects conclude that introducing privacy policies would favor consumers (Hermalin and Katz, 2006; S. N. Ali et al., 2020).

Other works have instead suggested acting on privacy policies to increase welfare. However, there is disagreement regarding the optimal level of enforcement: while some papers advocate for a total ban of specific data uses (Shy and Stenbacka, 2016; Anderson et al., 2019), like price discrimination or ad targeting, others assess how limiting their use, like through the GDPR, would give better results (Mauring, 2021; Belleflamme and Vergote, 2016). However, some recent empiric papers argue that soft privacy intervention could disproportionally harm smaller firms, increasing asymmetries in the market. Batikas et al. (2020) highlight this phenomenon in the market of web providers after implementing GDPR: the policy introduction lowered the market shares of all firms in favor of the market leader, Google. Garrett et al. (2022) also observed a similar result regarding web technology vendors (e.g., advertising, web hosting, audience measurement). After implementing GDPR, websites reduced their demand for vendors while the concentration in the vendors' market increased. However, most of these shocks were reabsorbed by 2018 due to the market's growth or lack of enforcement related to GDPR.

Finally, another critical policy aspect regards market concentration. Due to the increased competition, some authors find that data use could hinder newcomers' entrance or how the number of firms in the market would reduce after adopting data (Bergemann and Bonatti, 2011; Taylor and Wagman, 2014). This effect can facilitate mergers between firms, which in turn can create asymmetries that could result in dominant positions amplified by the data use. Binns and Bietti (2019) show how the market of online third-party trackers significantly increased its concentration over time (with Alphabet firms being present in more than 70% of the analyzed sample). The same trend is observed by Batikas et al. (2020) in the web provider market.

3. Data from consumers

A second strand of literature has considered cases where data are endogenously created. For example, firms can gather consumer data on their first interaction with them, which can then be used later; consumers can actively decide how much data they want to share with firms. The endogenous creation or diffusion of data strengthens the link between firms and consumers: firms need to consider their externalities on consumers to predict their behavior and act accordingly. Regarding the data effects, a substantial part of this literature focuses on behavior-based price discrimination: firms store data regarding past purchases and can use this information to distinguish recurring consumers from new ones.

As this data effect is intertemporal, the consumers' ability to anticipate the effects of their data disclosure on their welfare in later periods is crucial in assessing their strategy. In particular, the seminal work of Coase (1972) highlights how consumer patience can effectively hinder a monopolist's ability to price discriminate in the case of durable goods. The intuition is that consumers can always anticipate the monopolist's strategy of lowering the price in later periods. Thus, if they are infinitely patient, they force the monopolist to always sell at a low price. In the following subsection, I give particular attention to the model's characteristics that effectively allow behavior-based price discrimination, such as limited time periods and consumers' naïveté.

3.1. Behaviour-based price discrimination

Behavior-based price discrimination (referred to as BBPD in this section) is a relatively recent form of price discrimination, which does not fall under the standard degrees of price discrimination described in the seminal work of Pigou (1920). The novelty of this practice lies in its dynamic feature: firms need to observe consumers' past purchasing behavior to discriminate them in a later period. The observation of consumer behavior can be seen as a data collection stage, where firms compete to obtain information that they can use later. Thus, BBPD falls within the boundaries of this chapter.

In the first works in this strand of literature, BBPD was defined as the ability to 'segment customers based on their purchasing histories and to price discriminate accordingly' (R. B. Esteves, 2009b). This type of price discrimination differs from those presented in Section 2.1, where data allowed firms to pinpoint consumers' preferences (first-degree price discrimination) or learn their characteristics (third-degree price discrimination), as BBPD only allowed firms to distinguish between new and recurring consumers. However, in more recent works, the term BBPD has been used more broadly to describe models where firms acquire data from consumers and use it to price discriminate, regardless of the type of price discrimination (see, for example, Choe et al. (2018)). To better follow the developments in this literature, in this section, I analyze models that follow the more recent and broader definition of BBPD. Comprehensive surveys on the subject can be found in Fudenberg and Villas-Boas (2004) and R. B. Esteves (2009b).

The first studies on BBPD focused on competitive two-period settings where consumers can switch firms between periods (R. B. Esteves, 2009a). In these models, firms compete in a first stage and distinguish between previous and new consumers in the second. Firms thus offer better deals to their rival's previous customers without lowering the profits made on their previous consumer segment. This strategy has been described as paying customers to switch (Y. Chen, 1997) or customer poaching (Fudenberg and Tirole, 2000). Offering discounts to the rival's customers increases competition, lowering firms' profits. Moreover, the increased customer switching leads to higher deadweight loss, lowering total welfare (Y. Chen, 1997; R. B. Esteves, 2010). Choe et al. (2018) expand on this model by allowing firms to operate first-degree price discrimination on the consumers they identified in the first period. Their results highlight how this scenario leads to two asymmetric equilibria where one of the two firms prices more aggressively than the other to maximize its market share in the first period. However, both firms are still worse off with respect to the basic model, where firms can only distinguish between new and recurring consumers, and total welfare is reduced with respect to the model without BBPD. The reduction in total welfare linked to the increased consumer switching is robust to various extensions, such as considering an infinitely-lived game with two firms (Villas-Boas, 1999) or considering behavior-based customization of a product instead of BBPD (Zhang, 2011). On the other hand, consumers do not switch in equilibrium if the switching costs are split between transaction costs (that consumers pay whenever they switch) and learning costs (that consumers only pay during their first interaction with a firm). An increase in transaction costs results in consumer harm since firms increase their price as they realize consumers face a higher cost when switching (Nilssen, 1992).

Other settings have instead highlighted how BBPD can benefit both firms and consumers. BBPD allows firms to escape the Bertrand paradox (i.e., pricing at marginal cost) and obtain positive profits while increasing consumer surplus in a homogenous good market. Conversely, total welfare still decreases due to excessive switching (Taylor, 2003; R. B. Esteves, 2009a). Another viable strategy is allowing firms to offer long-term contracts that lock consumers for both periods under a breach penalty threat. This strategy reduces switching, increasing consumer surplus and total welfare (Fudenberg and Tirole, 2000). Moreover, in an experience goods market, a monopolist can maximize his profits through BBPD depending on consumers' mean evaluation of the good (Jing, 2011). BBPD can also drive mergers: when BBPD is allowed, a two-to-one merge in a triopoly setting allows the merged firm to extract more surplus. The merger harms consumers but does not cause deadweight loss (R. B. Esteves and Vasconcelos, 2015).

Other authors have extended the basic setting by introducing features for either or both actors: these include allowing data sharing between firms, allowing more nuanced consumer behavior, and allowing consumers to have more control over their data, letting them decide over the amount of shared data or allowing them to use anonymizing technologies to avoid being identified. Data sharing can allow firms to identify consumers better, enabling them to increase their focus on the high willingness to pay ones. Data sharing has proven beneficial when firms' products are positively correlated (Taylor, 2004) and when firms are uncertain about the substitutability or complementarity of their products (B. C. Kim and Choi, 2010). Data sharing can also help firms infer consumers' willingness to pay if they learn how much a consumer paid for a rival's product (Baye and Sappington, 2020). If firms collect data from consumers in the first period and operate first-degree price discrimination in the second, then data sharing only occurs if firms are (enough) vertically differentiated. In particular, the low-quality firm always sells its dataset to the high-quality one. The sale occurs because some consumers purchase from the low-quality firm in the first period, even if they have a stronger preference for the high-quality one. Thus, in the second period, the high-quality firm benefits from identifying those consumers by purchasing its rival's dataset. While data sharing is Pareto-improving for firms, it harms consumers, who would be better off if data sharing was banned (Liu and Serfes, 2006). Similarly, data sharing harms consumers when it occurs between identical banks that compete on credit contracts, and consumers (entrepreneurs) incur positive switching costs. Under this scenario, banks can operate third-degree price discrimination on consumers they serve in the first period, identifying if their customers are talented or untalented entrepreneurs. By sharing their datasets, banks can target their offers exclusively to talented entrepreneurs, increasing their aggregate profits. Moreover, as banks anticipate data sharing in the second period, competition in the first period is also relaxed, and talented entrepreneurs pay higher prices (Gehrig and Stenbacka, 2007).

Instead, the conditions under which data sharing is profitable change when a consumer sequentially buys correlated products from two sellers, and the first one can sell information to the second. The first seller is better off by not sharing data if (i) its profits are independent of the second seller's profits, (ii) products' valuations are positively correlated, and (iii) the optimal contract between the second seller and the consumer is independent on the decisions of the first seller (i.e., preferences in the downstream relation are additively separable). In this case, the effect of data sharing is ambiguous: while it increases efficiency in the downstream contract, it reduces it in the upstream contract (Calzolari and Pavan, 2006). Argenziano and Bonatti (2021) expand on this model by allowing for heterogenous sources and uses for data and allowing the consumer to distort the information revealed to the first agent as a way to influence the second agent's behavior. Their main result highlights how the consumer can benefit from data sharing when the quality offered by the two agents is similar, as this scenario minimizes the consumer's incentive to distort his revealed information.

Considering more nuanced consumer behavior also gives valuable results. A first strand of literature has introduced forward-looking consumers in a BBPD setting: these consumers anticipate firms' BBPD in the second period and update their first period behavior accordingly. Suppose that consumers' preferences for two products are correlated and that the first firm can sell the purchasing history of its customer base to the second. Then, some high-valuation consumers would prefer not to buy from the first firm to avoid being identified from the second one (Taylor, 2004). Similarly, forward-looking consumers can also use future data sharing to credibly signal their low willingness to pay, leading to price concessions (Baye and Sappington, 2020). A related aspect of consumer behavior is analyzing how it changes when a firm discloses that it adopts BBPD. If the firm does disclose this information, results align with those of Fudenberg and Tirole (2000) and Fudenberg and Villas-Boas (2006): consumers' anticipation of the firm's BBPD in the second period drives the firm's profits downwards. If, instead, the firm does not disclose this information, consumers form beliefs about the firm's use of BBPD based on the price observed in the first period. In this scenario, the firm would be better off by committing not to adopt BBPD. However, consumers form the belief that the firm will use BBPD in the second period, as the firm cannot credibly commit to not using it, and the firm's profits are still lower than the benchmark case where BBPD is not feasible or where he can credibly commit to not adopting it (Li et al., 2020).

Similarly, when considering a product that gets improved between two periods, a firm would be better off without using BBPD: the segmentation would incentivize consumers to delay their purchase to only period 2, decreasing firms' profits (Fudenberg and Tirole, 1998). The same result holds when considering an infinite game without product improvement (Villas-Boas, 2004) or a situation where goods are horizontally differentiated in period one and homogenous in period two (Jeong and Maruyama, 2009). Firms can also use consumers' anticipation of their strategies to induce self-selection: in a market with loyal consumers and lowest-price buyers, setting high prices in the first period allows firms to drive the lowest-price buyers out of the market, softening competition. This strategy allows firms to expand their market, with ambiguous effects on consumer welfare (Y. Chen and Zhang, 2009).

Other additions to consumer behavior regard heterogeneity in purchase quantity and stochastic preferences. The former captures an empirically observed characteristic: a low share of consumers contributes to a large share of profits (Schmittlein et al., 1993). The latter better describes how consumer preferences change over time due to the specific purchase situation (Wernerfelt, 1994). When consumers show high heterogeneity or preferences exhibit enough stochasticity, BBPD can increase firms' profits even when consumers are forward-looking. Moreover, when consumers are heterogeneous, and their preferences change over time, then firms use BBPD as a tool to reward their own best consumers and prevent poaching; instead, if only one of the two persists, then firms are better off by rewarding their rivals' consumers (Shin and Sudhir, 2010). Shin et al. (2012) expand on this model by modeling consumer heterogeneity as heterogeneity on the cost to serve a specific consumer. The intuition is that some consumers overuse the services they pay for, resulting in a firm's profit loss over those specific consumers. Their result shows that if customer heterogeneity in their cost to serve is sufficiently high, then BBPD is profitable for the firm as it allows it to 'fire' the high-cost customers. On the other hand, low levels of heterogeneity drive the firm's profits downwards, as in Villas-Boas (2004).

A more recent strand of literature has instead focused on consumers' information disclosure when data improves matching and allows the firm to operate price discrimination. This literature adopts a setup similar to Bergemann and Bonatti (2015) but allows consumers to disclose data instead of exogenously giving it to the firm. On the one hand, consumers are incentivized to disclose information to allow the firm to offer them a more valuable good. On the other hand, information disclosure also allows the firm to tailor its prices and extract more surplus from the trade. In equilibrium, consumers disclose the least informative segmentation that still guarantees a trade with the firm. However, total welfare would be higher with more information disclosure, as the consumer-product matching would be improved (Hidir and Vellodi, 2021). Ichihashi (2020) expands on this model by allowing the firm to commit not to use consumer information for pricing. Through this commitment, consumers are incentivized to share more data, and the additional revenues generated from better matching more than offset the losses caused by not being able to price discriminate. Perhaps surprisingly, consumers are worse off under this commitment, as they cannot strategically disclose data to influence the firm's prices downwards.

Adding an anonymizing cost, together with the possibility of enhancing the product in period two and consumer heterogeneity, leads instead to ambiguous results: selling an improved product to high-valuation consumers can improve both firms' profits and welfare thanks to induced self-selection. However, BBPD is not always convenient: its profitability depends on the ratio between consumers' product valuations in both periods (Acquisti and Varian, 2005). Similar results are observed when past purchases create a consumer score, which is correlated to their willingness to pay: consumers can benefit from BBPD when the willingness to pay is high. However, this situation hurts consumers when their scores are concealed from them (Bonatti and Cisternas, 2020). Other studies also show how lowering the hiding cost can improve consumer welfare (Conitzer et al., 2012).

Finally, a crucial feature that has been recently studied is that of data externality, i.e., the correlation between data of different users. This topic has seen significant media coverage, primarily because of the Cambridge Analytica scandal, where the firm could infer data of 87 million users while only 320 thousand had given their consent (Schneble et al., 2018). This effect significantly impacts the market outcomes. Suppose that consumer data collected by a firm in the first period can also disclose information about consumers who have not shared their data with that firm and interact with it in the second period. Consumers do not anticipate the negative effect their disclosing has on other consumers, so their data disclosing strategy leads to consumer harm (Garratt and Van Oordt, 2021). Other works regarding data externalities that do not focus on BBPD are described in the next section.

3.2. Data monetization and as a revenue increasing factor

While price discrimination has been the most common focus when considering data collection from consumers, other relevant effects have been analyzed in the literature. Here I offer a brief overview of some of the most notable ones: data as a valuable good (i.e., data monetization), data as a quality-increasing factor, and data as a more general revenue-increasing factor. Data monetization has been primarily analyzed in online platform settings, where it is the primary source of revenue for firms. The basic result is that consumers can't correctly evaluate the value of their data in the transactions, leading platforms to collect too much data compared to what is socially desirable (Fainmesser et al., 2020). One suggested solution to this inefficiency is shifting data ownership to consumers (Dosis and Sand-Zantman, 2019; Jones and Tonetti, 2020). However, recent literature highlighted that shifting data ownership is not always welfare-enhancing once the public benefit of data is considered (i.e., data generated from one consumer benefits other consumers). Depending on the magnitude of this public benefit with respect to the individual disutility created by data, either the firm's ownership or consumers' ownership can be welfare enhancing (Markovich and Yehezkel, 2021). Another tool to temper data overcollection would be to enforce stricter privacy policies, such as default opt-out policies regarding data usage (Economides and Lianos, 2021). Another valuable scenario is where firms collect consumer data by providing a service at a price and then monetize data. This leads firms to have two possible revenue streams, as they can either focus on making profits through their service by charging positive prices or subsidize consumers to maximize data collection. Under vertical differentiation, a high-quality firm opts for the former, while a low-quality firm opts for the latter. The effects of competition on consumer privacy (i.e., the amount of disclosed data) are ambiguous: while fewer data are disclosed under competition than under monopoly, a high level of competition leads to more disclosed data, as low-quality firms increase subsidies to consumers (Casadesus-Masanell and Hervas-Drane, 2015). Zogheib, Bourreau et al. (2021) expand on this scenario by studying the effects of consumer multi-homing with horizontally differentiated firms. The introduction of multi-homing reduces the data monetization channel, as firms are no longer exclusive dealers of those data points. Then, the firms' choice of privacy regime crucially depends on the value that consumers obtain from multi-homing and the value of multi-homing consumers' data. In particular, the authors highlight how firms may end up sharply increasing their data disclosure levels, especially if the value of multi-homing consumers' data is not too low when compared to those of single-homing consumers.

The concept of firms having two distinct profit channels (namely, the sale of the product and data monetization) is pivotal when analyzing the effects of data portability, as mandated by Article 20 of the European GDPR.[2] In particular, it is important to note that data portability only applies to user-generated data and not to data that is derived by firms. On the one hand, data portability alleviates switching costs between competing firms, as consumers do not need to provide their data *ex novo* to a new entrant. On the other hand, as consumers are less preoccupied of lock-in effects, they would tend to increase their data sharing with the incumbent. The increase in user-generated data sharing also improves the incumbent's ability to derive proprietary data, which in turn gives it a competitive advantage over entrants. When data are valuable enough, the latter dominates the former, and data portability reduces entry (Lam and Liu, 2020). From a consumer perspective, data portability skews the trade-off between the value of providing their data and their privacy concern towards the former, as the transaction costs to provide data are reduced. If firms anticipate this, they are, in turn, incentivized to collect and sell more data than without data portability. As the data monetization

[2]A detailed description of the economic implications of data portability can be found in Krämer (2021).

channel becomes more profitable, firms are less willing to fiercely compete over the price of the good they provide, leaving some consumers worse off as they set higher prices (Krämer and Stüdlein, 2019).

Data can also improve services' quality: the most common example is search engines, where both individual and collective search histories improve search accuracy. When considering markets for homogenous products in an infinitely lived game, firms benefit more from a steeper learning curve than a larger starting data stock. In this situation, consumers would benefit from data sharing since both across-user and within-user learning would be improved. However, if firms can anticipate such a policy, they opt to soften competition and decrease consumer subsidies, leading to a decrease in welfare (Hagiu and Wright, 2020). A similar application is when consumers' gross utility depends on the data they disclose and a firm's investment in quality: for example, a social media platform could create new and more effective sharing tools, leading consumers to increase their use of the platforms and share more data. Under this scenario, the platform under-provides quality and over-collects data with respect to the social optimum. A regulator could impose a cap on consumer data disclosure levels (e.g., setting limits on which data types may be disclosed or posing conditions regarding the type of third parties the platform can sell data to). However, the effect of a disclosure cap on welfare is ambiguous, depending on whether the cap introduction reduces consumer participation on the platform and on the complementarity between disclosed data and quality investment in determining consumers' utility (Lefouili and Toh, 2017).

Finally, data has also been modeled as a more generic revenue-increasing factor, allowing for more general insights that abstract from specific effects. In particular, some of these works have focused on data externality, highlighting additional implications that stem from this effect. First, consumers can anticipate how data sharing can be harmful to them but do not consider the (positive or negative) spillover effect their sharing has on other consumers, leading to potentially inefficient outcomes. Moreover, data externality also affects non-users, as the firm infers non-users' data from its users. This leads to excessive loss of privacy compared to the social optimum (Choi et al., 2019). Second, consumers with a low valuation of privacy share their data more easily; however, data externality undermines the value of privacy for all users, as firms can infer high-valuation consumers' data without their consent. Thus, consumers' data valuation is reduced, again leading to too much data collected in the market compared to the social optimum. An effective solution to overcollection would be to add a third-party mediator between the user and the firm. If a user wants to share its data with the firm, the third-party mediator could transform these data so they are still informative regarding the user but do not reveal information on users who do not want to share their data. This intermediation would effectively avoid harmful data externalities, increasing welfare (Acemoglu et al., 2022).

3.3. Discussion and existing evidence

In this section, I have reviewed models where firms obtain data from consumers: data can then be used in subsequent periods or directly be the firms' income source (i.e., data monetization). One of the most analyzed data effects is Behaviour-Based Price Discrimination (BBPD): firms collect data on consumers who buy from them and can thus distinguish recurrent consumers in subsequent periods. In recent years, the term BBPD has instead been used more broadly to identify models where prior interaction with consumers allows some form of price discrimination from firms, regardless of the type of discrimination. BBPD leads to increased competition in markets with horizontal differentiation, dissipating profits (Y. Chen, 1997; Fudenberg and Tirole, 2000; R. B. Esteves, 2010; Choe et al., 2018). However, BBPD can be beneficial when considering homogenous goods (Taylor, 2003; R. B. Esteves, 2009a) or experience goods markets(Jing, 2011). Moreover, BBPD can also favor mergers in high-concentration markets (R. B. Esteves and Vasconcelos, 2015). Most of these models also describe a deadweight loss caused by excessive consumers' switching when firms can distinguish previous consumers.

The increase in competition caused by BBPD benefits consumers, who obtain lower prices. However, firms can gain an advantage by sharing their data: this strategy allows for better targeting, hurting consumers (Taylor, 2004; B. C. Kim and Choi, 2010; Baye and Sappington, 2020; Liu and Serfes, 2006; Gehrig and Stenbacka, 2007). This outcome thus emphasizes the role of consumer privacy in market outcomes; a survey on this subject can be found in (Acquisti et al., 2016). Many of the models where data sharing is analyzed called for stricter policies on consumer privacy to rebalance the market outcomes in favor of consumers. On the other hand, a study from Aridor et al. (2020) shows how implementing more straightforward and more efficient ways to protect consumer privacy, such as the explicit opt-out introduced by the GDPR, can have unintended consequences: by abandoning more inefficient ways to protect their privacy, like browser-based privacy protection, privacy-conscious consumers reduce the noise created by short and spurious consumer histories, increasing the value and the traceability of the remaining consumers.[3] However, in markets where data can improve goods, such as search engines, mandatory data sharing between firms can benefit consumers, as a lack of starting data can act as an entry barrier, limiting competition in the market (Hagiu and Wright, 2020; Schäfer and Sapi, 2020). However, empirical evidence shows how the search engine's learning curve has significantly more impact than the initial data stock (Chiou and Tucker, 2017): thus, mandatory sharing could result in ambiguous effects.

Finally, a relevant aspect when firms obtain data from consumers is data externality: this can easily undermine consumers' valuation of privacy, leading to excessive data in the market (Choi et al., 2019; Acemoglu et al., 2022). This issue raises concerns about privacy policies: if firms can infer consumer data from users who are correlated with them, local policies such as GDPR would be heavily hindered. The topic of data externality has been central in discussing the Google/Fitbit merger case, which the European Commission cleared in 2020. Indeed, one of the main points of the merger's opposers is that Google's acquisition of Fitbit's health data could result in Google inferring information regarding non-users, leading to consumer harm (Bourreau et al., n.d.-a). Moreover, the efficiency gains from the merger could result in a drop in prices in the market where data is collected (i.e., Fitbit) and in a rise in prices in the market where data is applied (i.e., Google advertising or the digital health sector), where personalized pricing increases the per-consumer revenue. If the efficiency gains are significant enough, the merger could lead to monopolies in both markets (Z. Chen et al., 2022). On the other hand, the DGCOMP Chief Economist agreed with the Commission's decision, stressing how the Commission posed remedies to possible threats, such as Google limiting interoperability of Fitbit's wearables with non-Android devices by asking Google for a commitment to maintaining interoperability for at least ten years. Moreover, he highlights how the theory of harm regarding data externality lacks evidence, as there are no studies on the marginal value of data on advertising revenues, let alone on digital health data (Regibeau, 2021).

4. Data from intermediaries

The most recent and steadily growing literature strand focuses on data intermediaries: these firms collect massive amounts of data that they then sell in downstream markets. Their upstream position allows them to consider data externalities to their full extent:

[3] For a survey regarding empiric studies on consumer attention towards privacy, see Acquisti et al. (2015).

while an individual firm aims at maximizing its profits, a data intermediary's goal is to maximize the value of data in the downstream market that it can then extract from purchasing firms. In particular, the literature has been split into two major strands: one regarding data brokers and the other regarding attention platforms. While these data intermediaries gather massive amounts of data that they can then strategically sell in downstream markets, their main difference lies in the data collection process.

As highlighted by a recent Commission (n.d.) report, data brokers' data collection process rarely includes direct interaction with web users, as they collect consumer data through online and offline means (Bounie et al., 2021b). Consumers are often unaware that their personal data is present in their databases, and data brokers do not have to consider this negative externality when collecting data. Transferring this approach in theoretic models, data brokers are usually modeled as actors who already have consumer data or collect them by paying a marginal cost: the strategic interaction only happens between the data broker and downstream firms, not consumers.

On the other hand, attention platforms are at the forefront of data collection, as their business model focuses on gathering data by providing services, usually for free (Prat and Valletti, 2022). A distinctive trait of these platforms is that they usually need to strike a balance between the service quality they provide and their revenue stream: while increasing the latter would result in higher profits in the short term, it could result in less traffic on the platform in the long run, leading to fewer data gathered and also fewer transactions (Evans, 2019). Attention platforms thus need to consider both consumer-side externalities in the data collection process and firm-side externalities in the data sale since a misuse could drive users away from the platform.

4.1. Data brokers

The inclusion of data brokers in competition models introduces an additional level of strategic decisions, regardless of the effect of data at hand. Data brokers can serve more firms that belong to the same market: as such, these actors strategically choose to whom and which data to sell. Compared to settings where firms obtain data without strategic interactions (as in Section 2), data brokers consider additional externalities when choosing their strategy. The main difference is that data brokers consider how selling data to one firm impacts the others, as this externality impacts firms' data valuation and, in turn, the data broker's profits. A first strand of literature has focused on monopolistic data brokers: this assumption is not far-fetched, as the data broker industry is highly concentrated, and their combined market value is estimated at USD 156 billion per year (Pasquale, 2015). In these scenarios, data brokers choose their strategy by only focusing on the resulting outcomes in the downstream market, as they do not have to worry about competitors. In particular, their strategy crucially depends on the amount of market power they have, which is usually tied to the selling mechanism through which they sell data.

As a guiding example, consider a setting as in Thisse and Vives (1988),[4] but where firms can buy data regarding all consumers from a data broker instead of having them exogenously like in the basic model. Intuitively, a firm's willingness to pay for data will be equal to the difference in its profits between buying or not buying them. First, suppose that the data broker offers the dataset to both firms. This strategy decreases the firm's willingness to pay for data, as when both firms obtain the dataset, they dissipate profits in price wars as in Thisse and Vives (1988). This result holds for a variety of selling mechanisms, such as auctions, take-it-or-leave-it offers, and sequential bargaining (Bounie et al., 2020).

Second, suppose instead that the data broker only wants to serve one of the downstream firms. In such a scenario, the selling mechanism becomes crucial in assessing the data broker's preferred strategy and the related market outcomes. If the data broker can auction the whole dataset so that only the winning firm will obtain it, winning the auction (and, in turn, obtaining the dataset) also result in the rival not obtaining it, and vice-versa. Thus, firms' willingness to pay for the dataset is at its highest, as winning implies being informed and competing against an uninformed rival, and losing implies being uninformed and competing against an informed one. Thus, if the data broker can change the offer he makes to a firm conditional on the other firm's choice, he will prefer to underserve the market and only sell data to one firm (Montes et al., 2019). In particular, if firms are vertically differentiated, the data broker opts to sell data to the firm which is more advantaged with respect to the quality-adjusted cost differential (F. and Valletti, n.d.).

Subsequent literature has also shown that the data broker can be better off by selling smaller partitions of the dataset instead of selling it whole. The intuition is that, even if only one firm obtains data in equilibrium, the *competition effect* of data dissipates profits as the uninformed firm lowers its price, trying to compete against the informed

[4]For a detailed description of the model, see Section 2.1.

one. Instead, selling partitions that only contain the location of consumers who are close to a firm's position tempers the *competition effect*, leading to higher profits and, in turn, a higher data price (Bounie et al., 2021b). Regardless of the change in the data broker's selling strategy, the importance of the data broker's selling mechanism remains unchanged. If the data broker can set up an auction mechanism or bargain sequentially between the two firms, he always prefers to only serve one of the two in equilibrium. Conversely, if the data broker cannot change the offers made to a firm based on the rival's behavior, such as with take-it-or-leave-it offers, he instead opts to serve both firms in the downstream market (Bounie et al., 2020).

The previous studies focused on duopoly settings and highlighted how exclusive deals in data sales emerge when DBs have enough bargaining power. However, expanding this setting to a circular city with three firms highlights how a DB always benefits from selling to more than one of them (Delbono et al., 2022). Moreover, allowing endogenous firm entry might lead to different results. Under this scenario, the high data price results in a barrier to entry, reducing competition in the downstream market and harming consumers regardless of the data broker's bargaining power (Abrardi, Cambini, Congiu et al., 2022). Consumer data availability can also ambiguously influence mergers between firms: in a two-to-one merger, the presence of a data broker exacerbates the anti-competitive effect, leading to consumer harm. Instead, in a three-to-two merger, consumers benefit from the increased efficiency caused by the merger and data. However, the data broker still benefits by only serving the merged firm, while consumers would be better off if he also serves the non-merged firm (J. H. Kim et al., 2019).

The incentive for the data broker to temper the data-induced rise of competition remains consistent even when analyzing different data effects. When data allows for improving firms' products through differentiation, a data broker opts only to sell data that increase the product value for consumers who are loyal to one of the two firms. This strategy allows him to temper the rise in competition given by introducing data in the downstream market. On the other hand, if forced to sell data to both firms, the data broker would also sell competition-increasing data that would allow a firm to conquer its rival's customers. However, this strategy does not raise competition in the downstream market, as firms are better off buying all the data and then refraining from using competition-increasing ones. The fact that their rival has all the data is used by the data broker as a tool to threaten a potential non-buyer, increasing firms' willingness to pay for data (**iyer2000marketser**).

Another available strategy is separating access to consumers from the sharing of their information. For example, a data broker can do this by auctioning ad slots and deciding whether to grant access to consumers or share consumer information with the winning firms. Disclosing consumer data implicitly reveals information regarding their valuation of the winning firm's product to its rivals, increasing prices in the market. As such, information disclosure can be beneficial for the data broker, especially when the number of downstream firms is large (De Corniere and De Nijs, 2016). When firms are heterogeneous, or the data broker has incomplete information regarding them, the data broker can also guide the market outcome by offering a menu of contracts to achieve firms' self-selection (Arora and Fosfuri, 2005; Bergemann et al., 2018). Another way to handle firms' heterogeneity by the data broker is to sell access to individual consumers through tailored queries (i.e., cookies). Through this selling mechanism, firms can opt for positive targeting (buying information of consumers they want to reach) or negative targeting (buying information of consumers they want to avoid), and the data broker achieves their self-selection (Bergemann and Bonatti, 2015).

On the other hand, when considering long-lived data that can be used over multiple periods, a data broker opts to sell low-accuracy data in the first period to temper the competition with his future self (Cespa, 2008). Toggling the data accuracy can also be used to reduce competition in the downstream market to extract more profits: a data broker would opt to sell low-accuracy data to all firms or high-accuracy data only to a subset of them (Garcia and Sangiorgi, 2011; Kastl et al., 2018). Similarly, a search engine is incentivized to provide suboptimal matching to competing firms: even if the marginal cost of advertising is passed to consumers by firms, the increase in competition given by optimal matching would still result in a decrease in firms' profits and, in turn, of the search engine's ones (De Corniere, 2016).

Choosing the correct level of data accuracy also depends on the type of competition. When firms' actions are strategic complements (i.e., competition is à la Bertrand), the data broker opts for high accuracy, as data increases firms' profits that he can then extract through the price of data. On the other hand, when firms' actions are strategic substitutes (i.e., competition is à la Cournot), the data broker opts for low accuracy, as data would negatively affect firms' profits due to a considerable increase in downstream competition (Bimpikis et al., 2019).

Other strategies emerge when considering peculiar settings for data usage. A first example is when a data broker offers a search service that allows consumers to buy from their preferred stores. As the data broker is paid depending on the total visits in both stores, he has the incentive to divert consumers towards their least favorite store: this increases the number of consumers who visit both stores and influences firms' pricing strategies downward as they try to retain some of the misplaced consumers (Hagiu and Jullien, 2011). A similar setting is also studied when downstream firms can pay the data broker with their own consumer data instead of money to gain more prominence and attract more consumers: consumer data is then sold in an external market by both the firms and the data broker, providing an additional level of competition. Moreover, the data broker already has some consumer data: thus, data obtained from firms are deemed exclusive if the data broker does not possess them and non-exclusive otherwise. While both exclusive and non-exclusive data are valuable in the external market, exclusive data grant a higher marginal revenue. When the value of non-exclusive data is high, the data broker can achieve higher profits by making firms pay for prominence with consumer data rather than money, while the opposite occurs if the value of non-exclusive data is low. In both cases, introducing paid prominence weakens the competition between firms, leading to lower consumer surplus (Bourreau et al., n.d.-b).

Finally, a recent strand of literature considers competition in the data brokers' market. Bounie et al. (2021a) obtained a first set of results, expanding the setting in Bounie et al. (2021b) by introducing competition between DBs. Their main findings highlight how competition between DBs increases competition between firms, benefiting consumers and lowering the amount of data collected. Other works have instead focused on the non-rivalrous nature of data, considering the possibility of overlaps and synergies between two data brokers' datasets. First, the degree of data accuracy and the correlation between datasets influences the buyers' behavior: a high correlation, and likewise a high degree of accuracy, would result in firms seeing the datasets as substitutes, increasing the competition between data brokers. Moreover, a high level of competition between downstream firms further increases the degree of datasets' substitutability: as such, data brokers would prefer opting for exclusive deals with downstream firms to obtain higher profits. This strategy is reinforced by the fact that a first-mover advantage is observed: as such, a firm would pay a higher price for data if it can guarantee that its rival remains uninformed (Sarvary and Parker, 1997; Xiang and Sarvary, 2013).

Another crucial topic is that of dataset merging. Indeed, different datasets could, for example, be sub-additive if they contain overlapping information. However, merging datasets could also be supra-additive, thanks to complementarities and synergies with data types. As a guiding example, suppose that a dataset only contains the browser's histories and IP addresses of a set of consumers while another contains IP addresses and the respective e-mails. If a firm only obtains one of the two datasets, it could only obtain an indication of the average tastes of the consumer set (first dataset) or a way to reach them individually without knowing their tastes (second dataset). However, if the two datasets are merged, the firm would then be able to price discriminate consumers as it would learn their tastes and also obtain a mean to reach them individually. Moreover, data brokers can anticipate these data synergies and choose whether to compete or co-operate based on their datasets' complementarity. If data are sub-additive, data brokers prefer to merge their datasets: as the downstream firms see the datasets as substitutes due to data overlaps, merging them allows the data brokers to avoid fierce competition. Instead, when data are supra-additive, data brokers opt not to merge them as firms see them as complements (Gu et al., 2022). The complementarity between datasets is crucial in explaining why data brokers are often reported to trade data and start partnerships among each other (Commission, n.d.).

Other models have instead focused on data effects resulting from types of competition typical of attention platforms, without however modeling the interaction between the data intermediaries and consumers. A first notable effect is the control data brokers have over the firm's ability to reach consumers. The typical real-world examples are social networks, where platforms allow firms to show targeted ads to specific consumer segments. While platforms are paid per ad, firms' valuation of the ads depends on the competition they face for consumers: platforms face a trade-off between the number of ads sold and their individual value. If competition between platforms is mild, platforms opt to limit the ad supply, artificially creating a bottleneck that maximizes the ads' value at the expense of entrant firms (Prat and Valletti, 2022).

Another peculiar practice recently analyzed in the literature is that of social logins. This practice offers consumers a fast registration channel to many online services and allows information sharing between these services and the corresponding platform that offers the social login. Information sharing has ambiguous effects on online services, depending on the targeting improvement. Moreover, competition between platforms that offer social logins benefits consumers and increases the number of situations where social logins are offered to online services (Kramer et al., 2019).

4.2. Attention platforms

The business model of data-centered attention platforms may seem relatively simple: the platform offers content, usually for free, to consumers who agree to disclose some of their data to the platform. The platform then monetizes this information by selling datasets or selling access to the identified consumers to advertisers. This second strategy usually involves targeted advertising: the platform auctions a series of ad slots shown to specific consumer segments, and advertisers compete to obtain them. However, this simple mechanism hides a series of nuances that must be considered to model its functioning correctly.

First, the content offered by the platform is valuable for consumers, and they choose which platform to home based on this evaluation. Second, advertisements are a nuisance to consumers, as they get in the way of the content's consumption. As such, the platform must consider the trade-off between the increasing revenues given by additional ads and the decreasing consumers' utility. Third, the advertisement demand by advertisers and the content supply by the platform are positively correlated, creating a feedback loop: a higher demand for ads increases the value of individual ads and thus incentivizes the platform to increase the amount of content provided to attract more consumers (Evans, 2019, 2020).

Starting from this basic mechanism, the existing literature has expanded upon it by introducing additional factors observed in the real world. First, consumers can have negative externalities on their peers when disclosing their data: this happens when an individual's information is predictive of the behavior of others. This negative externality can allow the platform to obtain consumer information at a lower cost, as it reduces consumers' data evaluation. On the other hand, the platform opts to sell data to advertisers at an aggregate level to partially preserve consumers' privacy. This strategy allows the platform to capture the total value of information as the number of consumers becomes large (Bergemann et al., 2020).

Second, allowing advertisers to also reach consumers through secondary channels instead of only through the platform allows them to better assess the platform's influence on the market. In this scenario, advertisers can better target consumers by serving them through the platform but must pay a fee in return. On the other hand, consumers can decide how much information to share with the platform and whether they prefer buying through it or the outside market. As information becomes more precise, the value of data increases: consumers prefer to buy on the platform more, and the platform's market power over firms increases. This leads to two peculiar effects. First, as advertisers' profits in the offline market decrease, showing up on the platform becomes necessary. As such, the platform has almost complete control over their access to consumers, leading to a gatekeeper effect. Second, if the platform can compete with the advertisers it hosts, a copycat effect is observed: the platform can use the competitive advantage given by data to outclass them on their respective products, further decreasing their profits. Combining these two effects leads to a decrease in advertisers' participation in the market, as only a subset of them can sustain the cost of being active on the platform (Kirpalani and Philippon, 2020).

Another strand of literature has instead focused on competition between platforms. A key characteristic is whether consumers are single-homing (i.e., they only use one of the platforms) or multi-homing (i.e., they can use more than one platform). If consumers single home, they self-select into the platforms depending on the content they provide: this, in turn, creates a competitive bottleneck, as both platforms are the only channel through which an advertiser can reach a specific consumer. Perhaps surprisingly, if the two platforms were allowed to merge, this would increase the number of ads, while the effect on total welfare is ambiguous. Moreover, the entry of a new platform would result in higher prices for advertisement, as each platform still holds an exclusive channel to a subset of consumers (Anderson and Coate, 2005). Two key assumptions mainly drive these results: consumers can absorb any number of ads without them losing their value, and they single-home.

First, assuming that consumers can absorb any number of ads does not consider consumers' limited attention span. Introducing a limit on the number of ads that consumers register creates an additional externality between platforms: increasing the number of ads on a given platform negatively affects all the other platforms, as consumers reach their limit faster. As such, the entry of a new platform reduces the value of advertisement, as the higher the number of firms, the less they internalize the negative congestion externality, leading to a lower price per ad. The opposite is observed if two firms merge

(Anderson, 2012). By also introducing a limit on the time available to consumers, increased platform entry has an ambiguous effect on consumer welfare. On the one hand, consumers benefit from the increased differentiation; on the other, more entry increases the number of ads, thus reducing the time that consumers can spend consuming content. Which of the two effects dominate depends on the level of platforms' asymmetry (Anderson and Peitz, 2020).

Second, while the single-homing assumption can be a good approximation for some markets, like traditional TV broadcasting, the rapid growth of online advertising and streaming allows a sizeable share of consumers to multi-home. Allowing for consumer multi-homing solves the aforementioned competitive bottleneck, as advertisers now have multiple channels to reach specific consumers. For example, consumers could single-home in a given period but switch platforms between periods. This creates an overlap, as advertisers could reach the same consumers multiple times, wasting some advertisements. In turn, also advertisers must choose between single-homing and multi-homing: specifically, high-value advertisers opt for multi-homing, while low-value single-home (Athey et al., 2013). Allowing for multi-homing in the same period also brings similar results. In particular, introducing a correlation between consumer tastes shows how a positive correlation between platforms' contents increases the share of multi-homing consumers, leading to lower advertising levels (Ambrus et al., 2016). Similarly, allowing platforms to choose the genre of their content shows how platforms would opt for a high level of horizontal differentiation to maximize their share of single-homing consumers (Anderson, 2018).

However, this last result depends on the assumption that the amount of multi-homing consumers is exogenous. If, instead, each consumer can choose whether to single or multi-home, platforms opt for head-on competition by reducing differentiation. While this strategy lowers platforms' demand, only a handful of consumers will multi-home, as the platforms' services are close substitutes. As such, platforms obtain larger shares of single-homing consumers, granting them market power when bargaining with advertisers (Haan et al., 2021).

Other remarkable insights emerge when considering the addictiveness of platforms. This feature is different from the content quality provided by platforms: while content quality influences consumers' utility of participation on a platform, addictiveness influences their marginal utility in allocating attention. Platforms can control the level of addictiveness through many UI choices, such as intrusive notifications systems or infinite scrolling. When platforms can endogenously choose the level of addictiveness, they opt to sacrifice quality for attention when consumers have a tight constraint on their attention level: in this situation, addictiveness does not change total attention. However, it influences how consumers split their attention between platforms: as such, increased competition could incentivize platforms to raise addictiveness and steal consumers from their rivals (Ichihashi and Kim, 2022).

Finally, consumer multi-homing can also be detrimental when platforms anticipate it and know whether advertisers' use of consumer data benefits or harms them. If data are non-rivalrous, consumers can share their data with multiple platforms and earn compensation from all of them. Anticipating this, platforms compete less aggressively for data to reduce the amount of overlap between their datasets and increase their market power with intermediaries. This strategy allows platforms to earn positive profits at the expense of consumers (Ichihashi, 2021).

4.3. Discussion and existing evidence

In this section, I have analyzed models where data intermediaries can strategically collect and sell data in a downstream market and how their presence can influence market outcomes. Due to their position in the market, data intermediaries internalize a larger share of data externalities than other actors such as consumers and downstream firms (Bergemann and Bonatti, 2019). As observed in most of the literature, data often have a pro-competitive effect on firms since they allow them to identify consumers better (Thisse and Vives, 1988). This increase in competition can effectively deplete firms' profits, reducing their willingness to pay for data. Thus, data intermediaries adopt strategies that allow them to temper competition in the downstream market to increase firms' profits that the intermediaries can then extract.

Data intermediaries can temper competition in multiple ways: an intermediary could, for example, sell data of different consumers to different firms (Iyer and Soberman, 2000), only serve a subset of firms (F. and Valletti, n.d.; Montes et al., 2019; Bounie et al., 2021b, 2021a; Abrardi, Cambini, Congiu et al., 2022), or achieve firms' self-selection through various levers when firms are heterogeneous (Arora and Fosfuri, 2005; Bergemann and Bonatti, 2015; Bergemann et al., 2018). In all these cases, firms' profits are reduced as they either spend most of their profits on buying data or compete against rivals who obtain data and thus have a competitive advantage over them. This effect can effectively limit firms' entry into the market, further decreasing competition and harming consumers Abrardi, Cambini, Congiu et al., 2022.

On the other hand, data intermediaries can also have the incentive to increase competition, depending on how firms pay for their services. If a data intermediary is paid depending on the number of consumers he brings to the firms, he opts to divert consumers towards their less preferred firm so that a single consumer visits both firms, allowing double marginalization (Hagiu and Jullien, 2011). In this setting, the increase in competition between downstream firms is beneficial for the data intermediary, as it attracts more consumers that use his service to reach firms. This practice of consumer steering has also been observed in empirical works when a platform can compete with the firms it hosts. For example, Amazon products sold over Amazon are more recommended than its competitors' products; moreover, third-party products get fewer recommendations whenever Amazon runs out of the corresponding product (N. Chen and Tsai, 2019). A similar practice of steering has also been observed as a tool to maximize the value of ads: Facebook advertisements are skewed towards certain demographic groups despite neutral targeting settings, leading to discriminatory ad delivery (M. Ali et al., 2019).

Strategic interactions between data intermediaries also allow them to temper competition: the literature highlights scenarios in which data intermediaries can increase their profits by merging their datasets (Gu et al., 2022). If firms view data intermediaries' datasets as substitutes, a merger becomes a strategic tool to temper substitutability and, thus, competition. These results align with the observed practice of data trading between intermediaries (Commission, n.d.). Although the literature agrees on the inherent risks stemming from the figure of data intermediaries, empirical evidence is mostly lacking. This is due to the secretive nature of data transactions (Montes et al., 2019) and the black-box mechanisms through which platforms assign ads to consumers (Pasquale, 2015). Some recent studies have first tried to isolate the effect of data on both firms' performance and ads accuracy, to better understand the amount of market power data intermediaries have. First, they highlighted how product data could improve demand forecast, especially if the dataset contains many time periods (Bajari et al., n.d.). Second, they showed how targeting consumers through data intermediaries is more inaccurate than commonly assumed in the literature, with platforms outperforming data brokers regarding accuracy levels (Neumann et al., 2019).

The dominant positions of data intermediaries in the digital markets have raised policymakers' concerns on whether and how to intervene, bringing data intermediaries to the forefront of competition policy. Antitrust authorities have thus opened many investigations and requested various reports on data brokers and attention platforms to better assess the underlying market dynamics and their perils (for a survey, see Lancieri and Sakowski (2021)). One of the main problems in these analyzes is the zero-price nature of attention platforms toward consumers: this feature disables many instruments of antitrust analysis, such as the SSNIP test, that rely on pricing dynamics (Newman, 2020). Recent literature has thus advanced some proposals to overcome this issue.

First, an Attention-SSNIP test has been proposed as a tool to better assess market relevance when considering attention platforms. Increasing the consumers' nuisance when using a free product, for example, introducing a mandatory time-out before launching a search query on Google, would allow mapping how consumers react to this figurative increase in cost and which services they opt for instead (Wu, 2017).

A second proposal has been to use the amount of data shared by consumers as a proxy for the service's price. An exploratory study in this direction showed that Facebook overcharges consumers with data, hinting at a possible dominant position in the market (Summers, 2020).

Finally, a very recent development aims at subverting the way these platforms are analyzed. By focusing on consumer attention instead of the derived data, attention can be seen as a scarce, rivalrous, and tradeable product, more in line with standard economic definitions. Moreover, situating humans as attention producers and data intermediaries as distributors allow advertisers to be seen as the final consumers, making attention markets closely resemble familiar top-down distribution systems (Newman, 2020). These changes in how we think about platforms' business models could help policymakers better frame their market dynamics and better understand instances of platforms' abuse.

5. Conclusions

In this chapter, I have reviewed the effects of data on market outcomes. This theme is especially relevant due to the fast expansion of the digital economy, which heavily relies on data collection and consumption, and the growing concerns regarding the market leaders controlling much of the data market. By organizing the literature based on how data collection is modeled, I extracted more general insights that can help inform policymakers on if and how to intervene in case of controversies. Moreover, this organization can help authors better identify gaps in the literature to guide future research.

When data are widely available, they often have a pro-competitive effect in the market. This can, however, hinder newcomers' entrance into the market, leading to higher concentration. Firms also have a limited internalization of data externalities, often leading to data overuse: this raises privacy concerns that policymakers should consider when accounting for the effects of data.

I also observe similar outcomes when firms need to strategically interact with consumers to obtain data. However, these models repeated nature also highlights how data can facilitate market tipping, especially if firms are asymmetric in their starting positions. In these situations, data use exacerbates the advantage of the high-value (or better-informed) firm, increasing concentration. Firms can also strategically trade data to limit their interaction with consumers, allowing them to extract more profits: this is especially relevant when consumer data are correlated since a small dataset allows firms to infer information regarding large shares of the consumer base.

Finally, the introduction of data intermediaries further underlines the perils of unregulated data collection and usage. Data intermediaries can strategically sell their datasets to temper competition in the downstream market, allowing them to extract more profits at the expense of both firms and consumers. Their high market power stems from their quasi-monopolistic positions and the ability to strategically coordinate their actions in the case of competition between intermediaries. These concerns emerge regardless of the specific data effect, suggesting that further research is needed to better regulate these actors.

Overall, the theoretical literature is mostly in line with the scarce empirical evidence on the subject: however, the high specificity of most models limits their applicability, as it is difficult to untangle the individual effects of the various assumptions. From this point of view, models that make milder assumptions on both the type of competition and the effects of data allow for broader insights that can be extremely helpful when trying to understand the overall effect of data (for example, see De Corniere and Taylor (2020)). Further research is also needed on the empirical side, as real-world evidence can help uncover novel mechanisms that theoretical models can then investigate.

Bibliography

Abraham, I., Athey, S., Babaioff, M., and Grubb, M. D. (2020). Peaches, lemons, and cookies: Designing auction markets with dispersed information. *Games and Economic Behavior*, *124*, 454–477.

Abrardi, L., Cambini, C., Congiu, R., and Pino, F. (2022). User data and endogenous entry in online markets. *Available at SSRN*, *4256544*.

Abrardi, L., Cambini, C., and Rondi, L. (2022). Artificial intelligence, firms and consumer behavior: A survey. *Journal of Economic Surveys*, *36*(4), 969–991.

Acemoglu, D., Makhdoumi, A., Malekian, A., and Ozdaglar, A. (2022). Too much data: Prices and inefficiencies in data markets [forthcoming]. *American Economic Journal: Microeconomics*.

Acquisti, A., Brandimarte, L., and Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509–514. https://doi.org/10.1126/science.aaa1465

Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, *54*(2), 442–92. https://doi.org/http://dx.doi.org/10.1257/jel.54.2.442

Acquisti, A., and Varian, H. R. (2005). Conditioning prices on purchase history. *Marketing Science*, *24*(3), 305–523. https://doi.org/10.1287/mksc.1040.0103

Agrawal, A., Gans, J., and Goldfarb, A. (2019). *The economics of artificial intelligence: An agenda*. University of Chicago Press.

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., and Rieke, A. (2019). Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 1–30.

Ali, S. N., Lewis, G., and Vasserman, S. (2020). Voluntary disclosure and personalized pricing. *Proceedings of the 21st ACM Conference on Economics and Computation*, 537–538.

Ambrus, A., Calvano, E., and Reisinger, M. (2016). Either or both competition: A "two-sided" theory of advertising with overlapping viewerships. *American Economic Journal: Microeconomics*, *8*(3), 189–222. https://www.aeaweb.org/articles?id=10.1257/mic.20150019

Anderson, S. P. (2012). Foros, a˜., kind, h. *J., Peitz, M. " Media market concentration, advertising levels, and ad prices, International Journal of Industrial Organization*, *30*(3), 321–325. https://doi.org/10.1016/j.ijindorg.2011.11.003

Anderson, S. P. (2018). Foros, a˜., kind, h. *J. Competition for Advertisers and for Viewers in Media Markets". The Economic Journal*, *128*, 34–54. https://doi.org/10.1111/ecoj.12428

Anderson, S. P., Baik, A., and Larson, N. (2019). *Price discrimination in the information age: Prices, poaching, and privacy with personalised targeted discounts* (tech. rep.). CEPR Discussion Paper No. DP13793. Available at SSRN. https://ssrn.com/abstract=3428313

Anderson, S. P., and Coate, S. (2005). Market provision of broadcasting: A welfare analysis. *The Review of Economic Studies*, *72*(4), 947–972. http://www.jstor.org/stable/3700696

Anderson, S. P., and Peitz, M. (2020). *Ad clutter, time use, and media diversity* (tech. rep.). CEPR Discussion Paper No. DP15130, Available at SSRN. https://ssrn.com/abstract=3674907

Argenton, C., and Pr A¼fer, J. (2012). Search engine competition with network externalities. *Journal of Competition Law and Economics*, *8*(1), 73–105. https://doi.org/10.1093/joclec/nhr018

Argenziano, R., and Bonatti, A. (2021). *Data linkages and privacy regulation* [Working Paper].

Aridor, G., Che, Y.-K., Nelson, W., and Salz, T. (2020). *The economic consequences of data privacy regulation: Empirical evidence from gdpr* (Discussion Paper No. 26900). National Bureau of Economic Research.

Armstrong, M., and Vickers, J. (2001). Competitive price discrimination. *The RAND Journal of Economics*, *32*(4), 579–605. https://doi.org/10.2307/2696383

Armstrong, M., and Zhou, J. (2010). Conditioning prices on search behaviour. *MPRA Paper 19985 University Library of Munich, Germany.*

Arora, A., and Fosfuri, A. (2005). Pricing diagnostic information. *Management Science*, *51*(7), 1015–1064. https://doi.org/10.1287/mnsc.1050.0362

Athey, S., Calvano, E., and Gans, J. (2013). The impact of the internet on advertising markets for news media. *SSRN Electronic Journal.* 10.2139/ssrn.2180851

Athey, S., and Gans, J. S. (2010). The impact of targeting technology on advertising markets and media competition. *The American Economic Review*, *100*(2), 608–613. http://www.jstor.org/stable/27805067

Bajari, P., Chernozhukov, V., and Horta A§suA., J. (, Suzuki. (n.d.). The impact of big data on firm performance: An empirical investigation. *AEA Papers and Proceedings*, *109*, 33–37. https://www.aeaweb.org/articles?id=10.1257/pandp.20191000

Batikas, M., Bechtold, S., Kretschmer, T., and Peukert, C. (2020). *European privacy law and global markets for data* (tech. rep.). CEPR Discussion Paper No. DP14475. Available at SSRN. https://ssrn.com/abstract=3560282

Baye, M., and Sappington, D. (2020). Revealing transactions data to third parties: Implications of privacy regimes for welfare in online markets. *Journal of Economics Management Strategy*, *29*(2), 260–275. https://doi.org/10.1111/jems.12337

Belleflamme, P., Lam, W. M. W., and Vergote, W. (2020). Competitive imperfect price discrimination and market power. *Marketing Science*, *39*(5), 996–1015.

Belleflamme, P., and Vergote, W. (2016). Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters*, *149*. 10.1016/j.econlet.2016.10.027

Bergemann, D., and Bonatti, A. (2011). Targeting in advertising markets: Implications for offline versus online media. *The RAND Journal of Economics*, *42*(3), 417–443. http://www.jstor.org/stable/23046807

Bergemann, D., and Bonatti, A. (2015). Selling cookies. *American Economic Journal: Microeconomics*, *7*(3), 259–294. https://doi.org/http://dx.doi.org/10.1257/mic.20140155

Bergemann, D., and Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics*, *11*, 85–107.

Bergemann, D., Bonatti, A., and Gan, T. (2020). *The economics of social data* (tech. rep.). Cowles Foundation Discussion Paper No. 2203R, Available at SSRN. https://ssrn.com/abstract=3548336

Bergemann, D., Bonatti, A., and Smolin, A. (2018). The design and price of information. *American economic review*, *108*(1), 1–48. https://www.aeaweb.org/articles?id=10.1257/aer.20161079

Bester, H., and Petrakis, E. (1996). Coupons and oligopolistic price discrimination. *International Journal of Industrial Organization*, *14*(2), 227–242. https://doi.org/10.1016/0167-7187(94)00469-2

Bimpikis, K., Crapis, D., and Tahbaz-Salehi, A. (2019). Information sale and competition. *Management Science*, *65*(6), 2646–2664. https://doi.org/10.1287/mnsc.2018.3068

Binns, R., and Bietti, E. (2019). *Acquisitions in the third-party tracking industry: Competition and data protection aspects* [Preprint]. https://doi.org/10.31228/osf.io/fe8u7

Bonatti, A., and Cisternas, G. (2020). Consumer scores and price discrimination. *The Review of Economic Studies*, *87*(2), 750–791. https://doi.org/10.1093/restud/rdz046

Bounie, D., Dubus, A., and Waelbroeck, P. (2020). *Market for information and selling mechanisms* (tech. rep.). CESifo Working Paper No. 8307, Available at SSRN. https://ssrn.com/abstract=3618830

Bounie, D., Dubus, A., and Waelbroeck, P. (2021a). *Competition and mergers with strategic data intermediaries*. https://ssrn.com/abstract=3918829

Bounie, D., Dubus, A., and Waelbroeck, P. (2021b). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, *52*, 283–313. https://doi.org/10.1111/1756-2171.12369

Bourreau, M., Caffarra, C., Chen, Z., Choe, C., Crawford, G. S., and Duso, T. V. T. ((n.d.-a). *" google/fitbit will monetise health data and harm consumers"*. Centre for Economic Policy Research.

Bourreau, M., Hofmann, J., and Krämer, J. (n.d.-b). Prominence-for-data schemes in digital platform ecosystems: Economic implications for platform bias and consumer data collection. Available at SSRN 3867580.

Calzolari, G., and Pavan, A. (2006). On the optimality of privacy in sequential contracting. *Journal of Economic Theory*, *130*(1), 168–204. https://doi.org/10.1016/j.jet.2005.04.007

Campbell, J., Goldfarb, A., and Tucker, C. (2015). Privacy regulation and market structure. *Journal of Economics Management Strategy*, *24*, 47–73. https://doi.org/10.1111/jems.12079

Casadesus-Masanell, R., and Hervas-Drane, A. (2015). Competing with privacy. *Management Science*, *61*(1), 229–246.

Cespa, G. (2008). Information sales and insider trading with long-lived information. *The Journal of Finance*, *63*(2), 639–672. https://doi.org/10.1111/j.1540-6261.2008.01327.x

Chen, N., and Tsai, H. (2019). *Steering via algorithmic recommendations.* Available%20at%20SSRN:%20https://ssrn%22.com/abstract=3500407

Chen, Y. (1997). Paying customers to switch. *Journal of Economics and Management Strategy*, *6*(4), 877–897. https://doi.org/10.1111/j.1430-9134.1997.00877.x

Chen, Y., and Iyer, G. (2002). Consumer addressability and customized pricing. *Marketing Science*, *21*(2), 197–208. http://www.jstor.org/stable/1558067

Chen, Y., Narasimhan, C., and Zhang, Z. J. (2001). Individual marketing with imperfect targetability. *Marketing Science*, *20*(1), 23–41. https://doi.org/10.1287/mksc.20.1.23.10201

Chen, Y., and Zhang, Z. J. (2009). Dynamic targeted pricing with strategic consumers. *International Journal of Industrial Organization*, *27*(1), 43–50. https://doi.org/10.1016/j.ijindorg.2008.03.002

Chen, Z., Choe, C., Cong, J., and Matsushima, N. (2022). *Data-driven mergers and personalization".* The RAND Journal of Economics. https://doi.org/10.1111/1756-2171.12398

Chen, Z., Choe, C., and Matsushima, N. (2020). Competitive personalised pricing. *Management Science*, *66*(9), 4003–4023. https://doi.org/10.1287/mnsc.2019.3392

Chiou, L., and Tucker, C. (2017). *Search engines and data retention: Implications for privacy and antitrust* (tech. rep.). NBER Working Papers 23815, National Bureau of Economic Research, Inc.

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

Choi, J. P., Jeon, D. S., and Kim, B. C. (2019). Privacy and personal data collection with information externalities. *Journal of Public Economics*, *173*, 113–124.

Coase, R. H. (1972). Durability and monopoly. *The Journal of Law and Economics*, *15*(1), 143–149.

Colombo, S., Graziano, C., and Pignataro, A. (2021). *History-based price discrimination with imperfect information accuracy and asymmetric market shares* (tech. rep.). CESifo Working Paper No. 9049, Available at SSRN. https://ssrn.com/abstract=3837777

Commission, F. T. (n.d.). Data brokers: A call for transparency and accountability. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Conitzer, V., Taylor, C. R., and Wagman, L. (2012). Hide and seek: Costly consumer privacy in a market with repeat purchases. *Marketing Science*, *31*(2), 277–292. http://www.jstor.org/stable/41488023

De Corniere, A. (2016). Search advertising. *American Economic Journal: Microeconomics*, *8*(3), 156–88.

De Corniere, A., and De Nijs, R. (2016). Online advertising and privacy. *The RAND Journal of Economics*, *47*(1), 48–72.

De Corniere, A., and Taylor, G. (2020). *Data and competition: A general framework with applications to mergers, market structure, and privacy policy* (tech. rep.). CEPR Discussion Paper No. DP14446. Available at SSRN https://ssrn.com/abstract=3547379.

Delbono, F., Reggiani, C., and Sandrini, L. (2022). Strategic data sales to competing firms. *Available at SSRN*.

Dosis, A., and Sand-Zantman, W. (2019). *The ownership of data*. https://ssrn.com/abstract=3420680

Economides, N., and Lianos, I. (2021). *Restrictions on privacy and exploitation in the digital economy: A market failure perspective*. https://ssrn.com/abstract=3686785

Esteves, R. (2014). Price discrimination with private and imperfect information. *Scandinavian Journal of Economics*, *116*, 766–796. https://doi.org/10.1111/sjoe.12061

Esteves, R. B. (2009a). Customer poaching and advertising. *The Journal of Industrial Economics*, *57*(1), 112–146. http://www.jstor.org/stable/25483452

Esteves, R. B. (2009b). *A survey on the economics of behaviour-based price discrimination* (tech. rep.). NIPE Working Papers 5/2009, NIPE - Universidade do Minho.

Esteves, R. B. (2010). Pricing with customer recognition. *International Journal of Industrial Organization*, *28*(6), 669–681.

Esteves, R. B., and Vasconcelos, H. (2015). Price discrimination under customer recognition and mergers. *Journal of Economics Management Strategy*, *24*(3), 523–549. https://doi.org/10.1111/jems.12107

Evans, D. S. (2019). Attention platforms, the value of content, and public policy. *Review of Industrial Organization*, *54*, 775–792. https://doi.org/10.1007/s11151-019-09681-x

Evans, D. S. (2020). *The economics of attention markets*. https://ssrn.com/abstract=3044858

F., C. B., and Valletti, T. ((n.d.). Selling customer information to competing firms. *Economics Letters*, *149*, 10–14. https://doi.org/10.1016/j.econlet.2016.10.005

Fainmesser, I., Galeotti, A., and Momot, R. (2020). *Digital privacy* [Discussion paper, Johns Hopkins University].

Fudenberg, D., and Tirole, J. (1998). Upgrades, tradeins, and buybacks. *The RAND Journal of Economics*, *29*(2), 235–258. https://doi.org/10.2307/2555887

Fudenberg, D., and Tirole, J. (2000). Customer poaching and brand switching. *The RAND Journal of Economics*, *31*(4), 634–657. https://doi.org/10.2307/2696352

Fudenberg, D., and Villas-Boas, M. (2006). Behavior-based price discrimination and customer recognition. *Handbook on Economics and Information Systems*, *1*, 377–436.

Garcia, D., and Sangiorgi, F. (2011). Information sales and strategic trading. *The Review of Financial Studies*, *24*(9), 3069–3104. https://doi.org/10.1093/rfs/hhr041

Garratt, R. J., and Van Oordt, M. R. (2021). Privacy as a public good: A case for electronic cash. *Journal of Political Economy*, *129*(7), 2157–2180.

Garrett, J., Shriver, S., and Goldberg, S. (2022). *Privacy market concentration: Intended unintended consequences of the gdpr*. https://ssrn.com/abstract=3477686

Gehrig, T., and Stenbacka, R. (2007). Information sharing and lending market competition with switching costs and poaching. *European Economic Review*, *51*(1), 77–99. 10.1016/j.euroecorev.2006.01.009

Goldfarb, A., and Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3–43. https://doi.org/10.1257/jel.20171452

Gu, Y., Madio, L., and Reggiani, C. (2019). *Exclusive data, price manipulation and market leadership* (tech. rep.). CESifo Working Paper No. 7853. Available at SSRN. https://ssrn.com/abstract=3467988

Gu, Y., Madio, L., and Reggiani, C. (2022). Data brokers co-opetition. *Oxford Economic Papers*, *74*(3), 820–839.

Haan, M. A., Zwart, G., and Stoffers, N. (2021). *Choosing your battles: Endogenous multihoming and platform competition* (Discussion Paper No. DP2021-011, University of Groningen Faculty of Law Research Paper Forthcoming, Available at SSRN). TILEC. https://ssrn.com/abstract=3847216

Hagiu, A., and Jullien, B. (2011). Why do intermediaries divert search? *The RAND Journal of Economics*, *42*(2), 337–362. https://doi.org/10.1111/j.1756-2171.2011.00136.x

Hagiu, A., and Wright, J. (2020). *Data-enabled learning, network effects and competitive advantage* [Working Paper].

Hermalin, B. E., and Katz, M. L. (2006). Privacy, property rights and efficiency: The economics of privacy as secrecy. *Quantitative Marketing and Economics*, *4*, 209–239. https://doi.org/10.1007/s11129-005-9004-7

Hidir, S., and Vellodi, N. (2021). Privacy, personalization, and price discrimination. *Journal of the European Economic Association*, *19*(2), 1342–1363.

Hotelling, H. (1929). Stability in competion. *The Economic Journal*, *39*(153), 41–57.

Ichihashi, S. (2020). Online privacy and information disclosure by consumers. *American Economic Review*, *110*(2), 569–95. https://doi.org/10.1257/aer.20181052

Ichihashi, S. (2021). Competing data intermediaries. *The RAND Journal of Economics*, *52*, 515–537. https://doi.org/10.1111/1756-2171.12382

Ichihashi, S., and Kim, B. C. (2022). *Addictive platforms*. Management Science.

Iyer, G., and Soberman, D. (2000). Markets for product modification information. *Marketing Science*, *19*(3), 203–225. https://doi.org/10.1287/mksc.19.3.203.11801

Iyer, G., Soberman, D., and Villas-Boas, J. M. (2005). The targeting of advertising. *Marketing Science*, *24*(3), 461–476. https://doi.org/10.1287/mksc.1050.0117

Jeong, Y., and Maruyama, M. (2009). Commitment to a strategy of uniform pricing in a two-period duopoly with switching costs. *Journal of Economics*, *98*(1), 45–66. http://www.jstor.org/stable/41795638

Jing, B. (2011). Pricing experience goods: The effects of customer recognition and commitment. *Journal of Economics and Management Strategy*, *20*(2), 451–473. https://doi.org/10.1111/j.1530-9134.2011.00294.x

Johnson, J. P. (2013). Targeted advertising and advertising avoidance. *The RAND Journal of Economics*, *44*(1), 128–144. http://www.jstor.org/stable/43186411

Jones, C. I., and Tonetti, C. (2020). Nonrivalry and the economics of data. *American Economic Review*, *110*(9), 2819–2858.

Jullien, B., Lefouili, Y., and Riordan, M. H. (2020). Privacy protection, security, and consumer retention. *Available at SSRN:* https://doi.org/http://dx.doi.org/10.2139/ssrn.3655040

Kastl, J., Pagnozzi, M., and Piccolo, S. (2018). Selling information to competitive firms. *The RAND Journal of Economics*, *49*(1), 254–282. https://www.jstor.org/stable/45147433

Kim, B. C., and Choi, J. P. (2010). Customer information sharing: Strategic incentives and new implications. *Journal of Economics Management Strategy*, *19*, 403–433. https://doi.org/10.1111/j.1530-9134.2010.00256.x

Kim, J. H., Wagman, L., and Wickelgren, A. L. (2019). The impact of access to consumer data on the competitive effects of horizontal mergers and exclusive dealing. *Journal of Economics Management Strategy*, *28*(3), 373–391.

Kirpalani, R., and Philippon, T. (2020). *Data sharing and market power with two-sided platforms* (tech. rep.). 28023 NBER working paper available at. https://www.nber.org/papers/w28023

Kramer, J., Schnurr, D., and Wohlfarth, M. (2019). Winners, losers, and facebook: The role of social logins in the online advertising ecosystem. *Management Science*, *65*(4), 1678–1699. https://doi.org/10.1287/mnsc.2017.3012

Krämer, J. (2021). Personal data portability in the platform economy: Economic implications and policy recommendations. *Journal of Competition Law & Economics*, *17*(2), 263–308.

Krämer, J., and Stüdlein, N. (2019). Data portability, data disclosure and data-induced switching costs: Some unintended consequences of the general data protection regulation. *Economics Letters*, *181*, 99–103.

Lam, W. M. W., and Liu, X. (2020). Does data portability facilitate entry? *International Journal of Industrial Organization*, *69*, 102564.

Lancieri, F., and Sakowski, P. M. (2021). Competition in digital markets: A review of expert reports. *Stanford Journal of Law, Business Finance*, *26*(1), 65–170. https://doi.org/http://dx.doi.org/10.2139/ssrn.3681322

Lefouili, Y., and Toh, Y. L. (2017). *Privacy and quality* (tech. rep.). TSE Working Paper, 17 (795).

Li, X., Li, K. J., and Wang, X. (2020). Transparency of behavior-based pricing. *Journal of Marketing Research*, *57*(1), 78–99.

Liu, Q., and Serfes, K. (2004). Quality of information and oligopolistic price discrimination. *Journal of Economics Management Strategy*, *13*, 671–702. https://doi.org/10.1111/j.1430-9134.2004.00028.x

Liu, Q., and Serfes, K. (2006). Customer information sharing among rival firms. *European Economic Review*, *50*(6), 1571–1600. https://doi.org/10.1016/j.euroecorev.2005.03.004

Markovich, S., and Yehezkel, Y. (2021). *Data regulation: Who should control our data?* [Available at SSRN:]. https://ssrn%22.com/abstract=3801314

Mauring, E. (2021). *Search and price discrimination online* (tech. rep.). CEPR Discussion Paper No. DP15729. Available at SSRN. https://ssrn.com/abstract=3783955

Montes, R., Sand-Zantman, W., and Valletti, T. (2019). The value of personal information in online markets with endogenous privacy. *Management Science*, *65*(3), 955–1453. https://doi.org/10.1287/mnsc.2017.2989

Neumann, N., Tucker, C., and Whitfield, T. (2019). How effective is third-party consumer profiling and audience delivery? evidence from field studies. *Marketing Science*, *38*(6), 918–926. https://doi.org/10.1287/mksc.2019.1188

Newman, J. (2020). *Antitrust in attention markets: Definition, power, harm* (tech. rep.). University of Miami Legal Studies Research Paper No. 3745839, Available at. http://dx.doi.org/10.2139/ssrn.3745839

Nilssen, T. (1992). Two kinds of consumer switching costs. *RAND Journal of Economics*, *23*, 579–589.

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

Pigou, A. (1920). *The economics of welfare*. London: Macmillan.

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

Prüfer, J., and Schottmüller, C. (2021). Competing with big data. *The Journal of Industrial Economics*, *69*(4), 967–1008.

Regibeau, P. (2021). *Why i agree with the google-fitbit decision*. Available at. https://voxeu.org/article/why-i-agree-google-fitbitdecision

Roy, S. (2000). Strategic segmentation of a market. *International Journal of Industrial Organization*, *18*(8), 1279–1290. https://doi.org/http://dx.doi.org/10.1016/S0167-7187(98)00052-6

Rutt, J. (2012). Targeted advertising and media market competition. *SSRN Electronic Journal*. 10.2139/ssrn.2103061

Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics*, *10*(1), 141–156.

Sarvary, M., and Parker, P. M. (1997). Marketing information: A competitive analysis. *Marketing science*, *16*(1), 24–38. https://doi.org/10.1287/mksc.16.1.24

Schäfer, M., and Sapi, G. (2020). *Learning from data and network effects: The example of internet search* (Discussion Paper No. 1894. Available at SSRN). DIW Berlin. https://ssrn.com/abstract=3688819

Schmittlein, D. C., Cooper, L. G., and Morrison, D. G. (1993). Truth in concentration in the land of (80/20) laws. *Marketing Science*, *12*(2), 167–183.

Schneble, C. O., Elger, B. S., and Shaw, D. (2018). The cambridge analytica affair and internet-mediated research. *EMBO reports*, *19*(8).

Shaffer, G., and Zhang, Z. (2000). Pay to switch or pay to stay: Preference-based price discrimination in markets with switching costs. *Journal of Economics and Management Strategy*, *9*(3), 397–424. https://doi.org/10.1111/j.1430-9134.2000.00397.x

Shaffer, G., and Zhang, Z. J. (1995). Competitive coupon targeting. *Marketing Science*, *14*(4), 395–416. http://www.jstor.org/stable/184137

Shaffer, G., and Zhang, Z. J. (2002). Competitive one-to-one promotions. *Management Science*, *48*(9), 1143–1160. http://www.jstor.org/stable/822606

Shin, J., and Sudhir, K. (2010). A customer management dilemma: When is it profitable to reward one's own customers? *Marketing Science*, *29*(4), 671–689.

Shin, J., Sudhir, K., and Yoon, D. H. (2012). When to â€œfireâ€ customers: Customer cost-based pricing. *Management Science*, *58*(5), 932–947.

Shy, O., and Stenbacka, R. (2013). Investment in customer recognition and information exchange. *Information Economics and Policy*, *25*(2), 92–106. https://doi.org/10.1016/j.infoecopol.2013.03.002

Shy, O., and Stenbacka, R. (2016). Customer privacy and competition. *Journal of Economics Management Strategy*, *25*(3), 539–562. https://doi.org/10.1111/jems.12157

Summers, M. (2020). Facebook isn't free: Zero-price companies overcharge consumers with data. *Behavioural Public Policy*, 1–25. https://doi.org/10.1017/bpp.2020.47

Taylor, C. R. (2003). Supplier surfing: Competition and consumer behavior in subscription markets. *The RAND Journal of Economics*, *34*(2), 223–246. https://doi.org/10.2307/1593715

Taylor, C. R. (2004). Consumer privacy and the market for customer information. *The RAND Journal of Economics*, *35*(4), 631–650. https://doi.org/10.2307/1593765

Taylor, C. R., and Wagman, L. (2014). Consumer privacy in oligopolistic markets: Winners, losers, and welfare. *International Journal of Industrial Organization*, *34*, 80–84. https://doi.org/10.1016/j.ijindorg.2014.02.010

Thisse, J.-F., and Vives, X. (1988). On the strategic choice of spatial price policy. *The American Economic Review*, *78*(1), 122–137. http://www.jstor.org/stable/1814702

Tirole, J. (1988). *The theory of industrial organization*. MIT press.

Vickrey, W. S. (1964). *Microstatics*. Harcourt.

Villas-Boas, J. M. (1999). Dynamic competition with customer recognition. *The RAND Journal of Economics*, *30*(4), 604–631. https://doi.org/10.2307/2556067

Villas-Boas, J. M. (2004). Price cycles in markets with customer recognition. *The RAND Journal of Economics*, *35*(3), 486–501. https://doi.org/10.2307/1593704

Wernerfelt, B. (1994). On the function of sales assistance. *Marketing Science*, *13*(1), 68–82.

Wu, T. (2017). *Blind spot: The attention economy and the law*. Law Society: Public Law - Antitrust eJournal. Available at. https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=3030&context=faculty_scholarship

Xiang, Y., and Sarvary, M. (2013). Buying and selling information under competition. *Quantitative Marketing and Economics*, *11*(3), 321–351. https://doi.org/10.1007/s11129-013-9135-1

Zhang, J. (2011). The perils of behavior-based personalisation. *Marketing Science*, *30*(1), 170–186. https://doi.org/10.1287/mksc.1100.0607

Zogheib, J.-M., Bourreau, M., et al. (2021). *Privacy, competition, and multi-homing* (tech. rep.). University of Paris Nanterre, EconomiX.

CHAPTER 2

User Data and Endogenous Entry in Online Markets*

Laura Abrardi, Carlo Cambini, 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 defines the price and the partition of data sold to each firm. We show that the DB reduces firm entry and sells data to all entering firms. This reduces both downstream competition and consumer surplus. Results are robust to alternative selling mechanisms entailing different degrees of DB's bargaining power, and to the possibility by the DB to commit to the price of data before firm entry.

* A version of this chapter has been made available online, and can be retrieved at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4256544.

I 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 and feedbacks.

‡Politecnico di Torino, Department of Management, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy.

1. Introduction

Invisible hands move the market of data, but they might not be those of competition. 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 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 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., 2021; 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 (FTC, 2014; Crain, 2018). DBs' business model compounds both online and offline sources, collecting data from commercial, government, and other publicly available sources – e.g., blogs and 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: 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 that want to reach specific consumers with targeted offers.

[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. (2021) show that the algorithms used by the leading online grocers in the U.S. personalize prices at the delivery zip code level.

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., MontesEtAl2019) has pointed out how DBs have the incentive to increase some firms' market power by selling data selectively in downstream duopolistic markets. However, little is known about 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 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 BergemannBonatti2019)[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 TaylorWagman2014). Intuitively, the DB charges a price of data such that he extracts most of the firms' profits, thus reducing entry into 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 the whole dataset to each firm, allowing

[3]The use of direct sales when selling data has been documented by the United States subcommittee on antitrust (Judiciary Committee, 2020).

them to identify all 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.

We then extend the baseline model in two ways. First, we consider alternative selling mechanisms adopted by the DB, based on auctions with or without reserve prices (see, e.g., BounieEtAl2021a). We find that the entry barrier effect survives these alternative selling mechanisms. However, differently from the Take-It-Or-Leave-It offers of our baseline model, the DB prefers to sell data only to a subset of entering firms under the auction mechanisms. This is due to the fact that the possibility to compete having information precluded to rivals increases the firms' willingness to pay for data and, therefore, the DB's profits. We also find that the auction mechanisms entail a lower consumer surplus than Take-It-Or-Leave-It offers.

Second, we explore the situation in which the data sale occurs before the firms' entry. This is, for instance, the case of emerging digital markets, where potential entrants anticipate the value of consumer data and thus make their entry decision after having obtained (or not obtained) data. Under this alternative timing, entry is further reduced, leading to higher profits and lower consumer surplus compared to the baseline model.

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 (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 firms exogenously have data, the literature highlights a pro-competitive effect under competition.[4] As informed firms compete more fiercely, consumers benefit from lower prices. Although Taylor and Wagman (2014) show that the pro-competitive effect of data increases consumer surplus and limits firm entry, this is due to the erosion of profits stemming from the intense competition and not from the intervention of a DB who maximizes his profits. A more recent strand of literature has endogenized the information acquisition process either through firms' repeated interactions with consumers (Villas-Boas, 2004; Acquisti and Varian, 2005; Liu and Serfes, 2004; Bergemann and Bonatti, 2011; Hagiu and Wright, 2020) or by acquiring data from strategic actors (Bergemann and Bonatti, 2015; de Cornière, 2016; Gu et al., 2019; Choe et al., 2023). In particular, Braulin and Valletti (2016), Montes et al. (2019) and Bounie et al. (2021b) 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 effects 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 non-rivalrous 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 (Greenhut and Ohta, 1976; DeGraba, 1990; Tyagi, 1999; Cachon and Lariviere, 2005). 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 non-essential input, as firms can enter the market even when not purchasing data. Moreover, data acquisition only influences the firms' efficiency in extracting surplus from consumers.

We contribute to the existing literature in three ways. First, we extend the duopolistic setup to analyze how the number of competing firms in an oligopoly market influences

[4]See for instance Thisse and Vives (1988), Shaffer and Zhang (1995), Bester and Petrakis (1996), C. R. Taylor (2003), Liu and Serfes (2004, 2005), Taylor and Wagman (2014), Shy and Stenbacka (2016) and Chen et al. (2020).

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. Third, we study how different data selling mechanisms affect entry and market outcomes. We show that selling data through TIOLI offers tempers consumer harm, as opposed to auction-based selling mechanisms. This is due to the fact that, under TIOLI offers, data are sold to all firms, leading to higher competition.

From a policy perspective, a critical concern pertains to the concentration of the DBs' market and its effects on consumers. 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 2 presents the model, and Section 3 analyzes firms' equilibrium prices. Section 4 computes the DB's profits and his optimal strategy and discusses the consequent market outcomes. Section 5 extends the model along two main directions: by increasing the DB's bargaining power through the auction selling mechanisms and by allowing him to commit to data prices prior to firm entry. Section 6 concludes. All proofs and technical details are contained in the Appendix.

2. 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.

2.1. Consumers, firms and the Data Broker

We consider a free-entry game of a market represented by a circular city of length 1 (Vickrey, 1964; Salop, 1979), where firms sell competing products to consumers. Firms are indexed by $i \in \{0, 1, 2, \ldots, n-2, n-1\}$, where n 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. To maximize the value of data for firms – and thus, their willingness to pay – the partition sold to each firm contains its location, as in Bounie et al. (2021b). In particular, due to the symmetry of the market, the partition sold by each firm is centered on the firm's location. We denote with $d_i \in [0, 1]$ the partition of data offered to firm i. Consumers belonging to the arch d_i are identified by firm i, i.e., the firm knows their location and can perform on them first-degree price discrimination. The partition set containing the partitions offered by the DB to all firms is $\mathbf{P} = (d_0, d_1, d_2, \ldots, d_{n-1})$.

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 is offered the price $p_i(x) \in \{p_i^{\mathrm{T}}(x), p_i^{\mathrm{B}}\}$ from firm i and can either accept or reject it.

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

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

[5]As standard in the literature on markets with entry, we assume sequential entry to avoid coordination problems and ignore integer constraints on n. A similar approach has been recently adopted in Rhodes and Zhou (2021).

[6]We can think of F as the cost incurred in the process of digitization (see AndersonBedre-Defolie2021), such as the creation of an online retail shop.

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 $\widehat{x}_{i,i+1}$, such that $U(\widehat{x}_{i,i+1}, i) = U(\widehat{x}_{i,i+1}, i+1)$.

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 its 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 $\widehat{x}_{i,i+1} - \widehat{x}_{i-1,i} - d_i$ will receive its basic price. Then, its profits are equal to

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

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

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

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

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

2.2. The selling mechanism and timing

The DB sells data through simultaneous Take-It-Or-Leave-It (TIOLI) offers (Bergemann et al., 2018; Bounie et al., 2022).[7] 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:

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

[7]In Section 5.1 we extend the analysis to the case of a selling mechanism in which the DB can change the offer made to a firm on the basis of another firm's response, as in BounieEtAl2021a.

[8]By assuming public DB's offers, we rule out situations like secret contracting games as in Hart and Tirole (1988). In our model, firms are ex-ante identical and the DB's decision to sell data to any specific firm does not depend on the firm's identity, but possibly only on its location. See also Bounie et al. (2021b), Montes et al. (2019) and 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 $\mathbf{w} = (w_0, w_1, ..., w_{n-1})$. $\mathbf{P}$ 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 $\widetilde{\pi_i} = 0$. Consumer surplus is $\widetilde{CS} = v - \frac{5}{4}\sqrt{tF}$, which is also equal to total surplus (Taylor and Wagman, 2014).

3. Equilibrium prices

We rely on the equilibrium concept of Perfect Nash Equilibrium (PNE) and proceed by backward induction, starting by firms' pricing stage. While tailored prices are offered to the identified consumers, who are closest to the firm's location, farther consumers are instead 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]

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

If firm i buys data, it obtains the partition d_i, observes the competitors' basic prices and offers a tailored price $p_i^{\mathrm{T}}(x)$ to the identified consumers for each arch on which it

[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. (2021b) for an analogous approach.

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

competes. The tailored price matches the competitor's basic price in the utility level:

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

Note that tailored prices decrease as the rival's basic price decreases due to downstream competition. If the partition d_i is sufficiently large to allow firm i to identify all consumers it serves, its profits when it buys data, prior to paying for them, are given by (2). Using (6) and (7), we obtain:

$$\pi_i^{\mathrm{W}}(\mathbf{P}) = \frac{t}{2n^2} + \frac{(p_{i+1}^B + p_{i-1}^B)}{2n} + \frac{(p_{i+1}^B)^2 + (p_{i-1}^B)^2 - 2(p_i^B)^2}{4t} - F, \tag{8}$$

From (8), firm i's basic price negatively affects its profits, as it decreases its market share. Then, firms have the incentive to set their basic prices to the lowest possible level, i.e., zero, implying that profits in (8) do not depend on d_i. Indeed, once the data partition allows firm i to identify all consumers up to the indifferent one, any additional data do not allow to gain new consumers and have therefore no effect on profits.

Conversely, if the partition d_i allows firm i to serve both identified and unidentified consumers, 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

$$\pi_i^{\mathrm{W}}(\mathbf{P}) = \frac{d_i}{2n}\left(2t + np_{i-1}^{\mathrm{B}} + np_{i+1}^{\mathrm{B}} - ntd_i\right) + p_i^{\mathrm{B}}\left(\frac{n\left(p_{i+1}^{\mathrm{B}} + p_{i-1}^{\mathrm{B}} - 2p_i^{\mathrm{B}}\right) + 2t}{2nt} - d_i\right) - F, \tag{9}$$

where the first component represents the profit on the identified segment, whereas the second component represents the profit on the unidentified segment. Equation (9) highlights that the amount of data d_i has a direct effect on firm i's strategy by increasing the share of identified consumers. Furthermore, as we will see more in detail later on, the partition set $\mathbf{P}$ indirectly affects firm i's profit in (9) via the basic prices.

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

$$\pi_i^{\mathrm{L}}(\mathbf{P}) = p_i^{\mathrm{B}}\left(\frac{n\left(p_{i+1}^{\mathrm{B}} + p_{i-1}^{\mathrm{B}} - 2p_i^{\mathrm{B}}\right) + 2t}{2nt}\right) - F. \tag{10}$$

In this case, data have no direct effect on (10), which is affected only indirectly through the basic prices.

The first-order condition of (9) and (10) with respect to $p_i^{\mathrm{B}}(\mathbf{P})$ are, respectively:

$$p^{\mathrm{B}}_{i(\mathrm{W})} = \frac{t}{2n} - \frac{td_i}{2} + \frac{p^{\mathrm{B}}_{i+1} + p^{\mathrm{B}}_{i-1}}{4} \tag{11}$$

$$p^{\mathrm{B}}_{i(\mathrm{L})} = \frac{t}{2n} + \frac{p^{\mathrm{B}}_{i+1} + p^{\mathrm{B}}_{i-1}}{4}. \tag{12}$$

Equation (12) is the reaction function of the standard Salop (1979) model, whereas Equation (11) differs from it for the term $-\frac{td_i}{2}$.

From Equations (9) and (11), we find that data have two opposite effects on firms' profits. On the one hand, data allow firms to identify consumers and charge them with a tailored price, which 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 (9). 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, as shown in Equation (11). A lower basic price reduces firm profits by the second term in Equation (9), which constitutes the *competition effect* of data (Thisse and Vives, 1988).

Let us focus first on the subgame where all firms buy their respective partitions. The system of reaction functions 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, we have that:*
(i) $\frac{\partial p_i^{B*}}{\partial d_j} < 0$, $\forall i, j$ *(firm i's subgame equilibrium basic price is decreasing in* d_j*),*
(ii) $\frac{\partial \pi_i^{W*}(\mathbf{P})}{\partial d_j} < 0$, $\forall j \neq i$ *(firm i's subgame equilibrium profit is decreasing in* d_j*),*
(iii) There exists a threshold $\bar{d}_i \in [0, 1)$ *such that* $\frac{\partial \pi_i^{W*}(\mathbf{P})}{\partial d_i} > 0$ *if* $d_i < \bar{d}_i$*, and* $\frac{\partial \pi_i^{W*}(\mathbf{P})}{\partial d_i} \leq 0$ *otherwise,* $\forall i$.

Proof. See Appendix A. □

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 $d_i < \bar{d}_i$, 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 $d_i \geq \bar{d}_i$.

We now focus on the subgame where firm i does not buy data, and thus obtains profits as in Equation (10). The main properties of firms' equilibrium prices and profits in this subgame are described in the following lemma.

LEMMA 2. *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 Appendix A. □

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.

4. 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 $\mathbf{P}$. 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.

4.1. 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 the firm's 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 $\pi_i^{\mathrm{L}*}(\mathbf{P})$. Thus, the DB solves the following problem:

$$\max_{d_0,d_1,\ldots,d_{n-1}} \pi_{DB}(\mathbf{P}) = \sum_{i=0}^{n-1} \left(\pi_i^{\mathrm{W}*}(\mathbf{P}) - \pi_i^{\mathrm{L}*}(\mathbf{P})\right). \tag{13}$$

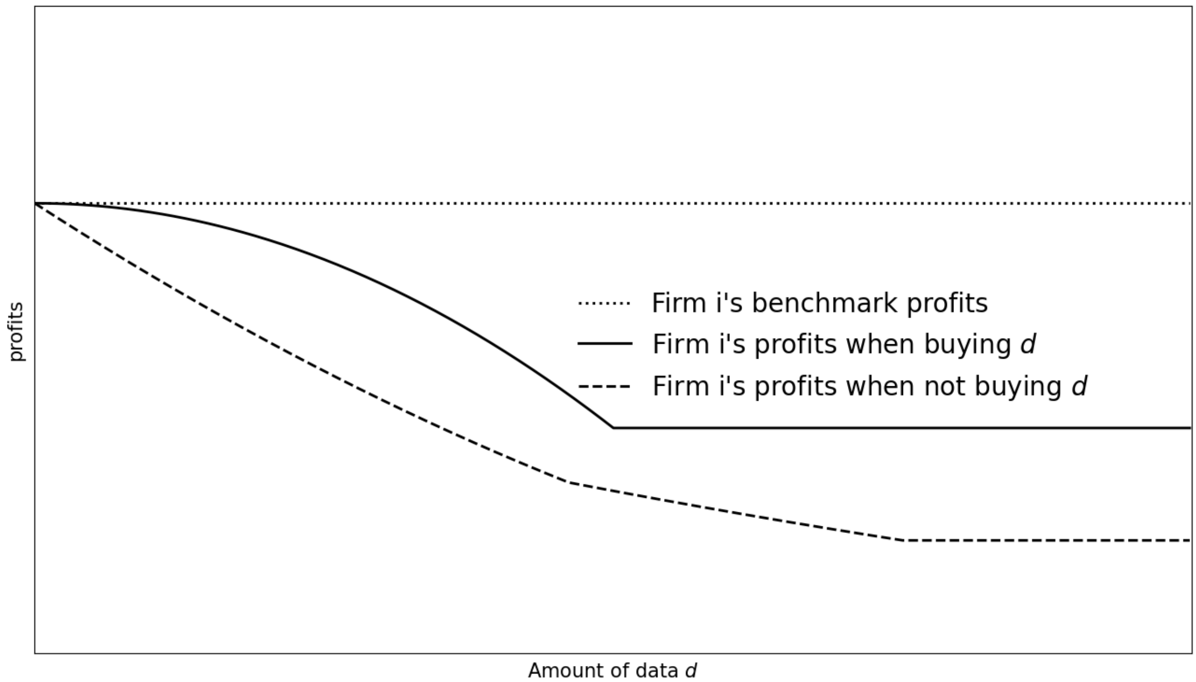
The following lemma illustrates the properties of the optimal partition chosen by the DB and its effects on firms' profits.

LEMMA 3. *In equilibrium, the DB offers same-sized partitions to all firms, i.e., $d_i = d$, for all i. Firms' profits when buying data are decreasing in d for $d \leq \frac{1}{n}$, and are constant otherwise. Firms' profits when not buying data are decreasing in d for $d \leq \frac{3}{2n}$ and are constant otherwise.*

PROOF. See Appendix A. □

Intuitively, due to the firms' symmetry, DB's profits are influenced by all the partitions he sells in the same way, and thus he offers symmetric partitions. Figure 1 provides a graphical representation of firms' profits functions with respect to d in the two subgames of buying and not buying data. Due to the competitive advantage provided by data, firms are always better off when buying data relative to not buying them.

FIGURE 1. Firms' profits with respect to d ($n = 4$ and $t = 20$)



When all firms have data, they set basic prices more aggressively in an attempt to expand their market share. The resulting competition effect is intensified by the symmetry between firms, thereby reducing profits. Once firms identify all consumers in their market segment, i.e., $d \geq \frac{1}{n}$, additional data allow firms to identify consumers that cannot be poached and, therefore, have no effect on either the basic pricing strategy or profits.

Conversely, if firm i does not purchase data, it will be at a disadvantage vis-à-vis its rivals, for two reasons. First, firm i cannot extract surplus from consumers via tailored prices. Second, firm i can only use the basic price for surplus extraction, whereas its competitors can also use the tailored price. Therefore, firm i sets a higher basic price than its informed rivals, thus losing market shares.[11] Since the uninformed firm's market share is lower due to its competitive disadvantage, its rivals need a larger partition ($d \geq \frac{3}{2n}$) to identify the indifferent consumer. Above this threshold, data have no additional effect on firm i's profits as they allow its rivals to identify consumers they cannot poach.

From Figure 1, we can clearly see how an uninformed firm's profits are always positive (prior to paying for entry), even when its rivals obtain the whole dataset. The reason is that, as firms are symmetric, the informed rivals are not able to overcome the positional advantage that the uninformed firm has over the consumers who are close to its location. In other words, we find that if firms are symmetric, perfect price discrimination is not enough to overcome the localized spatial competition that is typical of the Salop model, as informed firms are not able to poach consumers who are located after their rivals' locations.

4.2. Equilibrium data partitions

Given that the DB offers same-sized partitions in equilibrium from Lemma (3), we can rewrite his profit-maximization problem in (13) as

$$\max_{d} \pi_{DB}(\mathbf{P}) = n\left(\pi_i^{\mathrm{W}*}(\mathbf{P}) - \pi_i^{\mathrm{L}*}(\mathbf{P})\right) \tag{14}$$

The solution to the problem (14) is the optimal amount of data d^*, and it depends on the number of entering firms. We thus find the following proposition.

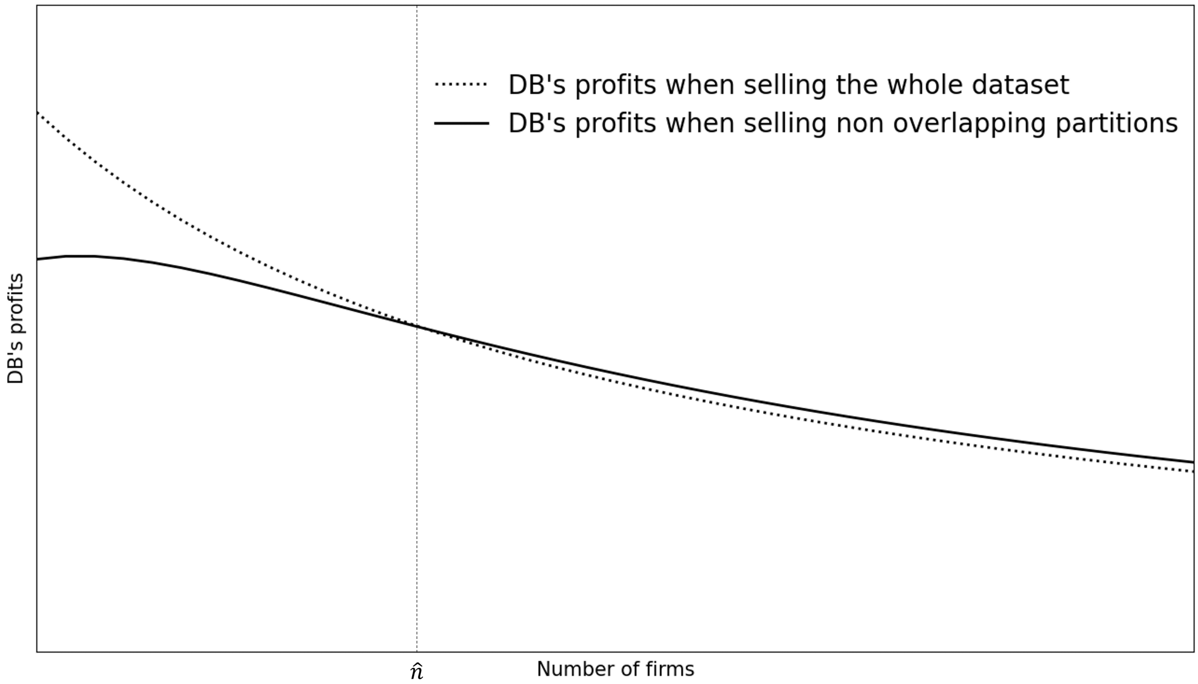
PROPOSITION 1. *There exists a number $\hat{n}$ of firms such that in equilibrium the DB offers $d^* = 1$ if $n < \hat{n}$, and non-overlapping partitions otherwise.*

PROOF. See Appendix A. □

Figure 2 illustrates the result in Proposition 1 by showing the DB's profits under the two strategies (selling the whole dataset or non-overlapping partitions) with respect to the number of firms in the market.

[11] The rivals' market shares are higher the closer they are to the uninformed firm. Then, the uninformed firm's profit function depends on whether its closest rivals serve unidentified consumers on both market segments (for d sufficiently low), on just one (for intermediate levels of d), or only identified consumers (for d sufficiently high). See Appendix A for details on the threshold levels.

FIGURE 2. DB's profits as a function of the number of entering firms ($t = 20$)



The intuition of the result in Proposition 1 lies in how data affect firms' profits and, in turn, the DB's profits. On the one hand, selling more data increases competition, reducing profits and their willingness to pay for data. On the other hand, selling more data increases the threat firms face when not buying data. This second effect increases firms' willingness to pay for data. Both effects depend on the *total* amount of data that are sold in the market. Moreover, these effects become constant once firms obtain sufficient data to identify all consumers they serve. In this case, additional data have no effect on the firms' and the DB's profits, and the effects reach their maximums. However, the number of firms influences how fast the second effect reaches its maximum.

Suppose that a low number of firms enter the market. If a firm chooses not to buy data, the total amount of data in the market greatly reduces, which in turn decreases the threat faced by the firm. A firm's profits when not buying data would thus be slowly decreasing with respect to the amount of data.

Conversely, suppose that a high number of firms enter the market. In this scenario, if a firm chooses not to buy data, the total amount of data in the market slightly reduces, and the non-buying firm still faces a considerable threat. The non-buying firm's profits thus quickly decrease with respect to data. Then, the DB opts to sell small quantities of data in the downstream market, as this strategy allows him to temper the first effect while ensuring that firms face a considerable threat when not buying.

If n is sufficiently low, the DB maximizes his profits by selling the whole dataset to all entering firms. Intuitively, when the downstream market is highly concentrated, the DB can extract more surplus from firms by maximizing the threat they face when not buying data. Conversely, when the number of firms is sufficiently high, the DB seeks to temper downstream competition, which is already high, by selling non-overlapping partitions.

The previous literature (Bounie et al., 2022) has shown that under TIOLI offers and assuming non-overlapping partitions, a DB sells data to both firms in a downstream duopoly. By allowing data partitions to overlap, we find that, while the DB still sells data to all firms in equilibrium, the amount of data he sells depends on the downstream market concentration, which in turn has implications on the number of entering firms and welfare.

4.3. 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.

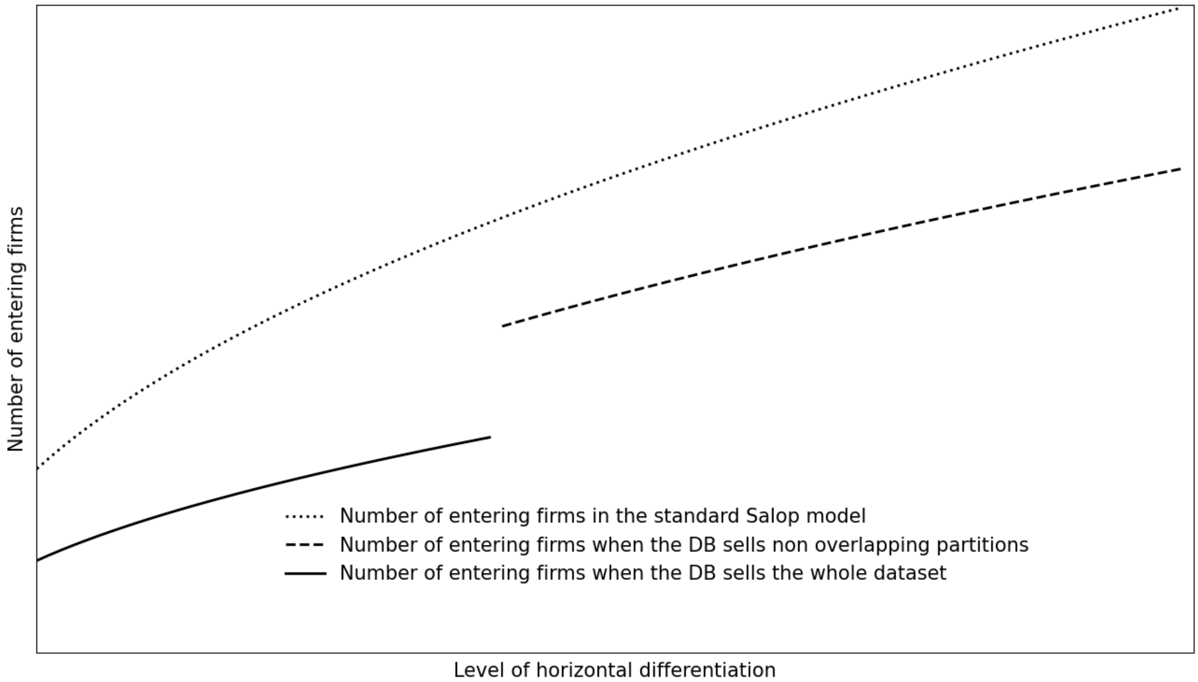
As a benchmark, recall that the number of entering firms in the standard Salop (1979) model (see Section 2.2), absent the DB, is $\widetilde{n} = \sqrt{\frac{t}{F}}$. In our model, the number of entering firms n^* is obtained through the free-entry condition, such that firms' profits after paying for 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 $t \leq \hat{t}$, and $n^* \approx \frac{3}{4}\sqrt{\frac{t}{F}}$ otherwise.*

PROOF. See Appendix A. □

Figure 3 illustrates the number of entering firms with respect to the level of horizontal differentiation in the downstream market.

Figure 3. Number of entering firms with respect to the level of horizontal differentiation t ($F = 1$)



The DB reduces firms' entry relative to the benchmark case where data are absent. Intuitively, such *entry barrier effect* arises because firms' profits are lower for two reasons. First, data increase the intensity of competition, leading firms to price more aggressively. Second, the need to pay for data further erodes firms' profits and leaves less room for entry. When $t \leq \hat{t}$, differentiation is low, and thus a small number of firms enter the market. Then, as derived from Proposition 1, the DB opts to sell the whole dataset, which further reduces firms' profits and entry as the competition effect of data is maximized. Conversely, if $t > \hat{t}$, competition is softer, leading to a large number of entering firms and incentivizing the DB to sell non-overlapping partitions, which in turn tempers the competition effect of data. The change in firms' profits when not buying data resulting from the change in the DB's strategy leads to the discontinuity in the number of entering firms that we observe in Figure 3. Interestingly, the number of entering firms in equilibrium cannot be lower than $\frac{\widetilde{n}}{2}$, i.e., the entry deterrence caused by data is limited by firm horizontal differentiation. Indeed, even when an uninformed firm faces perfectly informed rivals, it is still able to maintain a market share and obtain positive profits. 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.

Proposition 2 also highlights that entry depends on the degree of horizontal differentiation, with a discontinuity given by the DB changing his selling strategy, at some threshold level. When horizontal differentiation is low, competition is fierce. Fewer firms enter the market, leading the DB to sell the whole dataset (from Proposition 1). The sale of the whole dataset further intensifies competition, reducing firms' profits and, in turn, entry at $n^* = \frac{1}{2}\sqrt{\frac{t}{F}}$. Conversely, for high levels of horizontal differentiation, profits are higher so that more firms enter the market, and the DB sells non-overlapping partitions.

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:

$$CS_i = \int_{\hat{x}_{i-1,i}}^{\hat{x}_{i,i+1}} U(x,i)dx. \tag{15}$$

The total consumer surplus is $CS = \sum_i CS_i$. In particular, consumer surplus can be expressed as[12]

$$CS = \begin{cases} u - \frac{5t}{4n} + \frac{ntd^2}{2} & \text{for } \ d \in [0, \frac{1}{n}) \\ u - \frac{5t}{4n} + \frac{t}{2n} & \text{for } \ d \in [\frac{1}{n}, 1]. \end{cases} \tag{16}$$

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

$$TW = CS + \sum_{i=0}^{n-1} \pi_i + \alpha \pi_{\text{DB}}, \tag{17}$$

where $\alpha \in [0, 1]$ is the weight of DB's profits in the welfare function. The following proposition summarizes the impact of the DB's equilibrium strategy on consumer surplus and welfare.

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

Proof. See 6. □

Proposition 3 compares the equilibrium levels of consumer surplus and welfare in our model to those of the standard Salop model in which data are not available. We find that the entry barrier effect that arises because of the DB's presence lowers consumer

[12] As already shown in Lemma 3, all firms buy data in equilibrium and additional data have no effect when $d \geq \frac{1}{n}$. Thus, consumer surplus is constant for $d \geq \frac{1}{n}$.

surplus, regardless of the level of horizontal differentiation in the downstream market. To see why, consider Equation (16). The first two terms are the consumer surplus in the standard Salop model, and are positively related to the number of firms. The third term represents the effect of data on consumer surplus. A higher quantity of data intensifies competition between the entered firms and lowers basic prices, raising the surplus of (unidentified) consumers. The previous literature, focusing on an exogenous number of firms, has highlighted the positive effect of data on consumer surplus stemming from the fact that the competition effect of data outweighs the surplus extraction effect from a welfare perspective(Braulin and Valletti, 2016; Montes et al., 2019; Bounie et al., 2021b). However, our results highlight 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.[13] In particular, consumers are better off when the DB sells non-overlapping partitions, as this strategy entails a lower entry barrier effect, but the reduction in entry is still strong enough to cause consumer harm.

Proposition 3 also shows that if the weight α of the DB's profit in the welfare function is high enough, total welfare is higher than in the benchmark case, although it is mostly appropriated by the DB.[14] The DB's equilibrium strategy, either partially (when he sells non-overlapping partitions) or totally (when he sells the whole dataset), solves the excessive entry problem identified by Salop (1979). In this way, industry profits increase and can be subsequently extracted by the DB. This result shows how the increase in welfare is mainly driven by the increase in the DB's profits, causing redistributive concerns from a policymaking point of view.

5. Extensions

In this section, we extend our baseline model on two dimensions. First, we introduce variations in the DB's bargaining power by exploring different selling mechanisms for the data sale. Second, we analyze a scenario where the DB can commit to the price of data before firms' entry decision. This alternative timing allows the DB to take into account the effect of data on the number of entering firms when choosing his strategy, and provides him with more bargaining power than in the baseline model.

[13]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 discriminating through data increases consumer surplus by weakening the entry barrier effect (see, e.g., Taylor and Wagman (2014)).

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

5.1. Alternative selling mechanisms

In the baseline model, we assume that data are sold through Take-It-Or-Leave-It offers (TIOLI). This mechanism allows the DB to extract all firms' willingness to pay for data but does not allow him to change the offer he makes to a firm conditional on another firm's behavior. We now relax this assumption and consider two alternative selling mechanisms based on auctions, as in Braulin and Valletti (2016), Montes et al. (2019) and Bounie et al. (2021b, 2021a, 2022).[15]

First, we focus on the case where a DB can sell data through auctions with reserve prices (AR), as described by Jehiel and Moldovanu (2000). In particular, the DB chooses the partition set $\mathbf{P}$ and sets up n auctions. In each auction, all firms can participate. In any auction $i \in \{0, \ldots, n-1\}$, the DB offers a partition d_i and sets a reserve price v_i. To maximize the extraction of firm profits, the DB sets the reserve price in auction i equal to the highest willingness to pay for d_i among all firms. In particular, firm i has the highest valuation of partition d_i, as this particular partition is centered on that firm's location. The reserve price in auction i is

$$v_i = \pi_i^{\mathrm{W}}(\mathbf{P}) - \pi_i^{\mathrm{L}}(\mathbf{P}),$$

corresponding to firm i's willingness to pay for d_i, given $\mathbf{P}$. The vector of reserve prices is $\mathbf{v} = (v_0, v_1,\ v_2, \ldots, v_{n-1})$ and both $\mathbf{v}$ and $\mathbf{P}$ are common knowledge. Similar to Bounie et al. (2021b), the DB declares the maximum number of auctions he is going to fulfill, k, which is common knowledge prior to firms' bidding. A fulfilled auction is one where the transaction takes place. Declaring the maximum rather than the actual number of fulfilled auctions allows the DB to minimize firms' profits if they lose their auction.[16] The DB fulfils a subset $\mathbf{J}$ of auctions after the firms have placed their bids. Since some auctions may be unfulfilled, the partitions actually traded in equilibrium may differ from those initially offered in the auctions through $\mathbf{P}$. We denote with $\mathbf{P}^*$ the partition set *traded* in equilibrium, where $d_i \in \mathbf{P}^*$ is equal to $d_i \in \mathbf{P}$ if $i \in \mathbf{J}$ (i.e., the auction for d_i is fulfiled), and $d_i = 0$ otherwise (i.e., the auction for d_i is unfulfiled).

Second, we focus on auctions without reserve prices (AU), which thus decrease the DB's bargaining power with respect to the previously described mechanism. When the

[15] Note that the outcomes obtained under the auction mechanisms could also be achieved through richer contracts that allow the DB to change his offer conditional on firms' responses.

[16] In a duopoly Hotelling setup, where each firm has only one rival, as in Bounie et al. (2021a), this strategy corresponds to declaring to fulfill exactly $\frac{n}{2}$ auctions. The difference in our model is due to the fact that each firm has two direct rivals instead of one.

DB cannot set reserve prices, a firm can win its auction simply by bidding above the valuations of the other firms, which are lower than its own owing to their distance. Then, offering different partitions in auctions that the DB does not mean to conclude would increase firms' underbidding and erode the DB's profits.

Under both selling mechanisms, the updated timing of the model is as follows:

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

Stage 2. The DB chooses a partition set $\mathbf{P}$, the maximum number of auctions he will fulfil k and, under the AR mechanism, the reserve prices $\mathbf{v}$. Offers are non-renegotiable.

Stage 3. Firms that entered the market individually and simultaneously bid in the auctions.

Stage 4: The DB observes the bids and chooses a subset $\mathbf{J}$ of auctions to fulfill. The winning firms receive their respective partitions and pay their price to the DB, corresponding to $w_i = \pi_i^{\mathrm{W}}(\mathbf{P}) - \pi_i^{\mathrm{L}}(\mathbf{P})$.

Stage 5. Firms set basic prices p_i^{B} for the unidentified consumers.

Stage 6. Firms set tailored prices ${p_i}^{\mathrm{T}}(x)$ for the identified consumers if they have won an auction. Consumers purchase the product , and profits are made.

The previous literature on downstream Hotelling models (Braulin and Valletti, 2016; Montes et al., 2019; Bounie et al., 2021b) has shown how a monopolistic DB could either sell data to one or both downstream firms. Our circular city model can be seen as a concatenation of Hotelling segments, with symmetric firms located at the endpoints of each segment. Due to firms' symmetry, in equilibrium, the DB adopts the same strategy for all Hotelling segments, entailing either the sale of data to all firms (i.e., both firms in each market segment) or the sale of data to every other firm (i.e., only one firm in each market segment). In 6, we show that the sale of data to all firms leads to the same outcome as under TIOLI. Intuitively, if the DB wants to sell data to all firms, he needs to conclude all auctions, and thus he cannot condition his prices on a given firm's response. Therefore, we now focus on the sale of data to every other firm under the two selling mechanisms.

To streamline the analysis, we restrict it to the case in which the number of entering firms is even, allowing for a symmetric layout of firms on the circular market. In 6, we remove the hypothesis to include any number of entering firms, including odd values of

TABLE 1. Comparison between selling mechanisms

Selling Mechanism (SM)	Equilibrium strategy	Partition size	π_{DB}^{SM}	Share of identified consumers	$n_{SM}^{*}/\widetilde{n}$	$\widetilde{CS} - CS_{SM}$
AR	Every other firm	6/7n	$\frac{29}{28}\sqrt{tF}$	42.85%	1/2	$\frac{213}{196}\sqrt{tF}$
AU	Every other firm	4/3n	$\frac{4}{5}\sqrt{tF}$	66.67%	5/9	$\frac{29}{45}\sqrt{tF}$
TIOLI ($n^* < \hat{n}$)	All firms	3/2n*	$\frac{1}{2}\sqrt{tF}$	100%	1/2	$\frac{1}{4}\sqrt{tF}$
TIOLI ($n^* \geq \hat{n}$)	All firms	≈ 31/50n	$\approx \frac{1}{3}\sqrt{tF}$	≈ 62%	≈ 3/4	$\approx \frac{3}{20}\sqrt{tF}$

* overlapping partitions.

n. In the following proposition, we study the implications of the AR and AU selling mechanisms for equilibrium profits, the quantity of data sold, and consumer surplus.

PROPOSITION 4. *In equilibrium, the DB sells non-overlapping partitions in both AR and AU. The DB's equilibrium profits are such that* $\pi_{DB}^{AR*} > \pi_{DB}^{AU*} > \pi_{DB}^{TIOLI*}$. *Under AR and AU, the quantity of data sold to downstream firms and consumer surplus are lower than under TIOLI.*

PROOF. See 6. □

Table 1 summarizes the results of Proposition 4.

The DB's ability to extract surplus from firms is strengthened under the auction mechanisms, as he can maximize their winning profits and minimize their losing profits at the same time. Under AR, the DB offers the whole dataset in the auctions he does not want to fulfill so that a losing firm faces perfectly informed rivals. Instead, under AU, the DB offers same-sized partitions in all auctions but still only fulfills half of them. As there are no reserve prices, firms can win their auction by beating their rivals' offers, and thus the DB opts for same-size partitions to avoid underbidding.

While the auction-based selling mechanisms are more profitable for the DB than TIOLI, the higher price of data results in a reduction of firm entry. Moreover, under the auctions, the DB sells smaller partitions in order to temper downstream competition and increase firms' willingness to pay for data. The combination of these two effects results in a further decrease in consumer surplus, which is minimized under AR.

Our analysis shows that increasing the DB's bargaining power allows him to better extract surplus from firms, which drives up the entry barrier effect of data and, in turn, increases consumer harm. Forcing the DB to sell data through TIOLI offers, or to sell data to all entering firms, tempers the consumer harm but cannot completely prevent it.

5.2. Committing to the price of data: an alternative timing

In our baseline setup, firms enter the market in the initial stage and then participate in auctions to acquire data. This framework is, for example, consistent with traditional businesses that are already established before expanding into online markets and allow price discrimination through data. However, such timing may be less suitable to represent emerging digital markets, in which firms already know before entering that obtaining consumer data would give them an edge over the competition. In this Section, we explore the possibility that the DB sells data prior to firms' entry. As a consequence, firms make the entry decision only after observing the offer of data by the DB. In particular, the timing we analyze in this Section under AR is as follows:

Stage 1. The DB chooses a partition set $\mathbf{P}$, the reserve prices $\mathbf{v}$, and the maximum number of auctions he will fulfil k. Offers are non-renegotiable. All this information is common knowledge.

Stage 2. Firms individually and simultaneously bid in the auctions.

Stage 3. The DB observes the bids and chooses a subset $\mathbf{J}$ of auctions to fulfill. The winning firms receive their respective partitions and pay their price to the DB, corresponding to $w_i = \pi_i^{\mathrm{W}}(\mathbf{P}) - \pi_i^{\mathrm{L}}(\mathbf{P})$.

Stage 4. Firms enter the market and pay the fixed cost F.

Stage 5. Firms set basic prices p_i^{B} for the unidentified consumers.

Stage 6. Firms set tailored prices ${p_i}^{\mathrm{T}}(x)$ for the identified consumers if they have won an auction. Consumers purchase the product , and profits are made.

An analogous timing arises when data are sold through TIOLI offers or AU, the only difference being in stages from 1 to 3 that are specific to each selling mechanism.

In this setup, firms' equilibrium prices are defined as in the baseline timing, given that firms' price setting stage still takes place in the final stages of the game. However, the DB's strategy substantially departs from that of our baseline model. In fact, in the baseline model, the DB chooses his strategy by taking the number of entering firms as given. Conversely, under this alternative setting, the DB anticipates the effect of the data sale on firms' entry. We analyze the DB's strategy under all the selling mechanisms presented in Section 5.1. The main results in this setting are described in the following proposition.

Proposition 5. *If firms purchase data before entering the market, in equilibrium, the DB adopts the following strategies:*

TABLE 2. Comparison between selling mechanisms – Alternative timing

Selling Mechanism (SM)	Equilibrium strategy	Partition size	π_{DB}^{SM}	Share of identified consumers	$n^*_{SM}/\widetilde{n}$	$\widetilde{CS} - CS_{SM}$
AR	Every other firm	6/7n	$\frac{29}{28}\sqrt{tF}$	42.85%	1/2	$\frac{213}{196}\sqrt{tF}$
AU	Every other firm	3/2n*	$\frac{7}{8}\sqrt{tF}$	75%	1/2	$\frac{3}{4}\sqrt{tF}$
TIOLI	All firms	3/2n*	$\frac{1}{2}\sqrt{tF}$	100%	1/2	$\frac{1}{4}\sqrt{tF}$

* overlapping partitions.

(i) Under Take-It-Or-Leave-It offers (TIOLI), the DB opts for the sale to all firms and offers the whole dataset to all entering firms.
(ii) Under the auction with reserve prices (AR), the DB opts for selling to every other firm and offers d_H^* in the auctions he wants to fulfill and the whole dataset in the ones he does not want to fulfill;
(iii) Under the auction without reserve prices (AU), the DB opts for selling to every other firm and offers the whole dataset in all auctions;

The number of entering firms is $n^*_{\text{TIOLI}} = n^*_{\text{AR}} = n^*_{\text{AU}} = \frac{\widetilde{n}}{2}$. DB's profits are greater or equal, and consumer surplus is lower or equal than in the baseline model.

PROOF. See 6. □

Table 2 summarizes the findings of Proposition 5.

When the DB anticipates the effect of his strategy on firms' entry, the number of firms in equilibrium is the same, regardless of the selling mechanism. Such entry is equal to the one obtained in our baseline model under TIOLI when the downstream market is characterized by a low degree of horizontal differentiation and is the minimum that can be achieved in the market. The DB has the incentive to minimize entry to increase entering firms' profits (before paying for data), which he can then extract through the price of data. To better understand the implications of the different timing, let us focus on specific selling mechanisms.

Under TIOLI, the DB offers the whole dataset to all entering firms. By doing so, the DB increases the threat to firms that lose the auction, reducing their profits and thus entry into the market.

Under AR, the DB's strategy maximizes his profits for any given number of entering firms, as firms' profits when losing their auction do not depend on the partitions sold in equilibrium. In equilibrium, firms' expected profits when entering the market are the same as the previous extension, leading to the same market outcomes.

Finally, the alternative timing alters the DB's strategy under AU. Although he still opts for selling to every other entering firm, as in AR, he instead offers $d^*_{\text{AU}} = 1$ (i.e., the whole dataset), as opposed to the amount of data $d^*_{\text{AU}} = \frac{4}{3n}$ offered under our baseline timing. By doing so, the DB minimizes firms' profits when they lose and, in turn, the number of entering firms. Although the DB cannot maximize his profits by maximizing winning firms' profits, he can still do so by reducing competition in the downstream market.

To sum up, we find that if the data sale occurs before the firms' entry, the DB always maximizes his entry barrier effect. As his bargaining power is reduced, the DB floods the downstream market with data to reduce firms' profits after paying for data and, in turn, their entry. This strategy ultimately harms consumers, who are always worse off than in the benchmark due to the increase in the downstream market's concentration. Nonetheless, we find that consumer harm is lower under TIOLI than under the auctions mechanisms, consistent with the result obtained in Section 5.1.

6. 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.

Overall, our results show that consumer surplus is lower in the presence of a monopolistic DB, whereas the DB mostly appropriates total welfare. As a consequence, if the weight of the DB's profits in the welfare function is sufficiently low, the presence of a DB is welfare decreasing.

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. However, we also find that consumer harm can be decreased by properly regulating the DB's selling mechanism. In particular, a competition authority could either mandate the sale of data to all entering firms or enforce the use of direct sales (i.e., TIOLI offers). Such policies would effectively lower the DB's bargaining power but would also lead to an increase in the amount of data sold. Therefore, the ensuing increase in competition would also be accompanied by a lower degree of consumer privacy.

Finally, we find that the DB's negative effect on consumer surplus is stronger if he uses alternative selling mechanisms such as auctions or if firms purchase data before they decide to enter the market. In the latter scenario, the DB always chooses a strategy that minimizes firms' entry by flooding the downstream market with data.

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 further 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 into the downstream market and for consumers.

Acknowledgment: 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.

Bibliography

Acquisti, A., and Varian, H. R. (2005). Conditioning Prices on Purchase History. *Marketing Science*, *24*(3), 367–381. https://doi.org/10.1287/mksc.1040.0103

Aparicio, D., Metzman, Z., and Rigobon, R. (2021). *The pricing strategies of online grocery retailers* [NBER working paper n. 28639].

Armstrong, M., and Vickers, J. (2001). Competitive Price Discrimination. *The RAND Journal of Economics*, *32*(4), 579–605. https://doi.org/10.2307/2696383

Bergemann, D., Bonatti, A., and Smolin, A. (2018). The design and price of information. *American economic review*, *108*(1), 1–48. https://www.aeaweb.org/articles?id=10.1257/aer.20161079

Bergemann, D., and Bonatti, A. (2011). Targeting in advertising markets: Implications for offline versus online media. *The RAND Journal of Economics*, *42*(3), 417–443. https://doi.org/10.1111/j.1756-2171.2011.00143.x

Bergemann, D., and Bonatti, A. (2015). Selling Cookies. *American Economic Journal: Microeconomics*, *7*(3), 259–294. https://doi.org/10.1257/mic.20140155

Bester, H., and Petrakis, E. (1996). Coupons and oligopolistic price discrimination. *International Journal of Industrial Organization*, *14*(2), 227–242. https://doi.org/10.1016/0167-7187(94)00469-2

Bounie, D., Dubus, A., and Waelbroeck, P. (2021a). *Competition and mergers with strategic data intermediaries* (SSRN Scholarly Paper No. ID 3918829). Rochester, NY. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3918829

Bounie, D., Dubus, A., and Waelbroeck, P. (2021b). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, *52*(2), 283–313. https://doi.org/10.1111/1756-2171.12369

Bounie, D., Dubus, A., and Waelbroeck, P. (2022). Market for information and selling mechanisms. *Economics Working Paper Series*, *22/367*.

Braulin, F. C., and Valletti, T. (2016). Selling customer information to competing firms. *Economics Letters*, *149*, 10–14. https://doi.org/10.1016/j.econlet.2016.10.005

Cachon, G. P., and Lariviere, M. A. (2005). Supply chain coordination with revenue-sharing contracts: Strengths and limitations. *Management science*, *51*(1), 30–44.

Chen, Z., Choe, C., and Matsushima, N. (2020). Competitive Personalized Pricing. *Management Science*, *66*(9), 4003–4023. https://doi.org/10.1287/mnsc.2019.3392

Choe, C., Cong, J., and Wang, C. (2023). Softening competition through unilateral sharing of customer data. *Management Science.*

Clifford, S. (2012). Shopper alert: Price may drop for you alone. *The New York Times, https://www.nytimes.com/2012/08/10/business/supermarkets-try-customizing-prices-for- shoppers.html.*

Colombo, S. (2018). Behavior- and characteristic-based price discrimination. *Journal of Economics & Management Strategy*, *27*(2), 237–250. https://doi.org/10.1111/jems.12244

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society*, *20*(1), 88–104. https://doi.org/10.1177/1461444816657096

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 No. ID 3547379). Social Science Research Network. Rochester, NY. Retrieved March 1, 2022, from https://papers.ssrn.com/abstract=3547379

de Cornière, A. (2016). Search Advertising. *American Economic Journal: Microeconomics*, *8*(3), 156–188. https://doi.org/10.1257/mic.20130138

DeGraba, P. (1990). Input market price discrimination and the choice of technology. *The American Economic Review*, *80*(5), 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. Seville, Spain. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/strategic-data-sales-competing-firms

FTC. (2014). *Data brokers: A call for transparency and accountability* (tech. rep.). Federal Trade Commission, Washington, DC. Washington, DC.

Fudenberg, D., and Villas-Boas, J. M. (2006). Behavior-based price discrimination and customer recognition. *Handbook on Economics and Information Systems*, *1*, 377–436.

Greenhut, M. L., and Ohta, H. (1976). Related market conditions and interindustrial mergers. *The American Economic Review*, *66*(3), 267–277.

Gu, Y., Madio, L., and Reggiani, C. (2019). Exclusive Data, Price Manipulation and Market Leadership. *Cesifo Working Paper No. 7853.*

Hagiu, A., and Wright, J. (2020). Data-enabled learning, network effects and competitive advantage. *Working Paper.*

Hart, O. D., and Tirole, J. (1988). Contract Renegotiation and Coasian Dynamics. *The Review of Economic Studies*, *55*(4), 509–540. https://doi.org/10.2307/2297403

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

Jehiel, P., and Moldovanu, B. (2000). Auctions with Downstream Interaction among Buyers. *The RAND Journal of Economics*, *31*(4), 768–791. https://doi.org/10.2307/2696358

Johnen, J. (2020). Dynamic competition in deceptive markets. *The RAND Journal of Economics*, *51*(2), 375–401. https://doi.org/10.1111/1756-2171.12318

Judiciary Committee. (2020). *Investigation of Competition in Digital Markets: Majority Staff Report and Recommendations* (tech. rep.). US House of Representatives. https://judiciary.house.gov/uploadedfiles/competition_in_digital_markets.pdf

Liu, Q., and Serfes, K. (2004). Quality of Information and Oligopolistic Price Discrimination. *Journal of Economics & Management Strategy*, *13*(4), 671–702. https://doi.org/10.1111/j.1430-9134.2004.00028.x

Liu, Q., and Serfes, K. (2005). Imperfect price discrimination, market structure, and efficiency. *Canadian Journal of Economics/Revue canadienne d'économique*, *38*(4), 1191–1203.

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*, 79–84. https://doi.org/10.1145/2390231.2390245

Montes, R., Sand-Zantman, W., and Valletti, T. (2019). The Value of Personal Information in Online Markets with Endogenous Privacy. *Management Science*, *65*(3), 955–1453. https://doi.org/10.1287/mnsc.2017.2989

OECD. (2018). *Personalised pricing in the digital era* (Technical Report DAF/COMP(2018)13). OECD, Directorate for Financial and Enterprise Affairs Competition Committee. www.oecd.org/daf/competition/personalised-pricing-in-the-digital-era.htm

Rhodes, A., and Zhou, J. (2021). Personalized Pricing and Privacy Choice. *Working Paper*. https://www.economics.utoronto.ca/index.php/index/research/downloadSeminarPaper/997875056

Salop, S. C. (1979). Monopolistic Competition with Outside Goods. *The Bell Journal of Economics*, *10*(1), 141–156. https://doi.org/10.2307/3003323

Shaffer, G., and Zhang, Z. J. (1995). Competitive Coupon Targeting. *Marketing Science*, *14*(4), 395–416. https://doi.org/10.1287/mksc.14.4.395

Shy, O., and Stenbacka, R. (2016). Customer Privacy and Competition. *Journal of Economics & Management Strategy*, *25*(3), 539–562. https://doi.org/10.1111/jems.12157

Taylor, C. R. (2003). Supplier Surfing: Competition and Consumer Behavior in Subscription Markets. *The RAND Journal of Economics*, *34*(2), 223–246. https://doi.org/10.2307/1593715

Taylor and Wagman, L. (2014). Consumer privacy in oligopolistic markets: Winners, losers, and welfare. *International Journal of Industrial Organization*, *34*, 80–84. https://doi.org/10.1016/j.ijindorg.2014.02.010

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

Tyagi, R. K. (1999). On the effects of downstream entry. *Management science*, *45*(1), 59–73.

Valentino-DeVries, J., J. Singer-Vine, J., and Soltani, A. (2012). Websites vary prices, deals based on users' information. *The Wall Street journal*.

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), 486–501. https://doi.org/10.2307/1593704

Appendix

Proof of Lemma 1. The proof proceeds in two steps. First, we focus on the case where firm i serves both identified and unidentified consumers. Second, we analize the case in which firm i only serves identified consumers.

Step 1.

To find firms' equilibrium prices, we solve the equation system composed by the FOCs of firms' profits with respect to their basic prices (i.e., equation (11) $\forall i \in \{0, \ldots, n-1\}$). In matricial form we can write $\mathbf{A} * \mathbf{p} = \mathbf{b}$, 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. We invert this matrix by following Searle1979, finding

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

where $a_j = -\frac{1}{2\sqrt{3}} * \left(\frac{\left(2+\sqrt{3}\right)^j}{1-\left(2+\sqrt{3}\right)^n} - \frac{\left(2-\sqrt{3}\right)^j}{1-\left(2-\sqrt{3}\right)^n} \right)$. A property of circulant 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 $\mathbf{p} = \mathbf{A^{-1}} * \mathbf{b}$.

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

Thus, equilibrium basic prices are

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

Given that $\sum_{j=0}^{n-1} a_j = \frac{1}{2}$, we obtain

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

From (A.2), all basic prices are decreasing in all firms' partitions. Moreover, as a_j is decreasing in j, we find that firm i's profits are more influenced by partitions obtained by firms that are closer to its location.

To study the effects of data on firm i's profits, let us focus on (9). It is useful to rewrite firms' basic prices isolating the term with d_i from the sum:

$$
p_i^{\mathrm{B}*} = \frac{t}{n} - 2td_i a_0 - 2t \sum_{j=1}^{n-1} d_{i+j} a_j \tag{A.3}
$$

and

$$
p_{i+1}^{\mathrm{B}*} = p_{i-1}^{\mathrm{B}*} = \frac{t}{n} - 2td_i a_1 - 2t \sum_{j=0, j\neq 1}^{n-1} d_{i+j} a_j. \tag{A.4}
$$

For simplicity, let us denote with m the last term of (A.3), and with l the last term of (A.4). By substituting (A.3) and (A.4) in (9), we obtain

$$
\begin{aligned}
\pi_i^{W*}(\mathbf{P}) = -\frac{1}{n} + \frac{(l-m)m}{t} - \frac{1}{2}(1 + 4a_0(-1 + 2a_0 - 2a_1) + 4a_1)d_i^2 t + \\
\frac{t}{n^2} + \frac{d_i(((-1+2a_0)l + m - 4a_0 m + 2a_1 m)n + t - 2a_1 t)}{n} - F.
\end{aligned} \tag{A.5}
$$

From the Equation above, it is easy to see that all partitions $d_{j\neq i}$ decrease firm i's profits. To study the effects of d_i on $\pi_i^{W*}(\mathbf{P})$, we compute the FOC of the firm's profits with respect to d_i, and find that the FOC is positive whenever

$$
d_i < \frac{(-1+2a_0)ln + (1 - 4a_0 + 2a_1)mn + t - 2a_1 t}{(1 + 4a_0(-1+2a_0-2a_1) + 4a_1)nt},
$$

Note that $m < l$ as $a_0 > a_1$. By denoting with $\bar{d}$ the right-hand side of the equation, we conclude that when $d_i < \bar{d}$, d_i positively affects firm i's profits.

Step 2. When firm i can identify all the consumers it serves, its profits are

$$\pi_i^{\mathrm{W}}(\mathbf{P}) = \int_{\hat{x}_{i-1,i}}^{\hat{x}_{i,i+1}} p_i^{\mathrm{T}}(x)\, dx - F. \tag{A.6}$$

It is straightforward to see that p_i^B affects (A.6) only through the extremes of the integral. In particular, a decrease of p_i^B increases the integration interval and, in turn, firm i's profits. Thus, when firm i can identify all consumers, it sets its basic price as low as possible, i.e., $p_i^{B*} = 0$.

Proof of Lemma 2. When firm i does not buy data, basic prices are obtained by a slight variation of (A.1) in Lemma 1, whereby the i-th component of $\mathbf{b}$ is $\frac{2t}{n}$ instead of $\frac{2t}{n} - 2td_i$. The results of this subgame are denoted with the superscript L. Equilibrium basic prices are thus

$$p_i^{\mathrm{B\ L}*} = \frac{t}{n} - 2t\sum_{j=0}^{n-1} d_{i+j}a_j + 2td_ia_0 = \frac{t}{n} - m, \tag{A.7}$$

$$p_{i+1}^{\mathrm{B\ L}*} = \frac{t}{n} - 2t\sum_{j=0}^{n-1} d_{i+1+j}a_j + 2td_ia_1 = \frac{t}{n} - l, \tag{A.8}$$

$$p_{i-1}^{\mathrm{B\ L}*} = \frac{t}{n} - 2t\sum_{j=0}^{n-1} d_{i-1,1}a_j + 2td_ia_1 = \frac{t}{n} - l. \tag{A.9}$$

Basic prices are higher than in Lemma 1, as $d_i = 0$. By substituting (A.7), (A.8) and (A.9) in (10), we obtain

$$\pi_i^{L*}(\mathbf{P}) = \frac{t}{n^2} - \frac{l}{n} + \frac{(l-m)m}{t} - F. \tag{A.10}$$

By comparing (A.10) and firm's profits when obtaining data, we find that $\pi_i^{W*}(\mathbf{P}) > \pi_i^{L*}(\mathbf{P})$.

Proof of Lemma 3. DB's profits can be written as

$$\max_{d_0,d_1,\ldots,d_{n-1}} \pi_{DB}(\mathbf{P}) = \sum_{i=0}^{n-1} \left(\pi_i^{\mathrm{W}*}(\mathbf{P}) - \pi_i^{\mathrm{L}*}(\mathbf{P})\right). \tag{A.11}$$

By substituting the firm's profits when obtaining data and (A.10) in (A.11) and computing the FOCs with respect to $d_i \forall i \in \{0, \ldots, n-1\}$, we find that DB's profits are influenced in the same way by all the partitions he sells. Thus, due to simmetry, we can conclude that in equilibrium the DB sells equally sized to all firms, i.e., $d_i = d\ \forall i \in \{0, \ldots, n-1\}$.

Let us focus on firms' prices and profits when firm i buys data. By setting $d_i = d$ $\forall i \in \{0, \ldots, n-1\}$ in (A.3) and (A.4), we obtain $p_i^{B*} = \frac{t}{n} - td$ $\forall i \in \{0, \ldots, n-1\}$. As all basic prices are equal, indifferent consumers are located in the middle between firms' locations, i.e., $\hat{x}_{i,i+1} = \frac{2i+1}{2n}$. By substituting the equilibrium prices in (9), we obtain

$$\pi_i^{W*} = \frac{t}{n^2} - \frac{td^2}{2} - F. \tag{A.12}$$

Firms identify all consumers when $\frac{i}{n} + \frac{d}{2} \geq \hat{x}_{i,i+1}$, which we can rewrite as $d \geq \frac{1}{n}$. Thus, when $d \geq \frac{1}{n}$, all firms set their basic prices equal to 0. By replacing these basic prices in (8), we obtain

$$\pi_i^{W*} = \frac{t}{2n^2} - F. \tag{A.13}$$

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 (10). By setting $d_i = d$ $\forall i \in \{0, \ldots, n-1\}$ in (A.7), (A.8) and (A.9), we obtain

$$p_i^{\mathrm{B\ L}*}(\mathbf{P}) = \frac{t}{n} - td + 2tda_0 \quad \text{and} \quad p_{i-j}^{\mathrm{B\ L}*}(\mathbf{P}) = p_{i+j}^{\mathrm{B\ L}*}(\mathbf{P}) = \frac{t}{n} - td + 2tda_j. \tag{A.14}$$

By substituting (A.14) in (10), we find

$$\pi_i^{\mathrm{L}*}(\mathbf{P}) = \left(\frac{t}{n} - td + 2tda_0\right)\left(2d(a_1 - a_0) + \frac{1}{n}\right) - F. \tag{A.15}$$

(ii) From (A.14), 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.14), 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

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

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. Without loss of generality, we focus on firms $i-1$ and i. 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

$$\pi_{i-1}^{\mathrm{W}}(\mathbf{P}) = \int_{\widehat{x}_{i-2,i-1}}^{\frac{i-1}{n}} p_{i-1,i-2}^{\mathrm{T}}(x)\,dx + \int_{\frac{i-1}{n}}^{\frac{i-1}{n}+\frac{d}{2}} p_{i-1,i}^{\mathrm{T}}(x)\,dx$$
$$+ p_{i-1}^{\mathrm{B}}(\mathbf{P})\left(\widehat{x}_{i-1,i} - \frac{i-1}{n} - \frac{d}{2}\right) - F. \quad \text{(A.16)}$$

Firm i's profits are given by (10). The FOCs of (**??**) and (10) give us the equilibrium basic prices:

$$p_{i-1}^{\mathrm{B}*}(\mathbf{P}) = \frac{t\,(3-2nd)}{5n} \quad \text{and} \quad p_i^{\mathrm{B\ L}*}(\mathbf{P}) = \frac{t\,(4-nd)}{5n}. \quad \text{(A.17)}$$

Substituting (A.17) into (10), we obtain

$$\pi_i^{\mathrm{L}*}(\mathbf{P}) = \frac{t(nd-4)^2}{25n^2} - F. \quad \text{(A.18)}$$

(iii) From (A.17), informed firms set positive basic prices as long as $d < \frac{3}{2n}$. Thus, when $d \geq \frac{3}{2n}$, all informed firms set their basic prices equal to 0. Firm i's profits can thus be computed by setting $d = \frac{3}{2n}$ in (A.18), leading to

$$p_i^{\mathrm{B}*}(\mathbf{P}) = \frac{t}{2n} \quad \text{and} \quad \pi_i^{\mathrm{L}*}(\mathbf{P}) = \frac{t}{4n^2} - F. \quad \text{(A.19)}$$

Comparing (A.15), (A.18) and (A.19) we find that firm i's profits when not buying data are strictly decreasing in d if $d < \frac{3}{2n}$, and constant otherwise.

Proof of Proposition 1. By using the equations for $\pi_i^{\mathrm{W}*}(\mathbf{P})$ and $\pi_i^{\mathrm{L}*}(\mathbf{P})$ from Lemma 3, we can express the DB's profits as

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

which is continuous in d. We first prove that the strategy of setting $d \in [d_1, \frac{3}{2n})$ (corresponding to the second and third part of DB's profits) is always suboptimal. 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 would always prefer the first part to the second one. 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, the DB would always prefer the fourth part to the third one.

Let us now consider the two strategies $d < d_1$ and $d \geq \frac{3}{2n}$. If the DB sets $d < d_1$, his profits are maximized for

$$d^* = \frac{1 - 2a_1}{n\left(-8a_0^2 + a_0\left(8a_1 + 4\right) - 4a_1 + 1\right)}. \tag{A.20}$$

By substituting (A.20) in DB's profits, we obtain

$$\pi_{\mathrm{DB}^*} = \frac{t}{2}(1 - 2a_1)d^*. \tag{A.21}$$

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

$$\pi_{\mathrm{DB}}^* = \frac{t}{4n}. \tag{A.22}$$

We now compare (A.21) and (A.22). The DB sets $d = d^*$ if

$$\frac{(1 - 2a_1)^2}{(-8a_0^2 + a_0(8a_1 + 4) - 4a_1 + 1)} \geq \frac{1}{2}. \tag{A.23}$$

Numerical analysis allows us to find that inequality (A.23) 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.

If $n < \hat{n}$, the number of entering firms is given by the free entry condition

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

leading to $n^* = \frac{1}{2}\sqrt{\frac{t}{F}}$.

If instead $n \geq \hat{n}$, the number of entering firms is obtained from the condition

$$\pi_i^{\mathrm{L}*}(\mathbf{P}) = \left(\frac{t}{n} - td^* + 2td^*a_0\right)\left(2d^*(a_1 - a_0) + \frac{1}{n}\right) - F = 0, \tag{A.24}$$

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

$$d^* = \frac{\alpha(n)}{n}, \tag{A.25}$$

where

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

By substituting (A.25) in (A.24), we obtain

$$(1 - \alpha(n) + 2a_0\alpha(n))\,(2(a_1 - a_0\alpha(n) + 1) = \frac{F}{t}n^2. \tag{A.26}$$

Let us denote with $A(n)$ the left hand-side of (A.26). $A(n)$ is monotonically decreasing in n and quickly approaches an asymptote:

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

We approximate $A(n)$ with

$$\overline{A}(n) \approx \frac{1}{n^3} + \frac{36\sqrt{3}-99}{1644\sqrt{3}-2915} \quad \textit{and} \quad \underline{A}(n) \approx \frac{1}{n^3} + \frac{53}{100}. \tag{A.27}$$

The first (second) approximation overestimates (underestimates) the true value of $A(n)$ by less than 1% over its domain (i.e., $n \geq 2$). As the DB chooses his strategy given n, this approximation does not affect the DB's strategy, and it is only done to estimate the effect of the DB's strategies on firm entry. By substituting (A.27) in (A.26), we obtain

$$\frac{1}{n^3} + \frac{36\sqrt{3}-99}{1644\sqrt{3}-2915} - \frac{F}{t}n^2 = 0 \quad \textit{and} \quad \frac{1}{n^3} + \frac{53}{100} - \frac{F}{t}n^2 = 0. \tag{A.28}$$

To find an explicit solution to (A.28), we use the Newton-Raphson approximation method. We obtain

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

$$\textit{and} \quad \underline{n^*} \approx \frac{102400\frac{F}{t} + 11799\sqrt{\frac{t}{F}}}{200\left(\frac{512}{\frac{t}{F}^{\frac{3}{2}}} + 81\right)}, \tag{A.29}$$

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

Proof of Proposition 3. Since all firms set basic prices equal to 0 when $n < \hat{n}$, indifferent consumers are located at the center of each arch. We can thus write consumer surplus as

$$CS^{\text{TIOLI}}_{n<\hat{n}} = 2n^* \left(\int_{\frac{i}{n^*}}^{\frac{2i+1}{2n^*}} v - p_i^{\text{T}*}(x) - t\left(x - \frac{i}{n^*}\right) dx \right), \tag{A.30}$$

where

$$p_i^{\text{T}}(x) = -2tx + \frac{t}{n^*}(2i+1). \tag{A.31}$$

By replacing (A.31) in (A.30) we obtain

$$CS^{\text{TIOLI}}_{n<\hat{n}} = u - \frac{3t}{4n^*} = u - \frac{3}{2}\sqrt{tF}.$$

When $n \geq \hat{n}$, all firms offer equal basic prices and have equal market shares. 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

$$CS^* = 2n(\int_0^{\frac{d^*}{2}} u - tx - p_0^T(x)dx + \int_{\frac{d^*}{2}}^{\frac{1}{2n*}} u - tx - p_0^{B*}dx) = u - \frac{5t}{4n^*} + \frac{1}{2}n^*td^{*2} \quad \text{(A.32)}$$

Using (**??**) and (A.32), we find

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

By comparing the two results with the benchmark $\tilde{CS}$, 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. Then, the only components of total welfare are consumer surplus and the DB's profits. When $n < \hat{n}$, we have $CS^* = u - \frac{3}{2}\sqrt{tF}$ and $\pi^*_{DB} = \frac{1}{2}\sqrt{tF}$. By weighing the DB's profits for a coefficient α, total welfare is equal to

$$TW^* = u - \frac{3}{2}\sqrt{tF} + \frac{\alpha}{2}\sqrt{tF}$$

By setting $TW^* = \tilde{TW}$ and solving for α, we find that TW increases iff $\alpha \geq \frac{1}{2}$.

When instead $n \geq \hat{n}$, we have $CS^* = v - \frac{5t}{4n^*} + \frac{1}{2}n^*td^{*2}$ and $\pi^*_{DB} = \frac{t}{2}(1 - 2a_1)d^*$. By substituting the equilibrium values of n^* and d^*, and approximating the values of a_0 and a_1, we obtain

$$CS^* \approx u - \frac{7}{5}\sqrt{tF}$$

and

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

Total welfare is thus equal to

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

By setting $TW^* = \tilde{TW}$ and solving for α, we find that TW increases iff $\alpha \geq \frac{3}{7}$.

Proof of Proposition 4.

Sale to all firms. Suppose that the DB sells data to all firms, i.e., $\mathbf{P}^*_{\text{AR}} = \mathbf{P}^*_{\text{AU}} = (d, d, d, d, \ldots, d)$. Then, the DB sets up n auctions, with the aim of concluding all of them. Note that the DB cannot change the partitions he puts up for auction based on firms' bids, as he can only choose the number of auctions he wants to fulfil. Thus, the only way to obtain $\mathbf{P}^*$ is to offer $\mathbf{P} = \mathbf{P}^*$, and then concluding all of the auctions. This strategy corresponds to the one analyzed in Lemma 3, implying that the DB's equilibrium strategy and market outcomes under the auction mechanisms are the same than under the TIOLI sale.

Sale to alternating firms: Auction with reserve prices. In equilibrium, the DB offers a partition set $\mathbf{P}_{\mathrm{AR}}{}^{*} = (d, 0, d, 0, \ldots, 0)$. However, the DB can still set up an auction for firms who in equilibrium will not obtain data, and then he does not fulfil them. We show that the DB offers the whole dataset in these auctions, and thus the offered partition set is $\mathbf{P}_{\mathrm{AR}} = (d, 1, d, 1, \ldots, 1)$.

The DB's profits are

$$\pi_{\mathrm{DB}}(\mathbf{P}_{\mathrm{AR}}, \mathbf{J}) = \sum_{i \in \mathbf{J}} \left(\pi_i^{\mathrm{W}}(\mathbf{P}_{\mathrm{AR}}) - \pi_i^{\mathrm{L}}(\mathbf{P}_{\mathrm{AR}})\right). \tag{A.33}$$

Consider the set of data partitions $\{d_j\}_{j \notin \mathbf{J}}$. Since these partitions are offered in the auction, but are not actually sold in equilibrium, then $\pi_i^{\mathrm{W}}(\mathbf{P}_{\mathrm{AR}}) = \pi_i^{\mathrm{W}}(\mathbf{P}^{*}_{\mathrm{AR}})$ for all $i \in \mathbf{J}$, i.e., $\pi_i^{\mathrm{W}}(\mathbf{P}_{\mathrm{AR}})$ does not depend on $\{d_j\}_{j \notin \mathbf{J}}$. Hence, the DB chooses $\{d_j\}_{j \notin \mathbf{J}}$ so as to minimize $\pi_i^{\mathrm{L}}(\mathbf{P}_{\mathrm{AR}})$, given that doing so does not affect $\pi_i^{\mathrm{W}}(\mathbf{P}_{\mathrm{AR}})$.

Firm i's losing profits are

$$\pi_i^{\mathrm{L}}\left(\mathbf{P}_{\mathrm{AR}}\right) = p_i^{\mathrm{B}}\left(\mathbf{P}_{\mathrm{AR}}\right) \left(\frac{n\left(p_{i+1}^{\mathrm{B}}\left(\mathbf{P}_{\mathrm{AR}}\right) + p_{i-1}^{\mathrm{B}}\left(\mathbf{P}_{\mathrm{AR}}\right) - 2p_i^{\mathrm{B}}\left(\mathbf{P}_{\mathrm{AR}}\right)\right) + 2t}{2nt}\right) - F. \tag{A.34}$$

Equation (A.34), for any given p_i^{B}, is minimized when both $p_{i-1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}})$ and $p_{i+1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}})$ are minimized, that is, due to the non-negative constraint on prices,

$$p_{i-1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}}) = p_{i+1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}}) = 0.$$

This can be achieved by setting $d_{i-1} = d_{i+1} = 1$. In fact, focus on firm $i+1$. Then, by rearranging (8), its profit can be expressed as

$$\pi_{i+1}^{\mathrm{W}}(\mathbf{P}_{\mathrm{AR}}) = \int_{\hat{x}_{i,i+1}}^{\hat{x}_{i+1,i+2}} p_{i+1}^{\mathrm{T}}(x)\, dx - F. \tag{A.35}$$

Firm $i+1$ chooses $p_{i+1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}})$ so as to maximize (A.35). To this aim, note that function (A.35) is strictly decreasing in $p_{i+1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}})$. In fact, from (6) and (7), $p_{i+1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}})$ affects $\hat{x}_{i,i+1}$ and $\hat{x}_{i+1,i+2}$ but not $p_{i+1}^{\mathrm{T}}(\mathbf{P}_{\mathrm{AR}})$ in (A.35). In particular, a decrease of $p_{i+1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}})$ expands firm $i+1$'s market share by moving further away the two indifferent consumers. Given that function (A.35) is strictly decreasing in $p_{i+1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}})$, then he sets $p_{i+1}^{\mathrm{B}}(\mathbf{P}_{\mathrm{AR}}) = 0$. The same argument holds for $i-1$ by symmetry.

Having concluded that under AR the DB sets a partition set $\mathbf{P}_{\mathrm{AR}} = (d, 1, d, 1, \ldots, 1)$, and that in equilibrium he sells a partition set $\mathbf{P}_{\mathrm{AR}}{}^{*} = (d, 0, d, 0, \ldots, 0)$, we can now compute firms' profits.

Suppose that all firms who can obtain a partition of size d_{AR} centred on their location win their respective auctions, and that firm i is one of those. By following the same procedure as in the proof of Lemma 1, equilibrium basic prices are

$$p_{\mathrm{i}}^{\mathrm{B}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=\frac{t}{n}-2td_{\mathrm{AR}}\left(a_0+a_{\frac{n}{2}}+2\sum_{j=1}^{\frac{n}{4}-1}a_{2j}\right)=\frac{t}{n}-\frac{2}{3}td_{\mathrm{AR}} \tag{A.36}$$

and

$$p_{i+1}^{\mathrm{B}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=\ p_{i-1}^{\mathrm{B}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=\frac{t}{n}-\frac{1}{3}td_{\mathrm{AR}}. \tag{A.37}$$

Substituting (A.36) and (A.37) in (6), we obtain the indifferent consumers' locations:

$$\widehat{x}_{i-1,i}=\frac{2i-1}{2n}-\frac{d_{\mathrm{AR}}}{6}\quad\text{and}\quad\widehat{x}_{i,i+1}=\frac{2i+1}{2n}+\frac{d_{\mathrm{AR}}}{6}. \tag{A.38}$$

We can compute firm i's profits by substituting (A.36), (A.37) and (A.38) in (9), obtaining

$$\pi_i^{\mathrm{W}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=\frac{t}{n^2}+\frac{2d_{\mathrm{AR}}t}{3n}-\frac{7td_{\mathrm{AR}}^2}{18}-F. \tag{A.39}$$

Similarly, firm i+1's profits are

$$\pi_{i+1}^{\mathrm{L}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=\frac{t}{n^2}-\frac{2d_{\mathrm{AR}}t}{3n}+\frac{td_{\mathrm{AR}}^2}{9}-F. \tag{A.40}$$

We now focus on the case where winning firms only serve identified consumers. By comparing $\frac{i}{n}+\frac{d_{\mathrm{AR}}}{2}$ and $\hat{x}_{i,i+1}$, we find that firms only serve identified consumers when

$$d_{\mathrm{H}}\geq\frac{3}{2n}. \tag{A.41}$$

When (A.41) holds, firm i sets its basic price equal to 0 and its profits are given by (8). We find $p_{i+1}^{\mathrm{B}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)$ by solving the FOCs of (10) with $p_{i+2}^{\mathrm{B}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=p_i^{\mathrm{B}}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=0$, obtaining $p_{i+1}^{\mathrm{B}*}\ \left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=\frac{t}{2n}$. The same logic applies to firm i−1 by symmetry. By substituting the basic prices in the profits functions, we obtain

$$\pi_i^{\mathrm{W}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=\frac{9t}{8n^2}-F\quad\text{and}\quad\pi_{i+1}^{\mathrm{L}*}\left(\mathbf{P}_{\mathrm{AR}}^{*}\right)=\frac{t}{4n^2}-F.$$

Next, suppose that firm i loses its auction. Since the DB can fulfil up to $k=\frac{n}{2}+1$ auctions, he can now fulfil both firm $i+1$ and $i-1$'s auctions, that thus obtain the whole dataset. This subgame is the same as the case where firm i wins the auction and $d_{\mathrm{AR}}\geq\frac{3}{2n}$. As such, firm i's basic price and profits are equal to firm $i+1$'s ones in the previous subgame, leading to

$$p_i^{\mathrm{B}*}\left(\mathbf{P}_{\mathrm{AR}}\right)=\frac{t}{2n}\quad\text{and}\quad\pi_i^{\mathrm{L}*}\left(\mathbf{P}_{\mathrm{AR}}\right)=\frac{t}{4n^2}-F.$$

Turning to DB's profits, they can be computed as

$$\max_{d_{\mathrm{AR}}} \pi_{\mathrm{DB}} = \frac{n}{2}\left(\pi_i^{\mathrm{W}*}\left(\mathbf{P}_{\mathrm{AR}}\right) - \pi_i^{\mathrm{L}*}\left(\mathbf{P}_{\mathrm{AR}}\right)\right), \tag{A.42}$$

where

$$\pi_i^{\mathrm{W}*}\left(\mathbf{P}_{\mathrm{AR}}\right) = \begin{cases} \frac{t}{n^2} + \frac{2d_{\mathrm{AR}}t}{3n} - \frac{7td_{\mathrm{AR}}^2}{18} - F & \text{for } \ d_{\mathrm{AR}} < \frac{3}{2n} \\ \frac{9t}{8n^2} - F & \text{for } \ d_{\mathrm{AR}} \geq \frac{3}{2n} \end{cases} \tag{A.43}$$

$$\pi_i^{\mathrm{L}*}\left(\mathbf{P}_{\mathrm{AR}}\right) = \frac{t}{4n^2} - F \quad \text{for } \ 0 \leq d_{\mathrm{AR}} \leq 1.$$

We can rewrite DB's profits by substituting (A.43) in (A.42), obtaining

$$\max_{d_{\mathrm{AR}}} \pi_{\mathrm{DB}} = \begin{cases} \frac{n}{2}\left(\frac{3t}{4n^2} + \frac{2d_{\mathrm{AR}}t}{3n} - \frac{7td_{\mathrm{AR}}^2}{18}\right) & \text{for } \ d_{\mathrm{AR}} < \frac{3}{2n} \\ \frac{n}{2}\left(\frac{7t}{8n^2}\right) & \text{for } \ d_{\mathrm{AR}} \geq \frac{3}{2n}. \end{cases}$$

When the DB sets $d_{\mathrm{AR}} < \frac{3}{2n}$, the FOC of π_{DB} gives $d_{\mathrm{AR}}^* = \frac{6}{7n}$, resulting in profits equal to $\pi_{\mathrm{DB}}^* = \frac{29t}{56n}$. Conversely, when the DB sets $d_{\mathrm{AR}} \geq \frac{3}{2n}$, then his profits are constant with respect to d_{AR} and equal to $\pi_{\mathrm{DB}}^* = \frac{7t}{16n}$. By comparing the two results, we find that the DB maximizes his profits by setting $d_{\mathrm{AR}}^* = \frac{6}{7n}$. When the DB sells data to all firms, his profits are equal to $\pi_{\mathrm{DB}}^{\mathrm{TIOLI}*}(\mathbf{P}_{\mathrm{TIOLI}}) = \frac{t}{4n}$ when $n < \hat{n}$. By direct comparison, we find that, when $n < \hat{n}$, the DB prefers selling data to half of the entering firms.

When $n \geq \hat{n}$, DB's profits when selling to all firms are equal to

$$\pi_{\mathrm{DB}}^*(\mathbf{P}_{\mathrm{TIOLI}}) = \frac{t}{2}(1 - 2a_1)d_{\mathrm{TIOLI}}^*$$

where

$$d_{\mathrm{TIOLI}}^* = \frac{1 - 2a_1}{n\left(-8a_0^2 + a_0\left(8a_1 + 4\right) - 4a_1 + 1\right)}$$

The DB opts for selling data to half of the entering firms if

$$\frac{29}{56} > \frac{1}{2}\frac{(1 - 2a_1)^2}{-8a_0^2 + a_0\left(8a_1 + 4\right) - 4a_1 + 1}$$

Which is always satisfied for $n \geq 2$. We can thus conclude that the DB always opts for selling to half of the entering firms, regardless of n.

The number of entering firms will be such that their profits after paying for entry and data are 0. We obtain the number of entering firms by solving the free-entry condition

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

from which we find $n_{\mathrm{AR}}^* = \frac{1}{2}\sqrt{\frac{t}{F}}$.

Consumer utility in equilibrium is equal to

$$U(x,i) = v - p_i^*(x) - t\left(x - \frac{i}{n^*_{\text{AR}}}\right)$$

where $p_i^*(x)$ can either be $p_i^{\text{B}*}\left(\mathbf{P}^*_{\text{AR}}\right)$ if firm i does not identify the consumer located in x or $p_i^{\text{T}*}(x)$ if it identifies it. Due to the model's symmetry, we obtain total consumer surplus by computing it on the arch between firms i and $i+1$, and then multiplying it by n^*_{AR}.

Suppose that firm i wins its auction in equilibrium: as such, firm i and $i+1$ basic prices are given by (A.36) and (A.37), with $d^*_{\text{AR}} = \frac{6}{7n^*_{\text{AR}}}$. Firm i's tailored price is $p^{\text{T}*}_{i,i+1}(x) = p^{\text{B}*}_{i+1}\left(\mathbf{P}^*_{\text{AR}}\right) - 2tx + \frac{t}{n^*_{\text{AR}}}(2i+1)$, and the indifferent consumer is located in $\widehat{x^*}_{i,i+1} = \frac{2i+1}{2n^*_{\text{AR}}} + \frac{d^*_{\text{AR}}}{6}$. Consumer surplus is thus equal to

$$CS = n^*_{\text{AR}}\left(\int_{\frac{i}{n^*_{\text{AR}}}}^{\frac{i}{n^*_{\text{AR}}}+\frac{d^*_{\text{AR}}}{2}} v - p^{\text{T}*}_{i,i+1}(x) - t\left(x - \frac{i}{n^*_{\text{AR}}}\right)dx + \int_{\frac{i}{n^*_{\text{AR}}}+\frac{d^*_{\text{AR}}}{2}}^{\widehat{x^*}_{i,i+1}} v - p_i^{\text{B}*}\left(\mathbf{P}^*_{\text{AR}}\right)\right.$$
$$\left. - t\left(x - \frac{i}{n^*_{\text{AR}}}\right)dx + \int_{\widehat{x^*}_{i,i+1}}^{\frac{i+1}{n^*_{\text{AR}}}} v - p^{\text{B}*}_{i+1}\left(\mathbf{P}^*_{\text{AR}}\right) - t\left(\frac{i+1}{n^*_{\text{AR}}} - x\right)dx\right). \quad \text{(A.44)}$$

By substituting the prices and the indifferent consumer's location in (**??**), we obtain

$$CS = v - \frac{5t}{4n^*_{\text{AR}}} + \frac{\text{nt}{d^*_{\text{AR}}}^2}{9}, \quad \text{(A.45)}$$

which we can rewrite as

$$CS = v - \frac{229}{98}\sqrt{tF}. \quad \text{(A.46)}$$

Consumer surplus where data are absent is equal to $\widetilde{CS} = v - \frac{5t}{4\widetilde{n}}$, with $\widetilde{n} = \sqrt{\frac{t}{F}}$ (Salop, 1979). This implies $\widetilde{CS} = v - \frac{5}{4}\sqrt{tF}$, which is always higher than CS in (A.46).

Sale to alternating firms: Auction without reserve prices. The DB can offer a partition set $\mathbf{P}_{\text{AU}}$ which is different from $\mathbf{P}^*_{\text{AU}}$, as he can decide to not fulfil some of the auctions he sets up. However, without reserve prices, firms can win their auctions by beating their rivals' offers. We prove that the DB offers a partition set $\mathbf{P}_{\text{AU}} = (d_{\text{AU}}, d_{\text{AU}}, d_{\text{AU}}, d_{\text{AU}}, \ldots, d_{\text{AU}}, d_{\text{AU}})$.

The DB can offer a generic partition set $\mathbf{P}_{\text{AU}} = (d_{\text{AU}}, d_1, d_{\text{AU}}, d_3, \ldots, d_{\text{AU}}, d_{n-1})$. If firm 0 deviates, the DB would want to fulfil the auctions where he offers d_{n-1} and d_1. As d_{n-1} and d_1 have the same effect on firm 0's profits, the DB would set $d_{n-1} = d_1$. The same holds for any deviating firm, and thus the DB sets $d_1 = d_3 = \ldots = d_{n-1} = d$. If no even-indexed firm deviates, in equilibrium the DB fulfils the auctions where he offered d_{AU}. If one even-indexed firm deviates, the DB can instead fulfil all the auctions where he offered d. Without loss of generality, we focus on firms 0 and 1. Firms' willingness to pay for data are equal to

$$\pi_0^{\mathrm{W}}\left(d_{\mathrm{AU}}, 0, d_{\mathrm{AU}}, 0, \ldots, d_{\mathrm{AU}}, 0\right) - \pi_0^{\mathrm{L}}(0, d, 0, d, \ldots, 0, d)$$

and

$$\pi_1^{\mathrm{W}}(0, d, 0, d, \ldots, 0, d) - \pi_1^{\mathrm{L}}(d_{\mathrm{AU}}, 0, d_{\mathrm{AU}}, 0, \ldots, d_{\mathrm{AU}}, 0).$$

Suppose that $d > d_{\mathrm{AU}}$ (the opposite case is solved similarly): then it is straightforward to show that

$$\pi_1^{\mathrm{W}}(0, d, 0, d, \ldots, 0, d) - \pi_1^{\mathrm{L}}(d_{\mathrm{AU}}, 0, d_{\mathrm{AU}}, 0, \ldots, d_{\mathrm{AU}}, 0 > \\ \pi_0^{\mathrm{W}}\left(d_{\mathrm{AU}}, 0, d_{\mathrm{AU}}, 0, \ldots, d_{\mathrm{AU}}, 0\right) - \pi_0^{\mathrm{L}}(0, d, 0, d, \ldots, 0, d),$$

that is, firm 1's willingness to pay is higher than firm 0's one. Because there are no reserve prices, firm 1 can win its auction by offering firm 0's willingness to pay plus ε, where ε is an arbitrary small number. As such, the DB chooses d as low as possible to maximize firm 0's willingness to pay and, in turn, firm 1's: that is, he chooses $d = d_{\mathrm{AU}}$.

To find equilibrium prices and profits, we focus on a generic firm i. In equilibrium, firm i's basic prices and profits are the same as in the auction with reserve prices (see (A.39)). Instead, when firm i loses its auction, its basic price and profits are the same of firm $i+1$'s in the auction with reserve prices when firm i wins its auction (see (A.40)). We can write DB's profits as

$$\max_{d_{\mathrm{AU}}} \pi_{\mathrm{DB}}^{\mathrm{AU}} = \begin{cases} \frac{d_{\mathrm{AU}} t(8-3d_{\mathrm{AU}} n)}{12} & \text{for } \; d_{\mathrm{AU}} < \frac{3}{2n} \\ \frac{7t}{16n} & \text{for } \; d_{\mathrm{AU}} \geq \frac{3}{2n}. \end{cases} \tag{A.47}$$

When $d_{\mathrm{AU}} < \frac{3}{2n}$, DB's profits are maximized for $d_{\mathrm{AU}}^{*} = \frac{4}{3n}$. By replacing d_{AU}^{*} in (A.47), we find that the DB always opts for setting $d_{\mathrm{AU}}^{*} = \frac{4}{3n}$.

We now compare the DB's profits under the two strategies When the DB sells data to all entering firms and $n < \hat{n}$, the DB obtains $\pi_{\mathrm{DB}}^{\mathrm{TIOLI}^*}(\mathbf{P}_{\mathrm{TIOLI}}) = \frac{t}{4n}$. By directly comparing profits, we find that the DB prefers selling to half of the entering firms. When $n \geq \hat{n}$, the DB opts for selling to half of the entering firms if

$$\frac{4}{9} > \frac{1}{2} \frac{(1-2a_1)^2}{-8a_0^2 + a_0\left(8a_1 + 4\right) - 4a_1 + 1},$$

which is always satisfied for $n \geq 2$.

Firms remaining profits after entering and paying for data are equal to

$$\frac{t}{n^2} - \frac{2d^*_{\text{AU}}t}{3n} + \frac{t{d^*_{\text{AU}}}^2}{9} - F. \tag{A.48}$$

By solving the binding constraint (A.48) with respect to n gives us

$$n^*_{\text{AU}} = \frac{9\sqrt{tF} - 3d^*_{\text{AU}}t}{9F - t{d^*_{\text{AU}}}^2} = \frac{5}{9}\sqrt{\frac{t}{F}}, \tag{A.49}$$

leading to DB's profits being

$$\pi^{\text{AU}*}_{\text{DB}} = \frac{4}{5}\sqrt{tF}.$$

For consumer surplus, we can use (A.45), with $d^*_{\text{AU}} = \frac{4}{3n^*_{\text{AU}}}$ and $n^*_{\text{AU}} = \frac{5}{9}\sqrt{\frac{t}{F}}$. We obtain

$$CS^{\text{AU}} = u - \frac{341}{180}\sqrt{tF}.$$

Proof of Proposition 5. This proof proceeds in three steps: i) we solve the game when the DB sells data to all entering firms, which corresponds to the equilibrium under TIOLI, ii) we assess the DB's strategy under auction with reserve prices (AR), and iii) we assess the DB's strategy under auction without reserve prices (AU).

(i) Expressions of DB's profits when he sells to all entering firms are presented in the Proof of Proposition 1. First, we show that the strategies where the DB sets $d_1 \le d < \frac{3}{2n}$ are always dominated by the strategies where the DB sets $d \ge \frac{3}{2n}$. DB's profits are given by firms' willingness to pay times the number of entering firms. Under the alternative timing, the DB chooses his strategy by anticipating how d influences firm entry and, in turn, his profits. From the FOC of DB's profits, profits are decreasing in n. Firm entry is determined by firms' profits when losing, and is minimized for $d \ge \frac{3}{2n}$.[17] Thus, strategies where $d_1 \le d < \frac{3}{2n}$ can only be dominant if they allow to extract more surplus from individual firms. Suppose that n is given. Then, from the Proof of Proposition 1, we know that the DB prefers setting $d \ge \frac{3}{2n}$ instead of $d_1 \le d < \frac{3}{2n}$, as it allows him to extract more surplus from individual firms. Then, setting $d \ge \frac{3}{2n}$ dominates $d_1 \le d < \frac{3}{2n}$ even when the DB anticipates his effect on firm entry.

Second, we show that the strategy where the DB sets $d \ge \frac{3}{2n}$ always dominates the one where he sets $d \le d_1$. We refer to the DB setting $d \ge \frac{3}{2n}$ as d_{high}, whereas we refer to the DB setting $d < d_1$ as d_{low}. When the DB sets $d \ge \frac{3}{2n}$, he maximizes

$$\pi_{\text{DB}}({\mathbf{P}^*_{\mathbf{A}}}^{d_{\text{high}}}) = \frac{t}{4n} \quad s.t. \quad \frac{t}{4n^2} - F \ge 0. \tag{A.50}$$

[17] See also Figure 1 for a graphical solution.

By binding the constraint we obtain

$$n^*_{d_{\text{high}}} = \frac{1}{2}\sqrt{\frac{t}{F}}. \tag{A.51}$$

By replacing (A.51) in (A.50) we obtain

$$\pi_{\text{DB}}(\mathbf{P}^{*}_{\mathbf{A}}{}^{d_{\text{high}}}) = \frac{1}{2}\sqrt{tF}. \tag{A.52}$$

When the DB sets $d = d_{low}$, he maximizes

$$\begin{aligned}\pi_{\text{DB}}(\mathbf{P}^{*}_{\mathbf{A}}{}^{d_{\text{low}}}) &= t\left(d_{\text{low}}\left(1-2a_1\right) - n\frac{d^2_{\text{low}}}{2}\left(1+4\left(1-2a_0\right)\left(a_0-a_1\right)\right)\right)\\ s.t. &\quad \left(\frac{t}{n} - td_{\text{low}} + 2td_{\text{low}}a_o\right)\left(2d_{\text{low}}\left(a_1-a_0\right)+\frac{1}{n}\right) - F \geq 0. \end{aligned} \tag{A.53}$$

We want to show that

$$\pi_{\text{DB}}(\mathbf{P_A}{}^{d_{\text{low}}}) < \pi_{\text{DB}}(\mathbf{P_A}{}^{d_{\text{high}}}), \tag{A.54}$$

for all relevant values of d_{low}, t and F. In particular, we recall that $0 \leq d_{\text{low}} < \frac{1}{n}$, $t > 0$, $F > 0$, $t > F$. It is useful to express $\frac{F}{t} = k$, with $0 < k < 1$. We can rewrite (A.54) as

$$2d_{\text{low}}\left(1-2a_1\right) - nd^2_{\text{low}}\left(1+4\left(1-2a_0\right)\left(a_0-a_1\right)\right) < \sqrt{k}. \tag{A.55}$$

To solve (A.55), we bind the constraint in (**??**), find the number of entering firms and substitute it in (A.55). The constraint in (**??**) has no explicit solution. We thus want to find an approximated solution of n that overestimates the left-side of (A.55). By showing that the left side is smaller than the right side even after the round ups, we prove that also the original inequality holds.

The number of entering firms is given by binding the constraint in (**??**):

$$\pi_i{}^{\text{L}}(\mathbf{P_A}{}^{d_{\text{low}}}) = \left(\frac{t}{n} - td_{\text{low}} + 2td_{\text{low}}a_o\right)\left(2d_{\text{low}}\left(a_1-a_0\right)+\frac{1}{n}\right) - F = 0. \tag{A.56}$$

By substituting the explicit forms of a_0 and a_1 in (A.56), we can rewrite it as

$$\frac{1}{n^2} - \frac{2d_{\text{low}}}{\sqrt{3}n}f(n) + \frac{1}{3}d^2_{\text{low}}f(n)^2 - \frac{F}{t} = 0, \tag{A.57}$$

where

$$f(n) = \frac{\left(\sqrt{3}-1\right)\left(2+\sqrt{3}\right)^n + \left(1+\sqrt{3}\right)\left(2-\sqrt{3}\right)^n - 2\sqrt{3}}{\left(2+\sqrt{3}\right)^n + \left(2-\sqrt{3}\right)^n - 2}. \tag{A.58}$$

Although there is no explicit solution $n(d_{\text{low}})$, the expression is a second-order polynomial in d_{low}: then, we can obtain an explicit solution for $d(n)$. Solving (A.57) with respect to d we obtain

$$d_1^*(n) = \frac{\sqrt{3}\left(n\sqrt{k}+1\right)}{nf(n)} \quad \text{and} \quad d_2^*(n) = -\frac{\sqrt{3}\left(1-n\sqrt{k}\right)}{nf(n)}. \tag{A.59}$$

From Salop (1979) we know that, if a DB is absent, the number of entering firms is $n = \sqrt{\frac{t}{F}}$. As such, our solution must satisfy $d_{\text{low}}\left(\sqrt{\frac{t}{F}}\right) = 0$, which gives us $d_{\text{low}} = d_2^*(n)$. Having found $d_{\text{low}}(n)$, we need to invert the function to obtain $n(d_{\text{low}})$. To do so, we approximate $f(n)$ to find an explicit form of $n(d_{\text{low}})$. We recall that we want to round up $\pi_{\text{DB}}(\mathbf{P_A}^{d_{\text{low}}})$, which is inversely proportional to n. As such, we need to round down $n(d_{\text{low}})$, which requires rounding up $f(n)$. We find

$$f(n) \approx 0.6197\frac{1.0489n - 1.0566}{0.7806n - 0.4757}, \tag{A.60}$$

which overestimates $f(n)\ \forall\ n \geq 2$.

We can now substitute (A.59) in (A.57), obtaining

$$\begin{aligned} &n^2\left(0.6197 + 1.0489d_{\text{low}} + 0.7806\sqrt{3k}\right) \\ &\qquad - n\left(0.6197 * 1.0566d_{\text{low}} + 0.7806\sqrt{3} + 0.4757\sqrt{3k}\right) + 0.4757\sqrt{3} = 0, \end{aligned}$$

which has two solutions:

$$\begin{aligned} n\left(d_{\text{low}}\right) = &\frac{0.66d_{\text{low}} + 0.4757\sqrt{3k} + 0.7806\sqrt{3}}{1.3d + 1.56\sqrt{3k}} \\ &\pm \frac{1.31\sqrt{-0.277\sqrt{3}\left(2.6d_{\text{low}} + 3.12\sqrt{3k}\right) + \left(0.5d_{\text{low}} + 0.363\sqrt{3k} - 0.597\sqrt{3}\right)^2}}{2.6d_{\text{low}} + 1.56\sqrt{3k}}. \end{aligned}$$

Given that $n(0) = \sqrt{\frac{t}{F}} = \sqrt{\frac{1}{k}}$, the correct solution is the one with the positive sign. We have obtained $n\left(d_{\text{low}}\right)$ rounded down, which in turn rounds up $\pi_{\text{DB}}(\mathbf{P_A}^{d_{\text{low}}})$. Next, we round up the exponential terms present in the left side of (A.55) to increase it. We first focus on $(1-2a_1)$. This function is monotonically increasing in n, and its limit is

$$\lim_{n\to\infty}(1-2a_1) = 2 - \frac{2}{\sqrt{3}}.$$

As $(1-2a_1)$ increases the left side of (A.55), we approximate

$$(1-2a_1) \approx 2 - \frac{2}{\sqrt{3}}. \tag{A.61}$$

Next, we focus on $(1+4(1-2a_0)(a_0-a_1))$. This function is monotonically increasing in n, and decreases the left side of (A.55). We find

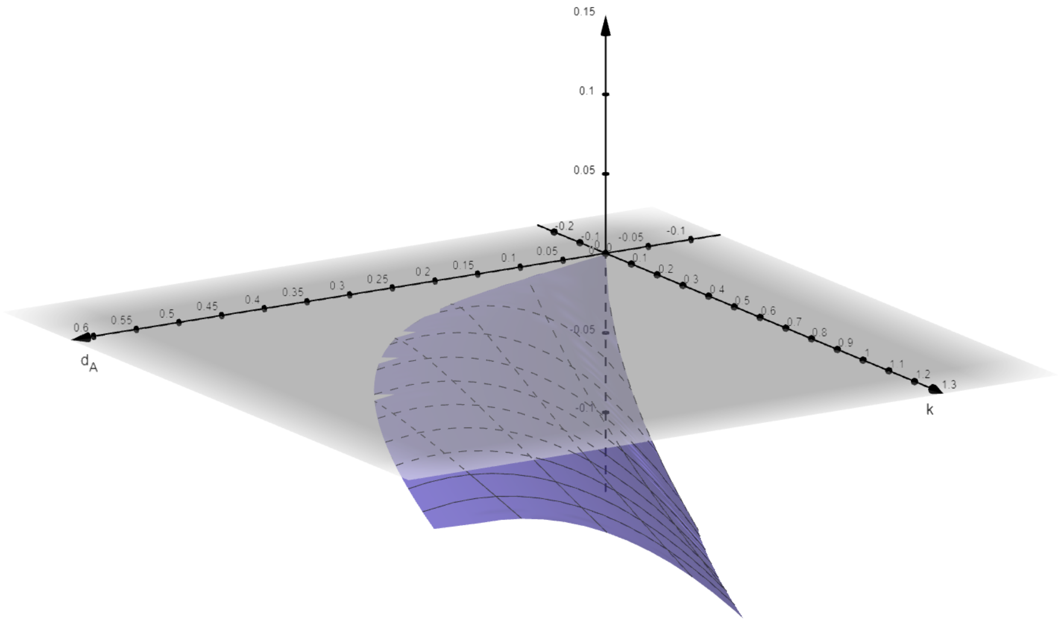
$$(1+4(1-2a_0)(a_0-a_1)) \approx 1.36\frac{8n-1}{8n+2}, \tag{A.62}$$

which underestimates the function and thus overestimates the left-side of (A.55). By replacing (A.61) and (A.62) in (A.55) and setting $n = n(d_{\text{low}})$ we obtain

$$\left(2 - \frac{2}{\sqrt{3}}\right) 2d_{\text{low}} - n(d_{\text{low}})\, d_{\text{low}}^2 \left(1.36 \frac{8n(d_{\text{low}}) - 1}{8n(d_{\text{low}}) + 2}\right) - \sqrt{k} < 0, \tag{A.63}$$

which only depends on d_{low} and k, with $0 < d_{\text{low}} < \frac{1}{n(d_{\text{low}})}$, and $0 < k < 1$. As such, we can plot the left-hand side of (A.63) for all the proper couples (d_{low}, k). As we can see in Figure A.1, the inequality always holds: thus, the DB's strategy is the same as the one adopted in the baseline model when $n < \hat{n}$, described in Proposition 1.

FIGURE A.1. Difference in DB's profits between selling d_{high} and d_{low}



This figure shows the left term of inequality (A.63) for $0 < d_{\text{low}} < \frac{1}{n(d_{\text{low}})}$ and $0 < k < 1$.

(ii) The difference between the original timing and this extension is that the DB's strategy is directly influenced by the number of entering firms, as he anticipates the effect of the data sale on it. As described in the previous step, DB's profits are decreasing in n. Since the DB's strategy under AR maximizes firms' willingness to pay and minimizes firm entry, The DB adopts the same strategy under the alternative timing.

(iii) Under AU, as already shown in the proof of Proposition 4, the DB offers same sized partitions in all the auctions. Without loss of generality, we focus our analysis on firm 0.

Firm 0's profits when winning and losing are the same as in in the proof of Proposition 4. Thus, DB's profits are

$$\max_{d_{\mathrm{AU}}} \pi_{\mathrm{DB}}^{\mathrm{AU}} = \begin{cases} \frac{d_{\mathrm{AU}} t(8-3d_{\mathrm{AU}} n)}{12} & \text{for } \; d_{\mathrm{AU}} < \frac{3}{2n} \\ \frac{7t}{16n} & \text{for } \; d_{\mathrm{AU}} \geq \frac{3}{2n}, \end{cases} \tag{A.64}$$

given that

$$\begin{cases} \frac{t}{n^2} - \frac{2d_{\mathrm{AU}} t}{3n} + \frac{\mathrm{t}d_{\mathrm{AU}}^2}{9} - F \geq 0 & \text{for } \; d_{\mathrm{AU}} < \frac{3}{2n} \\ \frac{t}{4n^2} - F \geq 0 & \text{for } \; d_{\mathrm{AU}} \geq \frac{3}{2n}. \end{cases} \tag{A.65}$$

When $d_{\mathrm{AU}} < \frac{3}{2n}$, we obtain the number of entering firms by binding the first part of the piecewise function (A.65), obtaining

$$n_{\mathrm{AU}}^{*} = \frac{9\sqrt{tF} - 3d_{\mathrm{AU}} t}{9F - td_{\mathrm{AU}}^2}. \tag{A.66}$$

When $d_{\mathrm{AU}} \geq \frac{3}{2n}$, the number of entering firms is constant and given by binding the second part of the piecewise function (A.65), obtaining

$$n_{\mathrm{AU}}^{*} = \frac{1}{2}\sqrt{\frac{t}{F}}. \tag{A.67}$$

By substituting (A.66) and (A.67) in (A.64), we obtain

$$\max_{d_{\mathrm{AU}}} \pi_{\mathrm{DB}}^{\mathrm{AU}} = \begin{cases} \frac{d_{\mathrm{AU}} t\left(72F + td_{\mathrm{AU}}^2 - 27d_{\mathrm{AU}}\sqrt{tF}\right)}{108F - 12td_{\mathrm{AU}}^2} & \text{for } \; d_{\mathrm{AU}} < \frac{3}{2n_{\mathrm{AU}}^{*}} \\ \frac{7}{8}\sqrt{tF} & \text{for } \; d_{\mathrm{AU}} \geq \frac{3}{2n_{\mathrm{AU}}^{*}}. \end{cases} \tag{A.68}$$

Computing FOCs of (A.68) for $d_{\mathrm{AU}} < \frac{3}{2n_{\mathrm{AU}}^{*}}$ with respect to d_{AU}, we find that DB's profits are monotonically increasing in d_{AU}. As such, the DB sets $d_{\mathrm{AU}}^{*} \geq \frac{3}{2n_{\mathrm{AU}}^{*}}$ and obtains profits

$$\pi_{\mathrm{DB}}^{\mathrm{AU}*} = \frac{7}{8}\sqrt{tF}. \tag{A.69}$$

As profits in (A.69) are higher than when selling data to all firms, in equilibrium the DB opts for selling to every other firm. To compute consumer surplus, we can use (**??**), as the difference between AR and AU with regard to consumer surplus only lies in the quantity of data sold. By replacing d_{AU}^{*} in (**??**), we obtain

$$CS^{\mathrm{AU}} = u - \frac{5t}{4n_{\mathrm{AU}}^{*}} + \frac{n_{\mathrm{AU}}^{*} td_{\mathrm{AU}}^{*2}}{9}. \tag{A.70}$$

where $n_{\mathrm{AU}}^{*} = \frac{1}{2}\sqrt{\frac{t}{F}}$ and $d_{\mathrm{AU}}^{*} = \frac{3}{2n_{\mathrm{AU}}^{*}}$, as it is the limit case after which data exhaust their marginal effect. We can rewrite (A.70) as

$$CS^{\mathrm{AU}} = u - \frac{t}{n_{\mathrm{AU}}^{*}} = u - 2\sqrt{tF}.$$

Odd number of entering firms

In this Appendix, we focus on the DB's equilibrium strategy under the selling mechanism AR when n is odd. In this case, selling data to every other firm results in at least two adjacent firms being both informed (or both uninformed).

While this problem shares some similarities with a Salop model with heterogeneous costs (see alderighi2012), in our setup the heterogeneity between firms also changes their outside option, thus preventing us from finding an explicit solution when n is odd.

Following Delbono et al. (2021), to address the case of an odd number of firms, we perform a numerical analysis that includes two specific DB's strategies, in addition to the sale of all firms. These two strategies do not emerge when the number of firms is even and are represented in Figure B.1. The first (in panel (a) of Figure B.1) is the sale of data only to even-indexed firms, and we denote this strategy as E. This strategy implies that the DB sells data to two adjacent firms, and to every other remaining firm. Specifically, in the figure, the DB sells data to firms 0, 2, 4 and 6. By doing so, the DB serves $\frac{n+1}{2}$ firms. The second (panel (b) of Figure B.1) entails the sale of data only to odd-indexed firms, and we denote this strategy as O. This strategy implies that the DB leaves two adjacent firms uninformed, and sells data to every other remaining firm. Specifically, in the figure, firms 1, 3 and 5 are informed. Note that, under these strategies, the data partitions sold in equilibrium are no longer symmetric, due to the asymmetric configuration of informed firms in the market.

FIGURE B.1. Asymmetric candidate equilibrium strategies when the number of entering firms is odd ($n = 7$).

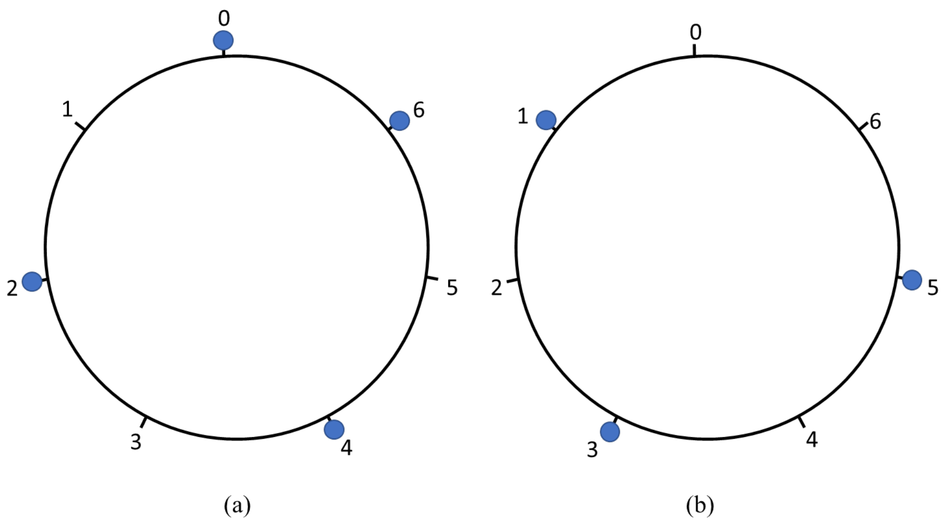
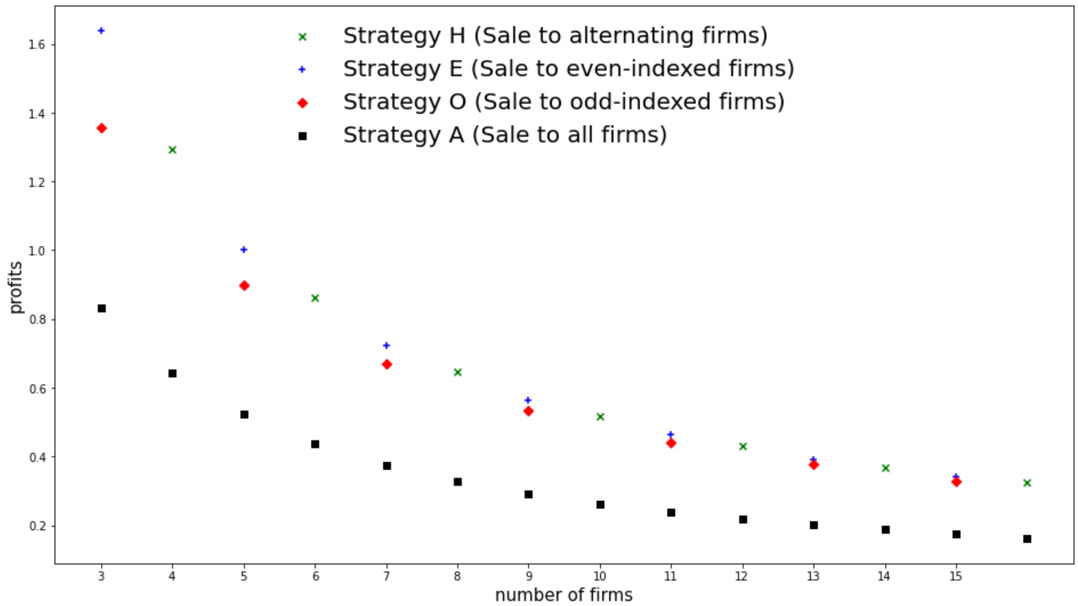


Figure B.2 shows the optimal DB's profits (on the vertical axis) for any number of entering firms (on the horizontal axis), under different configurations of informed firms. The blue points represent the DB's profit corresponding to his optimal strategy in the sub-game when the number of firms is odd, under strategy E. The red points represent the DB's maximum profit in the subgame in which the number of firms is odd under strategy O. We also report the strategies analyzed in our baseline model. Specifically, the green points are the DB's profits corresponding to his optimal strategy when the number of firms is even and he sells data to every other firm (strategy H). Finally, the black points represent the DB's profit under the sale to all firms (strategy A).

FIGURE B.2. DB's profits with odd number of firms. ($t = 10$)

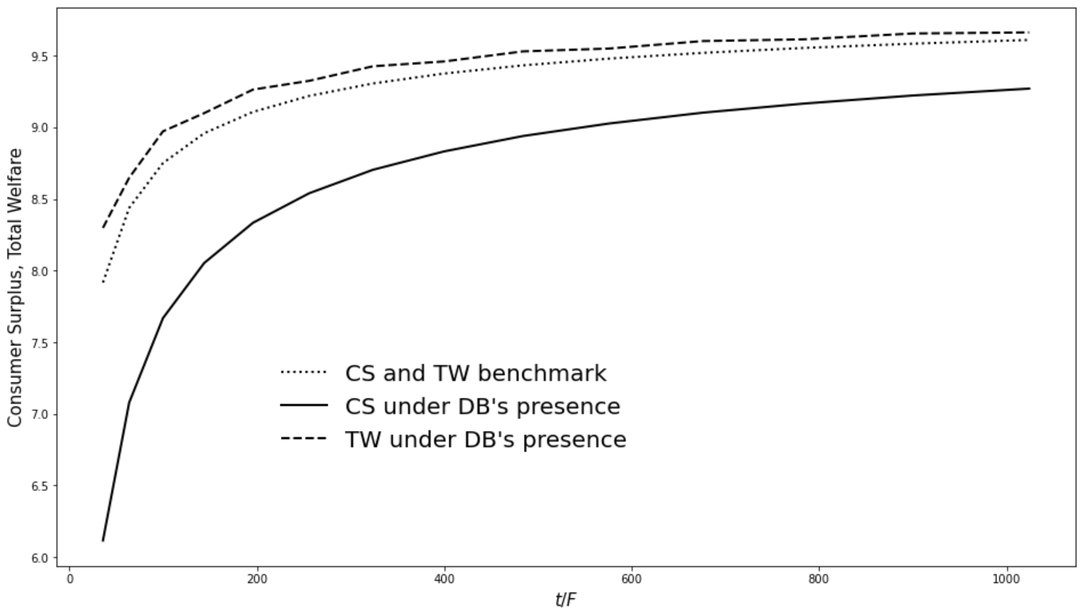


From the numerical analysis in Figure B.2, two results emerge. First, selling data to all firms is always a dominated strategy under the considered auction mechanism, regardless of whether the number of firms is even or odd. Second, when the number of entering firms is odd, the DB prefers strategy E over strategy O. Strategy E ensures that at least a portion of consumers is identified on every arch, which results in additional surplus extracted by firms that can then be captured by the DB.

We now assess the implications of strategy E on entry, consumer surplus and welfare. To this aim, note that under strategy E firms are heterogeneous in terms of equilibrium profits. To get an intuition in the most straightforward way, let us refer to the market configuration represented in panel (a) of Figure B.1. Firms 0 and 6 obtain higher equilibrium profits than firms 2 and 4, because the DB can threaten the latter with the sale of the whole dataset to their rivals on both sides. Conversely, the DB can threaten firms 0 and 6 with the sale of the whole dataset to their rivals on only one side. This implies that the free-entry condition is binding only for firms 2 and 4, which obtain equilibrium profits equal to $\frac{t}{4n^2} - F$. From this free-entry condition, we find that the optimal number of entering firms is equal to $n_E^* = \frac{1}{2}\sqrt{\frac{t}{F}}$. We thus conclude that the entry barrier effect identified under AR is unaffected when we remove the assumption of an even number of firms.

Although the quantity of data sold in equilibrium under strategy E for an odd number of firms is different from the quantity of data sold under strategy H for an even number of firms, the equal magnitude of the entry barrier effect implies that consumer surplus is lower and welfare is higher relative to the benchmark case of no data. Figure B.3 shows the levels of consumer surplus and welfare in function of t/F, from which we obtain the equilibrium number of firms $n^* = \frac{1}{2}\sqrt{\frac{t}{F}}$. When n^* is even, consumer surplus and welfare are obtained under the equilibrium strategy of selling data to every other firm. Conversely, when n^* is odd, consumer surplus and welfare are obtained under strategy E.

FIGURE B.3. Consumer surplus and welfare in function of t/F. ($t = 10$, $v = 10$)



CHAPTER 3

Data Broker competition and downstream market entry

Laura Abrardi, Carlo Cambini, Flavio Pino†

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.

†Politecnico di Torino, Department of Management, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy.

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 *à la Hotelling* (Montes et al., 2019; Bounie et al., 2021b, 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 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.

[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 (2009).

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

[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.

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.

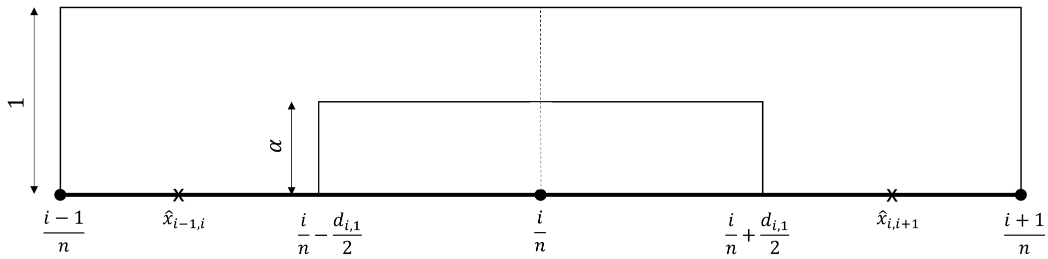


FIGURE 1. Firm i's market share when buying from DB_1. Firm i has a probability α of identifying consumers on the arch $[\frac{i}{n} - \frac{d_{i,1}}{2}, \frac{i}{n} + \frac{d_{i,1}}{2}]$ and offer them tailored prices, while it always offers his basic price on the consumers located on the arches $[\hat{x}_{i-1,i},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

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

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

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]

[5]The sequentially of Stages 4 and 5 is common in the literature (see Montes et al. (2019) and Bounie et al. (2021b, 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).

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 $p^B_{i,k}$ for unidentified consumers.

Stage 5. Firms that obtained a partition set tailored prices $p^T_{i,k}(x)$ 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 $\tilde{CS} = \tilde{TW} = v - \frac{5}{4}\sqrt{tF}$.

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 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

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

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 See Appendix A. $\square$

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:

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

Instead, DB_2 competes *à 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:

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

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. $\square$

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.

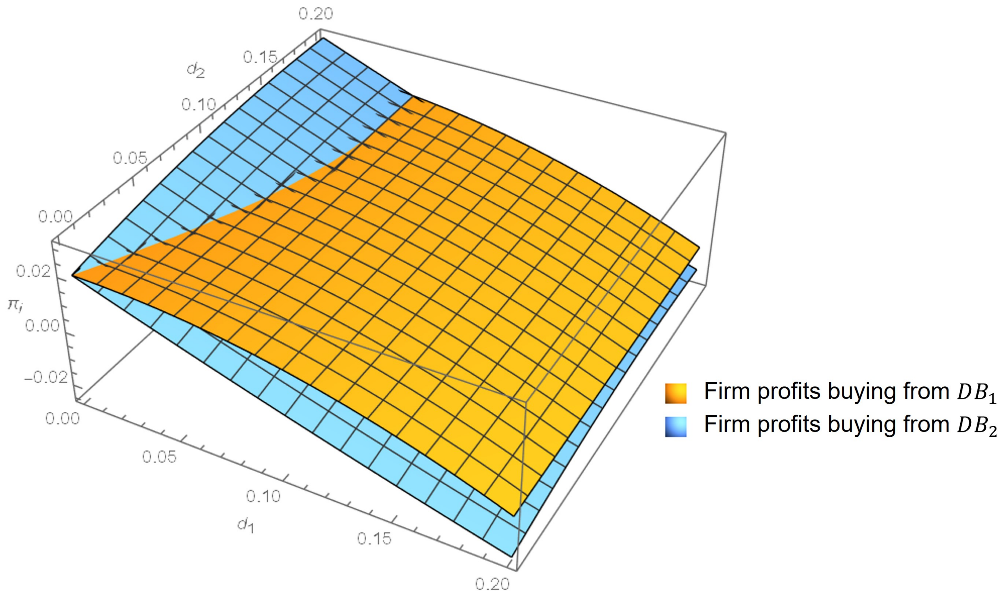


FIGURE 2. Firms' profits when buying and not buying data as a function of the partitions sold by DB_1 (d_1) and by DB_2 (d_2). $\alpha = 0.5, \beta = 0.5, t = 3, n = 5, F = 0.1$.

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]

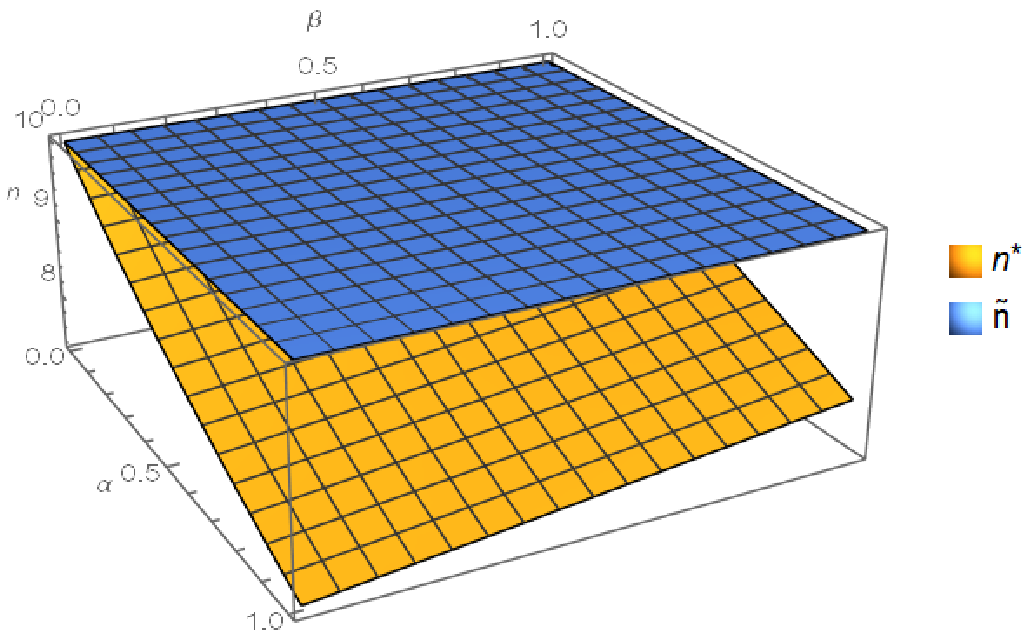


FIGURE 3. Number of entering firms in equilibrium and in the standard Salop model as a function of α and β. $u = 10, t = 10, F = 0.1$.

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.

[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.

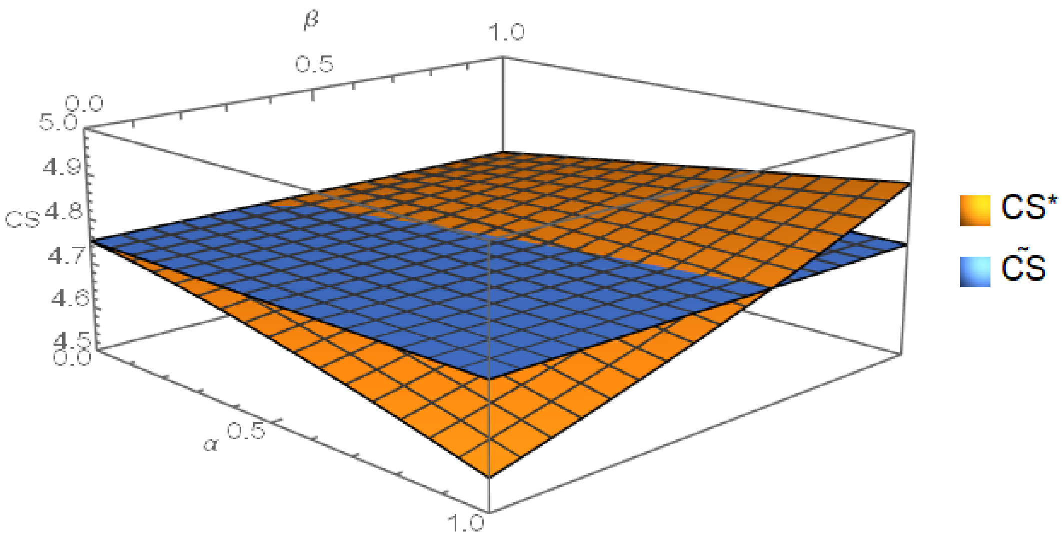


FIGURE 4. Consumer surplus in equilibrium and in the standard Salop model as a function of α and β. $u = 10, t = 10, F = 0.1$.

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. $\square$

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 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

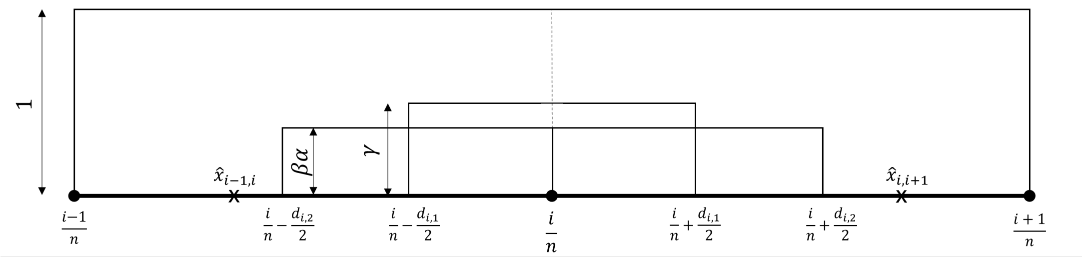


FIGURE 5. Firm i's market share when buying from both DB_1 and DB_2, assuming $d_2 > d_1$. Firm i has a probability γ of identifying consumers on the arch $[\frac{i}{n} - \frac{d_{i,1}}{2}, \frac{i}{n} + \frac{d_{i,1}}{2}]$ and offer them tailored prices, while it has a probability $\beta\alpha$ on consumers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

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 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. □

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. □

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.

Bibliography

Abrardi, L., Cambini, C., Congiu, R., and Pino, F. (2022). User data and endogenous entry in online markets. *Available at SSRN*, *4256544*.

Baik, S. A., Larson, N., et al. (2022). Price discrimination in the information age: Prices, poaching, and privacy with personalized targeted discounts. *The Review of Economic Studies*.

Belleflamme, P., Lam, W. M. W., and Vergote, W. (2020). Competitive imperfect price discrimination and market power. *Marketing Science*, *39*(5), 996–1015.

Bergemann, D., and Bonatti, A. (2011). Targeting in advertising markets: Implications for offline versus online media. *The RAND Journal of Economics*, *42*(3), 417–443. http://www.jstor.org/stable/23046807

Bergemann, D., and Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics*, *11*, 85–107.

Bounie, D., Dubus, A., and Waelbroeck, P. (2021a). *Competition and mergers with strategic data intermediaries*. https://ssrn.com/abstract=3918829

Bounie, D., Dubus, A., and Waelbroeck, P. (2021b). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, *52*, 283–313. https://doi.org/10.1111/1756-2171.12369

Chen, Z., Choe, C., and Matsushima, N. (2020). Competitive personalised pricing. *Management Science*, *66*(9), 4003–4023. https://doi.org/10.1287/mnsc.2019.3392

Esteves, R. B. (2009). *A survey on the economics of behaviour-based price discrimination* (tech. rep.). NIPE Working Papers 5/2009, NIPE - Universidade do Minho.

Fudenberg, D., and Villas-Boas, J. M. (2006). Behavior-based price discrimination and customer recognition. *Handbook on economics and information systems*, *1*, 377–436.

Goldfarb, A., and Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3–43. https://doi.org/10.1257/jel.20171452

Gu, Y., Madio, L., and Reggiani, C. (2022). Data brokers co-opetition. *Oxford Economic Papers*, *74*(3), 820–839.

Hagiu, A., and Wright, J. (2020). Data-enabled learning, network effects and competitive advantage. *Unpublished manuscript*.

Ichihashi, S. (2021). Competing data intermediaries. *The RAND Journal of Economics*, *52*, 515–537. https://doi.org/10.1111/1756-2171.12382

Iyer, G., Soberman, D., and Villas-Boas, J. M. (2005). The targeting of advertising. *Marketing Science*, *24*(3), 461–476. https://doi.org/10.1287/mksc.1050.0117

Liu, Q., and Serfes, K. (2004). Quality of information and oligopolistic price discrimination. *Journal of Economics Management Strategy*, *13*, 671–702. https://doi.org/10.1111/j.1430-9134.2004.00028.x

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*, 79–84.

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

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

Pino, F. (2022). The microeconomics of data–a survey. *Journal of Industrial and Business Economics*, *49*(3), 635–665.

Rhodes, A., and Zhou, J. (2021). *Personalized pricing and privacy choice* (tech. rep.). Working paper.

Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics*, *10*(1), 141–156.

Shaffer, G., and Zhang, Z. J. (1995). Competitive coupon targeting. *Marketing Science*, *14*(4), 395–416. http://www.jstor.org/stable/184137

Taylor, C. R., and Wagman, L. (2014). Consumer privacy in oligopolistic markets: Winners, losers, and welfare. *International Journal of Industrial Organization*, *34*, 80–84. https://doi.org/10.1016/j.ijindorg.2014.02.010

Thisse, J.-F., and Vives, X. (1988). On the strategic choice of spatial price policy. *The American Economic Review*, 122–137.

Vickrey, W. S. (1964). *Microstatics*. Harcourt.

Villas-Boas, J. M. (2004). Consumer learning, brand loyalty, and competition. *Marketing Science*, *23*(1), 134–145.

Appendix

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 **Searle1979invertingempty citation** 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

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

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

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

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

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

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

$$\pi^*_{i,1} - (\pi^*_{i,1} - \pi^D_{i,2}) = \pi^D_{i,2}.$$

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

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

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

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

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 (??)). 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

$$d^* = \frac{1 - 2a_1}{n} \tag{A.14}$$

and

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

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 (**??**) 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

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

We obtain TW by adding (A.18) with (**??**) 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 $TW^* = \tilde{TW}$ when $\alpha = \beta = 0$, and since $n^* \leq \tilde{n} \quad \forall \quad \alpha, \beta \in [0, 1]$ (as described in Proposition 2), $TW^* \geq \tilde{TW} \quad \forall \quad \alpha, \beta \in [0, 1]$.

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

[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 $d_{i,2} < d_{i,1}$.

$$
\pi_{i,12} = \gamma \int_{\frac{i}{n}-\frac{d_{i,1}}{2}}^{\frac{i}{n}+\frac{d_{i,1}}{2}} p^{\mathrm{T}}_{i,12}(x)\,dx + (1-\gamma)d_{i,1}p^{\mathrm{B}}_{i,12} +
$$

$$
\beta\alpha\left(\int_{\frac{i}{n}-\frac{d_{i,2}}{2}}^{\frac{i}{n}-\frac{d_{i,1}}{2}} p^{\mathrm{T}}_{i,12}(x)\,dx + \int_{\frac{i}{n}+\frac{d_{i,1}}{2}}^{\frac{i}{n}+\frac{d_{i,2}}{2}} p^{\mathrm{T}}_{i,12}(x)\,dx \right) + (1-\beta\alpha)(d_{i,2}-d_{i,1})p^{\mathrm{B}}_{i,12}
$$

$$
+ p^{\mathrm{B}}_{i,12}\left(\widehat{x}_{i,i+1} - \widehat{x}_{i-1,i} - d_{i,2}\right) - F, \quad \text{(A.20)}
$$

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

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

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

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

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:

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

$$
- 2ta_0\left(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})\right) + 2t\frac{a_0}{n}, \quad \text{(A.23)}
$$

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

$$
- 2ta_1\left(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})\right) + 2t\frac{a_1}{n}, \quad \text{(A.24)}
$$

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

$$
- 2ta_1\left(\frac{1}{n} - \gamma d_{i,1} - \beta\alpha(d_{i,2} - d_{i,1})\right) + 2t\frac{a_1}{n}, \quad \text{(A.25)}
$$

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

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

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

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

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

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

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

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

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

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

$$\pi^*_{i,2} = \frac{t}{n^2} - \frac{t}{2}\bigg(4(-1+2a_0)(a_0 - a_1)(\gamma - \beta\alpha)^2 d_1^2 + \beta\alpha d_2^2\bigg) - \frac{td_1(\beta\alpha - \gamma)(-1+2a_1)}{n} - F. \quad (40)$$

As described in Lemma 3, DBs set their prices as

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

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

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

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

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

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

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

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

Following the same approach as in Proposition 3, we obtain

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

By replacing n^{*sub} in (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 $\gamma > \bar{\gamma}$, $CS^{*sub} < \tilde{CS}$.

CHAPTER 4

Mandated data sharing in hybrid marketplaces

Federico Navarra*, Flavio Pino†, Luca Sandrini‡

Considering a monopolistic hybrid platform, we investigate the effect of a mandated data sharing policy on market outcomes. We show that the hybrid platform can use the per-transaction fee to effectively avoid direct competition with sellers, resulting in markets that are fully covered but where sellers opt to price as monopolists. When data sharing is mandated, the platform adjusts the fee so that he still avoids competition with sellers. However, as sellers can also price discriminate, consumers are ultimately harmed as both the platform and sellers can extract all of their surplus. We argue that in markets where competition is softer, mandatory data sharing may damage the very agents it is intended to protect.

*Department of Economics and Management, University of Padova (Italy)

†Politecnico di Torino, Department of Management, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy.

‡Quantitative Social and Management Sciences, Faculty of Economics and Social Sciences, Budapest University of Technology and Economics, Budapest (Hungary).

1. Introduction

The relevance of data in modern economies has constantly been increasing during the past years, primarily due to the importance of digital markets. Every interaction on digital platforms and websites is tracked and registered by companies, and large amounts of data are traded every moment. Recent estimates from the European Commission show that:

> "The value of the data economy for the EU27 has been estimated to have reached almost 400 billion in 2019 and 440 billion in 2021, with a year-on-year growth rate of 4.9% in 2021. The estimated share of overall impacts on GDP in the EU27 ranges from 3.1% in 2019 to 3.6% in 2021" (DATA Market Study 2021–2023, pg. 116)."[4]

Data is a core input factor for production processes, logistics, targeted marketing, smart products, and services. Moreover, they are fundamental to training Artificial Intelligence and refining algorithms. On top of that, data drive interoperability in interconnected environments and are expected to drastically impact specific sectors such as mobility and healthcare.

Data owners have a large competitive advantage over their market rivals. Hence, data are very relevant to competition and privacy authorities. Digital platforms may have the incentive to adopt potential anti-competitive practices, such as self-preferencing (Padilla et al., 2022) and bias-recommendation (Bourreau and Gaudin, 2022), or, more generally, they may abuse their data-driven dominant position.

To stay competitive, firms competing against digital platforms increasingly depend on timely access to relevant data and their ability to use those to develop innovative applications, services, and products. For these reasons, a widespread debate has emerged on whether – and under which conditions and legal bases – public intervention is required to ensure adequate and timely access to data. One of the proposed remedies is to mandate platforms to share with sellers and rival companies all or part of the consumers' data they possess. Data sharing is one of the pillars of the European strategy for data.[5] It is at the

[4] Available at https://digital-strategy.ec.europa.eu/en/library/results-new-european-data-market-study-2021-2023.

[5] Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN.

core of the Digital Market Act (DMA hereafter), the recently introduced EU competition law regulating large digital platforms' (gatekeepers) business conduct.[6]

This paper analyses the strategic interactions between a monopolistic platform and the many sellers operating within the digital marketplace. We model each market as a Hotelling line, with a seller located at one end and the platform that can choose to enter at the other end. In particular, we investigate the effects of mandated data sharing on market outcomes and social welfare when data can be used to price discriminate consumers. We focus on a setting where a digital platform mediates between many sellers and consumers. Sellers must pay a per-transaction fee to the platform to be included in the marketplace, whereas consumers do not pay any admission fee. The platform decides, in order, i) the size of the fee, ii) whether or not to directly produce some (or all) final goods and compete against the sellers, and iii), in case of entry, the price of each good it produces. This setup is consistent, as an example, with Amazon's product groups' referral fees. Amazon subdivides its marketplace into broad product groups such as "baby products" or "clothing and accessories". While these categories contain many sub-markets, Amazon sets a single per-transaction fee for every product group. Moreover, we assume the platform has an ex-ante data advantage against the sellers, meaning that it can use the data it owns exclusively. Thus, Mandated data sharing alters the interaction between the platform and sellers.

Our main result shows that mandating data sharing can have unintended consequences in hybrid marketplaces and could even lead to consumer harm. It is indeed true that under data sharing, sellers are better able to compete against the platform. However, the platform can keep the increase in competition under control by increasing the per-transaction fee that sellers pay to the platform. In particular, we show that the platform sets a high enough fee so that the sellers opt to act as monopolists, even if the markets are fully covered. This scenario, previously explored by Thépot (2007) and Bacchiega et al. (2021) is usually referred to as *monopolistic duopoly*, and also occurs in the standard Hotelling model. Indeed, by relaxing the standard full market coverage assumption, there exists an

[6]In particular, article 6 of the DMA states that: *"The gatekeeper shall provide business users and third parties authorized by a business user, at their request, free of charge, with effective, high-quality, continuous and real-time access to, and use of, aggregated and non-aggregated data, including personal data, that is provided for or generated in the context of the use of the relevant core platform services or services provided together with, or in support of, the relevant core platform services by those business users and the end users engaging with the products or services provided by those business users."* (DMA, Art. 6.10), available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R1925&from=EN.

interval of parameters such that the firms set prices so that the indifferent consumer is also indifferent between buying or not buying the product. When both the sellers and the platform can price discriminate, a monopolistic duopoly is particularly harmful to consumers, as both actors can perfectly extract surplus through tailored prices while avoiding price wars. Furthermore, the platform would actually be better off under mandated data sharing if he could change the per-transaction fee after the policy implementation. We thus argue that mandated data sharing can have unintended consequences that ultimately harm the consumers that the policy is supposed to protect.

Our second result pertains to the pricing strategies adopted by the platforms and the sellers. We identify conditions under which the platform has the incentive to give up competing in some market segments. In particular, the platform decides not to send any offer to some consumers as it would be less profitable than collecting the fee from the sellers. This result stems from the vertically integrated (i.e., hybrid) nature of the platform, which operates as an intermediary and a rival in the market. Earning revenues in both cases, the platform exploits its advantage and sets a per-transaction fee to lower the efficiency of rival sellers. Then, depending on whether the competitive price guarantees higher or lower net revenues than the transaction fee, the platform decides whether to win or lose the competition.

The paper is organized as follows. In the next section, we provide an overview of the existing literature on the topic and illustrate our main contributions. Section 3 describes the model, and in Section 3.1, we solve it without and with mandated data sharing. Finally, in Section 4, we compare our results with those of the related literature and draw concluding remarks.

2. Related literature

This paper contributes to four main strands of literature. The first one focuses on the effect of consumer data in digital economics. The use of data is widespread across every sector, thanks to their versatility; typical uses include improving products or services' quality and efficiency, personalization, matching, and discriminating between different consumer groups or individuals. Recent surveys (Goldfarb and Tucker, 2019; Bergemann and Bonatti, 2019; Pino, 2022) have thus focused on categorizing both the types and uses of data, trying to extract broader insights that hold across different models. Two typical data functions are those of allowing price discrimination on consumers and increasing the vertical differentiation between firms (either allowing for an improvement of the products or a reduction in their marginal cost of production).

Price discrimination has been observed in various markets: a typical example involves the use of geolocalization to tailor prices to different consumers (Mikians et al. (2012); Aparicio et al. (2021)). The literature has mostly focused on competition between informed firms, stemming from Thisse and Vives (1988) seminal work, and on the vertical relations between firms and a data seller (Montes et al. (2019); Bounie et al. (2021); Delbono et al. (2021); Abrardi et al. (2022)). The common insight of these models is that allowing all firms to obtain data benefits consumers while it harms both the firms and the data seller. Our main contribution highlights how, when vertical integration is introduced, mandating data sharing can benefit all actors, as firms can retain more profits due to their increase in competitiveness while the platform benefits from the overall increase in market efficiency.

The second strand of literature focuses instead on information sharing. Information sharing has been extensively studied in the literature: Raith (1996) describes a general model that summarizes many existing models to show the determinants of when and how firms are incentivized to disclose private information. The recent literature on digital economics is also gaining interest with regard to information sharing, with particular attention on consumer data. Krämer and Schnurr (2022) focus on market contestability with regard to data-rich incumbents and explore the possible effects of policy interventions such as data siloing, data sharing, and data portability. Regarding e-commerce, they stress the importance of sellers' data portability, as this policy would allow sellers to grow without having to lock in on a specific platform. Prüfer and Schottmüller (2021) study competition in a data-driven market where data reduce the cost of quality production. Their model shows how mandated data sharing does not reduce the dominant firm's incentive to innovate and also eliminates the risk of market tipping. Krämer and Shekhar (2022) expand on this topic by analyzing how the aforementioned policy interventions impact competition between platforms, modeling the effect of data as an improvement in the user experience on the platforms and allowing platforms to compete as well as set their investment levels. In particular, they show that mandated data sharing can reduce innovation investment by platforms, which in turn can hurt consumers when data externalities are large. De Corniere and Taylor (2020) analyze the effects of data sharing by using a competition-in-utilities approach, finding sufficient conditions under which data sharing would be unambiguously pro-competitive. Liu et al. (2021) focus instead on a retail platform that hosts sellers and can strategically disclose information to them: they find that the platform has the incentive to disclose information only to a subgroup of sellers. While our work focuses on an exogenous shock that mandates complete information sharing instead of allowing for a strategic decision by the platform, as far as we know, we are the first to allow the platform to vertically integrate, entering the downstream market and competing with sellers.

Related to the vertical integration aspect of our model, the third strand of literature concerns the classical questions regarding access pricing and sabotage. Indeed, our model resembles the typical setup of an upstream monopolist that controls infrastructure and can choose to integrate downstream. Economides (1998) shows how an integrated monopolist has the incentive to degrade the quality of the downstream input to raise the costs of its rivals until they are driven out of the market. Beard et al. (2001) expand on this topic by showing that the upstream monopolist is always willing to expand downstream but has the incentive to sabotage only when input price regulation is introduced. Our model presents a similar result: the platform can strategically use the per-transaction fee to increase downstream costs, allowing it to compete better against sellers. Moreover, the platform sets the fee such that sellers opt to price as monopolists, increasing surplus extraction from consumers and, in turn, the platform's profits.

Finally, the fourth strand of literature concerns hybrid marketplaces – platforms that allow transactions between sellers and buyers and where the platform can become a seller's competitor. This literature is becoming pivotal in policy discussion, as tech giants such as Amazon and Apple are themselves hybrid marketplaces. Empirical evidence suggests that the downstream entry of the platform, sometimes referred to as *dual mode*, usually takes place in successful markets and leads sellers to reduce their growth efforts in the platform (Zhu and Liu (2018)). Moreover, sellers tend to increase the prices in the markets where the platform enters while shifting their innovation investments elsewhere (Wen and Zhu (2019)). In particular, evidence suggests that complementary goods become the focus of innovation, as the platform entry usually expands the demand for that good (Foerderer et al. (2018)). From a theoretical perspective, the effects of a platform operating in dual mode are ambiguous. On the one hand, platform downstream entry could reduce sellers' market power and increase competition, benefiting consumers (Dryden et al. (2020); Etro (2021a)). Platform entry could also induce it to reduce its commission fees to further expand the market's reach (Etro (2021b)). On the other hand, a higher quality (or lower cost) of the platform's goods can incentivize it to increase its commission fees, ultimately harming consumers (Anderson and Bedre-Defolie (2021)). While our model assumes the presence of a hybrid marketplace, focusing instead on mandated data sharing, our results show that an increase in vertical differentiation by the platform leads indeed to an increase in the commission fee, which in turn can crowd out sellers.

3. Model set-up

Consider a digital marketplace owned by a platform. The marketplace groups together a unit mass of markets denoted by $\mathcal{I}$. In each market, $i \in \mathcal{I}$, a seller (s^i) and the platform (p) compete in prices for horizontally differentiated goods. In what follows, we will sometimes refer to the seller and the platform together as to *firms*.

We assume that in each market, there is a continuum of consumers uniformly distributed on the $[0-1]$ Hotelling line. They consume at most one unit of either the good sold by the seller or the one sold by the platform. A consumer (she/her) located in $x \in [0,1]$ derives constant utility $u > 1$ from consuming either of the two goods and pays the price p_k, where $k = s, p$. Also, she suffers a mismatch disutility $t|z_k - x|$ from consuming a variety that is not her favorite one, where $t > 0$ is the transportation cost, and z_k is the location of the variety consumed ($z_s = 0$ and $z_p = 1$). Throughout the model, we assume that t is sufficiently large so that no firm can, under any circumstance, cover the entire market alone. In each market i, the consumers' utility functions are:

$$U_s^i = u - p_s^i - t\,x \tag{1}$$

$$U_p^i = u - p_p^i - t(1-x) \tag{2}$$

$$U_{no}^i = 0 \tag{3}$$

where the subscript $_{no}$ stands for *no consumption*.

In any market $i \in \mathcal{I}$, the seller (it) and the platform (he/him) compete in prices and sell two varieties of one good. We assume production does not involve any fixed cost, but the seller active in i produces at a marginal cost $c_s^i \in [0,1]$. Instead, the platform produces all final goods at the same constant marginal cost $c_p \in [0,1]$. We assume that sellers are heterogeneous in their costs of production ($c_s^i \neq c_s^{-i}$). Moreover, c_s^i is uniformly distributed between 0 and 1, and each technology is allocated to only one seller. In other words, we can map each technology c_s^i to a market i.[7]

In addition to the marginal costs of production, all sellers have to pay the same per-transaction fee $f > 0$ to the platform in order to sell on the marketplace. Consequently, the payoffs of the two firms are:

$$\pi_s^i = D_s^i(p_s^i - c_s^i - f), \qquad \pi_p^i = D_p^i(p_p^i - c_p) + D_s^i f \tag{4}$$

where D_k^i indicates the demand of each firm $k = s, p$ in market i and includes all consumers who derive larger utility from consuming the good produced by k than by $-k$ or than not consuming at all. Notice that the platform earns revenues from per-transaction fees paid by the seller.

In each market, the seller must earn net revenues to stay active. It must set a price that is at least as high as the marginal costs of production, which are determined by the cost parameter c_s^i and, crucially, by the fee f set by the platform. By adjusting f, the platform can alter the price of the seller and the market demands, which are determined by the locations of the indifferent consumers.

Finally, we assume that the platform has access to information that allows him to operate first-degree price discrimination on all consumers. If data sharing is mandated, then sellers can also set tailored prices for individual consumers. We also assume that each consumer only observes one price from each firm: in other words, a consumer who observes a tailored price from a firm cannot also observe the uniform price.[8]

Market configurations. Three possible market configurations may emerge. First, the locations of the consumers indifferent between buying from either the platform or the seller and not buying at all are such that $\tilde{x}_{s,no} > \tilde{x}_{p,no}$. In this case, keeping in mind that the location of the two firms are $z_s = 0$ and $z_p = 1$, there exists a consumer $\tilde{x}_{s,p} \in (\tilde{x}_{s,no}, \tilde{x}_{p,no})$ who is indifferent between buying from the seller or the platform and derives positive utility in both cases. This is the standard Hotelling duopoly (hd) case

[7]This analysis allows us to explore both cases where the platform is less or more efficient than the sellers.

[8]This assumption resembles, for example, a consumer purchasing a good through an online portal. If the consumer is identified, the portal can then directly show her the tailored price, rendering the uniform price unavailable. Moreover, suppose instead that uniform prices are always observable: then, the tailored prices would only be attractive if they are lower than the uniform prices, i.e., tailored discounts. Recent work by Baik, Larson et al. (2022) has shown that in such a scenario, firms would set higher uniform prices to also increase the tailored prices they offer, leading to the same profits in equilibrium.

with full market coverage. Firms compete in the product market, and prices are strategic complements.

Second, the locations of the indifferent consumers are such that $\tilde{x}_{s,no} < \tilde{x}_{p,no}$. In this case, consumers located in $x \in (\tilde{x}_{s,no}, \tilde{x}_{p,no})$ prefer not buying at all, and the market is not covered. This is the local monopolies scenario (lm), in which firms do not compete against each other and set prices to maximize profits on their own turf.

Finally, the locations of the consumers indifferent between buying from firm i and not buying at all are such that $\tilde{x}_{s,no} = \tilde{x}_{p,no} = \tilde{x}_{s,p}$. In other words, there exists a range of values of v such that the platform and the seller can achieve higher profits by pricing like monopolists while the market is fully covered. This scenario is referred to in the literature as *monopolistic duopoly* (md) case (Thépot, 2007; Bacchiega et al., 2021). Prices are strategic substitutes, and firms adjust them strategically to ensure the market is just covered. Differently from the Hotelling duopoly market configuration, in this one, the indifferent consumer derives zero utility from consuming either good.

In the solution of the game, we will highlight the different strategies that emerge under these three market configurations and for which ranges of parameters such configurations exist.

Timing. The timing of the game is the following: $t = 0$) the policy maker introduces a data-sharing policy. If there is no data-sharing, the platform uses data exclusively. Otherwise, all sellers can also use data. $t = 1$), the platform sets a single per-unit linear fee which is the same in all markets. All sellers have to pay it to be allowed to sell their goods on the marketplace. $t = 2$) Given the fee f, sellers and the platform set prices simultaneously in each market. If data allow the data owners to price discriminate consumers, as they know their exact locations on the Hotelling line, we model price competition as in Montes et al. (2019), and Bounie et al. (2021), among others.[9] $t = 3$) Consumers observe the prices and decide if and what they consume. The solution concept is Subgame Perfect Nash Equilibrium, and the game is solved by backward induction.

For the sake of clarity, in what follows, we will omit the apex ^{i} when doing so does not create confusion.

[9]When the platform holds information about consumers' location, but the sellers don't, a well-known problem is the existence of a pure strategy Bertrand-Nash equilibrium (see Rhodes and Zhou, 2022, p.25). In order to ensure equilibrium existence, we assume that personalized price schedules are only set after uniform prices are set. Consistently, Amazon allegedly shows higher prices to Amazon Prime subscribers but compensates them with discounted services such as free shipping. See https://www.consumeraffairs.com/news/lawsuit-alleges-amazon-charges-prime-members-for-free-shipping-031414.html

3.1. Platform data advantage - No data sharing

Consider any given market i. Data ownership allows the platform to operate first-degree price discrimination on all consumers. The price set by the uninformed seller is uniform for all consumers, while the platform can offer tailored prices to each consumer. Recall that the platform sets his tailored prices as a function of the seller's uniform price when it competes with the seller or can extract all surplus from uncontested consumers. Thus, we can write the platform's prices as

$$p_p^{TC}(x) = p_s - t + 2tx, \qquad p_p^{TM}(x) = u + tx - t. \tag{5}$$

The apex ^{TC} stands for *Tailored under Competition*, suggesting a price that makes the consumer indifferent between buying from p or s. Instead, the apex ^{TM} indicates a price *Tailored under Monopoly* that makes the consumer indifferent between buying from the platform or not buying at all. To complete the analysis, we must find the seller's equilibrium price for any value of f. Moreover, for any level of f, we need to compare two scenarios: one where the seller sets its uniform price as a monopolist and one where it sets its price as if it was under direct competition with the platform. For simplicity, we drop the superscript i from the seller's functions, as we are focusing on a single market. Moreover, for better readability, the superscripts indicate the level of market coverage (hd - hotelling duopoly; dm - duopolistic monopoly; lm - local monopoly, respectively), while the subscript indicates the seller's pricing strategy (c - competitive; m - monopoly respectively). The step-by-step resolution is relegated to the Appendix. In the following paragraphs, we provide the intuitions behind the seller's and platform's strategies under the different scenarios.

Seller prices as a monopolist. Suppose that the seller sets its price as a monopolist and that f is high such that partial market coverage ensues.[10] Then, the seller's profits can be written as

$$\pi_{s_m}^{lm} = (p_{s_m}^{lm} - c_s - f) * \tilde{x}_{s,no}, \tag{6}$$

where $\tilde{x}_{s,no}$ is obtained by equating to zero the first equation in 1. By maximizing with respect to p_{sm}^{lm}, we obtain

$$p_{s_m}^{lm*} = \frac{c_s + f + u}{2}; \qquad \tilde{x}_{s,no}^{*} = \frac{u - c_s - f}{2t}; \qquad \pi_{s_m}^{lm*} = \frac{(c_s + f - u)^2}{4t} \tag{7}$$

[10]Recall that we assume that the transportation cost is high enough so that neither the seller nor the platform can fully cover the market on their own. Then, a high enough fee would limit the seller's market share and, in turn, lead to partial market coverage.

With regard to the platform, as we are under partial market coverage, his profits are equal to

$$\int_{\tilde{x}_{p,no}}^{1} p_p^{TM}(x) - c_p\, dx + D_s^i\, f, \tag{8}$$

which leads to

$$\tilde{x}^*_{p,no} = \frac{c_p + t - u}{t}; \qquad \pi^{lm*}_{p_m} = \frac{c_p^2 - f(c_s + f) - 2c_p u + u(f+u)}{2t}. \tag{9}$$

First, from the seller's equilibrium price in (7), it is clear that the seller cannot profitably serve any consumer if $f > u - c_s$. Thus, after this threshold, the seller would not stay in the market.

Second, by imposing $Tildex^*_{s,no} = \tilde{x}^*_{p,no}$ and solving for f, we can find the lower bound after which the market is fully covered. We obtain

$$f^{lm}_m = 3u - 2t - c_s - 2c_p. \tag{10}$$

When $f \leq f^{lm}_m$, market shares overlap, implying that $\tilde{x}_{p,no} < \tilde{x}_{s,no}$. Suppose that the seller continues to price as a monopolist. Instead, the platform must adjust his pricing strategy, as the consumers located in $[\ Tildex_{p,no},\ Tildex_{s,no}]$ are now contested by the seller. Then, the platform offers the price $p_p^{TM}(x)$ to consumers in $(\tilde{x}_{s,no}, 1]$, and the price $p_p^{TC}(x)$ to consumers located in $[\tilde{x}_{p,no}, \tilde{x}_{s,no}]$. By solving the model, we obtain that the indifferent consumer is located in

$$\tilde{x}^*_{s,p} = \frac{2c_p - c_s - f + 2t - u}{4t}, \tag{11}$$

which results in profits being equal to

$$\begin{aligned} \pi^{dm*}_{s_m} &= \frac{(c_s + f - u)(-2c_p + c_s + f - 2t + u)}{8t}; \\ \pi^{dm*}_{p_m} &= \frac{4c_p^2 - c_s^2 - 6c_s f - 5f^2 - 4tc_s + 4tf - 4t^2 + 2u(3c_s + f + 6t) - u^2 - 4c_p(c_s - f + 2t + u)}{16t}. \end{aligned} \tag{12}$$

Intuitively, lowering f allows the seller to decrease its uniform price, increasing the competition intensity with the platform. As the seller is never able to fully cover the market on its own. the profits above hold for any $f \in [0, f^{lm}_m]$.

Seller prices as under competition. Next, suppose instead that the seller sets its uniform price as if it always is under direct competition with the platform. For ease of exposition, assume that at first f is low enough that the standard Hotelling duopoly configuration ensues. Both the seller and the platform set their uniform prices simultaneously, and then the platform sets his tailored prices as a function of the seller's uniform price. By standard computations, we obtain equilibrium prices

$$p_{s_c}^{hd} = \frac{c_s + c_p + 2f + t}{2}; \qquad p_{p_c}^{hd} = cp + f. \tag{13}$$

As we can see, the platform sets his uniform price as low as possible. In particular, he sets it to recover the marginal cost of production and the fee that he would lose when poaching a consumer from the seller. In equilibrium, all the consumers the platform serves will purchase through tailored prices, which are equal to

$$p_{p_c}^{TC}(x) = \frac{c_s + c_p + 2f - t + 4tx}{2}. \tag{14}$$

These prices lead to the indifferent consumer being located in

$$\tilde{x}_{s,p}^{*} = \frac{c_p - c_s + t}{4t}, \tag{15}$$

which results in equilibrium profits

$$\pi_{s_c}^{hd*} = \frac{(c_p - c_s + t)^2}{8t}; \qquad \pi_{p_c}^{hd*} = \frac{(c_p - c_s)^2 + 2t(8f + 3(c_s - c_p)) + 9t^2}{16t}. \tag{16}$$

From the platform's equilibrium profits in (16), it is clear that the platform's profits are monotonically increasing in f. The intuition is that, as long as the market is covered, both the platform and the seller directly pass the fee onto consumers. Thus, the results in (16) hold as long as all consumers have a net utility after purchasing the good that is at least zero.

By analyzing consumers' net utility, we find that the consumer with the lowest net utility is located in $x = 1$. Indeed, such a consumer has a strong preference for the platform's product, and the platform can, in turn, extract most of her surplus. Thus, after a first threshold of f, the platform must adjust the tailored price he offers to consumers located near his location in order to ensure non-negative utility. However, this change in strategy only pertains to the platform's choice of tailored prices and does not influence the uniform prices choice or the location of the indifferent consumer.

However, as f increases further, the indifferent consumer $Tildex_{s,p}^{*}$ will also obtain a net utility equal to zero. After this threshold, duopolistic monopoly ensues: the seller loses market share, as it cannot profitably serve consumers too far from its location, whereas the platform can poach them through its tailored prices. As f continues to increase, we reach the point where $\tilde{x}_{s,no} = \tilde{x}_{p,no}$. After this threshold, the consumers lost by the seller are located too far from the platform to be profitably poached by the platform, and the market configuration becomes that of local monopolies. The platform extracts all available surplus from his consumers while the seller continues to price competitively by construction.

Finally, for extremely high values of f, the seller is not able to profitably serve any consumers and thus leaves the market.

Equilibrium strategy. In the previous paragraphs, we have analyzed the seller's and platform's equilibrium strategies under competitive and monopolistic strategies. However, for any level of f, the seller will either prefer the former or the latter. In the following Proposition, we describe the strategy chosen by the seller as a function of f.

PROPOSITION 1. *There exists a threshold $Barf$ such that, if $f < \bar{f}$, the seller sets its price competitively, and the market configuration is that of Hotelling duopoly. Otherwise, the seller sets its price as a monopolist, and the market configuration is either duopolistic monopoly or local monopolies.*

PROOF. See Appendix. □

The intuition behind the Proposition is that if f is low enough, the seller has a more leveled playing field when competing against the platform. Then, pricing competitively allows the seller to obtain a larger market share and, in turn, higher profits. Conversely, if the fee is too high, the seller is better off completely ignoring the platform and only focusing on consumers close to its location. Moreover, as the equilibrium market configuration is guided by the seller's pricing choice, the platform's profits present a discontinuity for $f = \bar{f}$.

Having obtained the equilibrium strategy for a given market i, we can now proceed to compute the total platform's profits across all markets, which are equal to

$$\Pi_p = \int_0^1 \pi_p^i(f)\, dc_s^i, \tag{17}$$

which the platform maximizes with respect to f. We find the following:

PROPOSITION 2. *Assume there is no data sharing. The platform sets a fee*

$$f^* = \frac{2(u + 2(t + c_p))) - 3}{10}$$

such that the equilibrium market configuration is monopolistic duopoly (md) in every market i. The prices are:

$$p_s^i = \frac{u + c_s^i + f^*}{2}; \quad p_p^i(x) = \frac{u + c_s^i + f^* + 2t(2x-1)}{2}$$

PROOF. See the Appendix. □

Figure 1 gives a visual representation of the seller's and platform's profits in a given market, as well as the equilibrium fee value f^*.

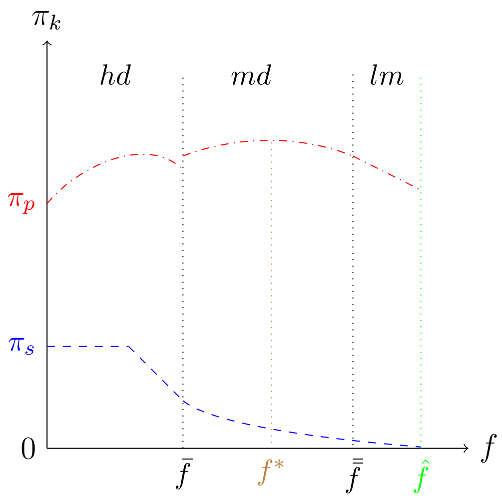


FIGURE 1. The equilibrium market configuration. $\forall\ c_s^i, c_p \in [0,1]$ the optimal platform sets a fee f^* such that firms act as monopolistic duopoly (md). The red dashdotted curve is the profit function of the platform, whereas the blue dashed curve is the profit function of the seller. Thresholds $Barf$ and $Bar\ Barf$ sort between Hotelling duopoly (hd) and monopolistic duopoly, and between monopolistic duopoly and local monopolies (lm), respectively. Above $\hat{f}$ the seller leaves the market.

The intuition behind this result is that the platform faces a trade-off when setting f. On the one hand, a higher fee results in higher agency profits and a lower degree of competition with the seller, which increases overall profits. On the other hand, a fee that is too high could lead to unserved consumers, which in turn reduces the overall surplus that can be extracted. Intuitively, a configuration such as duopolistic monopoly allows the platform to ensure full market coverage, maintain a low degree of competition against the sellers and extract all surplus from his captive consumers.

3.2. Data sharing

Suppose instead that data sharing is mandated so that sellers can also operate first-degree price discrimination. Depending on the market configuration, we can sort consumers into three groups. The first one includes those consumers that can only be profitably targeted by the seller. Consequently, the seller offers them its tailored price $p_s^{TM}(x)$ and extracts all the surplus from them. Similarly, the second group includes those consumers that can only be profitably targeted by the platform. To them, the platform offers a tailored price $p_p^{TMk}(x)$. Finally, the third group includes those consumers who can be reached by both the seller and the platform and are thus contested. Since both firms can price discriminate, they technically compete *à la Bertrand* for each consumer of the third group. To do so, they set a tailored price $p_k^{TC}(x)$. Thus, under data sharing, neither the sellers nor the platform serve consumers through their uniform prices, which in turn do not affect their strategies.

The intense competition to *conquer* contested consumers generates ambiguous incentives on the platform. Because of the transportation costs, the firms have to offer consumers prices that are decreasing in the preference mismatch (i.e., the distance between consumers and firms' locations). However, the platform earns f from every consumer who purchases the seller's variety. Consequently, the platform may adopt sophisticated pricing strategies to *regulate* competition with the seller for consumers in the third group (i.e., contestable ones). We derive the following lemma:

LEMMA 1. *The platform strategically gives up competition for consumers located in* $x^i \in [\tilde{x}^i_{p,no}, \tilde{x}^i_{s,no}]$. *It does so because the transaction fee earned by allowing the seller to serve those consumers is larger than the surplus the platform could directly extract through personalized pricing.*

PROOF. The proof of Lemma 1 stems from the following considerations. First, define $\tilde{x}_{p_2,no}$ as the last consumer from whom the platform can extract at least $c_p + f$. This consumer is necessarily closer to the platform than the consumer in $\tilde{x}_{p,no}$, by construction. It follows that if the consumer indifferent between the seller's good and the zero payoff is located in $\tilde{x}_{s,no} \in [\tilde{x}_{p,no}; \tilde{x}_{p_2,no}]$, then the platform earns $f > p_p^{TC}(x \in [\tilde{x}_{p,no}, \tilde{x}_{s,no}]) - c_p$ by renouncing to compete. □

To understand this seemingly counterintuitive result, consider a platform with a marginal cost c_p. If there exist some consumers who are contested, then the platform's standard strategy is to offer $p_p^{TC}(x)$ and undercut the rival. However, if $p_p^{TC}(x) < c_p + f$, the net revenues the platform earns from winning the price competition is $p_p^{TC}(x) - c_p < f$. Instead, by giving up those consumers and allowing the seller to serve them, the platform earns f without producing anything. Thus, the platform prefers strategically losing price competition because it is more efficient to let the seller serve that market segment, allowing it to fully extract the surplus from those consumers.

Giving up the price competition is a viable option if and only if those consumers are contested (third group). Otherwise, if they cannot be profitably targeted by the seller, the platform earns nothing from not offering them a tailored price. We define $\tilde{x}_{p2,no} > \tilde{x}_{p,no}$ the location of the last contested consumer that the platform can conquer with a price $c_p + f$. Figure 2 graphically describes the combination of market configuration and pricing strategies of the two firms.

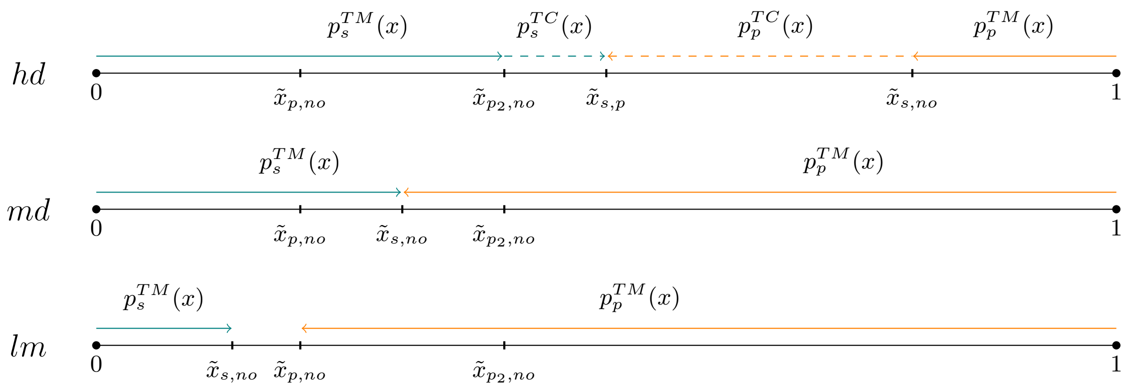


FIGURE 2. Market configuration and pricing strategies. When $f > \bar{\bar{f}}_{ds}$, firms are local monopolies (lm). If $f \in [\bar{f}_{ds}, \bar{\bar{f}}_{ds}]$, firms operate in monopolistic duopoly. Finally, if $f < \bar{f}_{ds}$, firms compete in a Hotelling duopoly. For illustrative reasons, the diagram shows the case $c_p = c_s^i$.

The intuition behind the different market configurations is similar to that presented under the no data sharing case. If f is high enough, then the seller's and platform's market segments do not overlap. As there are no contested consumers, each firm offers its monopoly tailored price to each of the consumers belonging to their turf.

As f decreases, market segments start overlapping, and a group of contested consumers emerges. However, the platform is better off leaving these potentially contested consumers to the seller, as poaching them would result in the platform obtaining individual profits lower than f. The platform thus does not contest such consumers, so the seller can still serve them with its monopoly-tailored price. Finally, for low enough values of f, the platform is incentivized to actually poach some of the consumers that can be reached by both firms. Thus, both firms will offer their monopoly-tailored prices to captive consumers and competitive-tailored prices to the contested consumers from whom the platform can extract individual profits higher than f.

In the Appendix, we identify the cut-off values of f such that the different market configurations emerge and solve the game under each configuration to derive the equilibrium in every market. Moreover, we prove that:

PROPOSITION 3. *Assume there is data sharing. The platform sets a fee*

$$f^*_{ds} = \frac{3u - t - c_p - 1}{3}$$

such that the equilibrium market configuration is monopolistic duopoly (md) in every market i. The equilibrium prices are:

$$p_s^{TM}(x) = u - t\,x \qquad p_p^{TM}(x) = u - t(1-x);$$

PROOF. See the Appendix. □

The intuition behind the results stated in Proposition 3 resembles the one described for the scenario with no data sharing: the platform can maximize its profits by ensuring full market coverage while simultaneously softening competition with the seller. As to the equilibrium fee, we find that f^*_{ds} can be lower or higher than f^*, depending on the parameters. In particular, when c_p is relatively low, the platform can serve most of the consumers in all markets. If this is the case and given the efficiency of the seller, the platform can strategically set a larger fee to raise its rival costs without uncovering the market. Instead, if c_p is relatively high, the platform may not be able to cover a large section of the market. Lowering the fee enables the sellers to serve more consumers and allows the platform to keep earning revenues from all consumers in the market. One may also notice that, due to data sharing, the equilibrium fee decreases faster in c_p than without data sharing. Indeed, data allow the sellers to price more efficiently and to cover larger sections of the market.

Finally, from a welfare perspective, we find that:

PROPOSITION 4. *Data sharing exerts a negative effect on consumers, who are strictly worse off, and a positive effect on firms' payoffs, which strictly increase. Total welfare increases as the latter effect dominates the former.*

PROOF. See the Appendix. □

Proposition 4 states that all consumers are worse off under mandated data sharing, whereas firms are better off. On the one hand, the platform can extract more profits through the per-unit fee while still using it to avoid direct competition with sellers. On the other hand, sellers become more efficient in extracting surplus and thus enjoy higher profits. Instead, all consumers obtain zero surplus in equilibrium. In fact, by controlling the fee, the platform sets up a monopolistic duopoly with the seller, and both of them can perfectly extract surplus through price discrimination, leaving consumers with no residual utility. This result is novel in the literature. Most models stemming from Thisse and Vives (1988) seminal work highlight how allowing both competitors to price discriminate leads to price wars, which largely benefit consumers. Instead, we show that, by strategically using the fee, the platform can avoid price wars altogether, leading to a fully covered market under monopolistic pricing.

4. Discussion and conclusions

Consumer data are becoming an essential input in the digital economy. In particular, hybrid marketplaces can collect vast amounts of data, transform them into valuable information and then use them to compete against the sellers they host. In this work, we investigate the effect of a mandated data sharing policy on market outcomes when data allow to price discriminate consumers.

The previous literature (Prüfer and Schottmüller, 2021; Krämer and Shekhar, 2022) has focused on the effects that mandated data sharing can have on platforms' incentives to innovate, highlighting how the effect on consumer surplus can be either positive or negative depending on the model's characteristics. When data enable price discrimination, allowing all firms to obtain them usually benefits consumers, as firms engage in price wars (see Montes et al. (2019), Bounie et al. (2021) and Abrardi et al. (2022) among others). However, as far as we know, we are the first to analyze the effects of mandated data sharing when the platform can compete with downstream sellers.

We find that mandated data sharing may instead negatively affect consumers. The reason is that the platform can avoid data-induced price wars by setting a high per-transaction fee, which incentivizes sellers to set monopolistic prices, even when a market is fully covered. By avoiding downstream competition, both the sellers and the platform can use data-enabled price discrimination to extract higher surplus from consumers, leaving them worse off.

These results highlight the complexity of the effects of a mandated data sharing policy, as they are ambiguous and hard to predict. Indeed, we argue that in markets imperfect competition takes place, mandatory data sharing may damage the very agents it is intended to protect, namely consumers.

Turning to platform profits, our analysis shows that they increase under data sharing. Then, a question naturally arises: if platforms unambiguously benefit from data sharing, why should a policymaker mandate it? We interpret this seemingly paradoxical result as the consequences of hidden costs we fail to model. Data sharing does not consist of a mere transfer of a file via email, but it entails investments in interoperability between sellers and buyers. Those costs could be non-negligible and could also entail competitive risks for the platform, as data sharing could stimulate entry by new platforms, exerting potential negative pressure on the incumbent. Indeed, the recitals of the DMA place market contestability among the important goals of the act. Further research is thus needed to better analyze these additional characteristics.

Bibliography

Abrardi, L., Cambini, C., Congiu, R., and Pino, F. (2022). User data and endogenous entry in online markets. *Available at SSRN*, *4256544*.

Anderson, S., and Bedre-Defolie, Ö. (2021). Hybrid platform model. https://doi.org/https://ssrn.com/abstract=3867851

Aparicio, D., Metzman, Z., and Rigobon, R. (2021). *The pricing strategies of online grocery retailers* [NBER working paper n. 28639].

Bacchiega, E., Carroni, E., and Fedele, A. (2021). Monopolistic duopoly. *Mimeo*.

Baik, S. A., Larson, N., et al. (2022). Price discrimination in the information age: Prices, poaching, and privacy with personalized targeted discounts. *The Review of Economic Studies*.

Beard, T. R., Kaserman, D. L., and Mayo, J. W. (2001). Regulation, vertical integration and sabotage. *The Journal of Industrial Economics*, *49*(3), 319–333.

Bergemann, D., and Bonatti, A. (2019). Markets for information: An introduction. *Annual Review of Economics*, *11*, 85–107.

Bounie, D., Dubus, A., and Waelbroeck, P. (2021). Selling strategic information in digital competitive markets. *The RAND Journal of Economics*, *52*(2), 283–313. https://doi.org/10.1111/1756-2171.12369

Bourreau, M., and Gaudin, G. (2022). Streaming platform and strategic recommendation bias. *Journal of Economics & Management Strategy*, *31*(1), 25–47.

De Corniere, A., and Taylor, G. (2020). Data and competition: A general framework with applications to mergers, market structure, and privacy policy.

Delbono, F., Reggiani, C., and Sandrini, L. (2021). *Strategic data sales to competing firms* (Technical Report JRC126568). JRC Digital Economy Working Paper, Seville, Spain. Seville, Spain. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/strategic-data-sales-competing-firms

Dryden, N., Khodjamirian, S., and Padilla, J. (2020). The simple economics of hybrid marketplaces. https://doi.org/https://ssrn.com/abstract=3650903

Economides, N. (1998). The incentive for non-price discrimination by an input monopolist. *International Journal of Industrial Organization*, *16*(3), 271–284.

Etro, F. (2021a). Product selection in online marketplaces. *Journal of Economics & Management strategy*, *30*(3), 614–637.

Etro, F. (2021b). Hybrid marketplaces with free entry of sellers. *Available at SSRN*.

Foerderer, J., Kude, T., Mithas, S., and Heinzl, A. (2018). Does platform owner's entry crowd out innovation? evidence from google photos. *Information Systems Research*, *29*(2), 444–460.

Goldfarb, A., and Tucker, C. (2019). Digital economics. *Journal of Economic Literature*, *57*(1), 3–43. https://doi.org/10.1257/jel.20171452

Krämer, J., and Schnurr, D. (2022). Big data and digital markets contestability: Theory of harm and data access remedies. *Journal of Competition Law & Economics*, *18*(2), 255–322.

Krämer, J., and Shekhar, S. (2022). Regulating algorithmic learning in digital platform ecosystems through data sharing and data siloing: Consequences for innovation and welfare. *Available at SSRN.*

Liu, Z., Zhang, D., and Zhang, F. (2021). Information sharing on retail platforms. *Manufacturing and Service Operations Management*, *23*(3), 547–730.

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*, 79–84. https://doi.org/10.1145/2390231.2390245

Montes, R., Sand-Zantman, W., and Valletti, T. (2019). The Value of Personal Information in Online Markets with Endogenous Privacy. *Management Science*, *65*(3), 955–1453. https://doi.org/10.1287/mnsc.2017.2989

Padilla, J., Perkins, J., and Piccolo, S. (2022). Self-preferencing in markets with vertically integrated gatekeeper platforms. *The Journal of Industrial Economics*, *70*(2), 371–395.

Pino, F. (2022). The microeconomics of data–a survey. *Journal of Industrial and Business Economics*, *49*(3), 635–665.

Prüfer, J., and Schottmüller, C. (2021). Competing with big data. *The Journal of Industrial Economics*, *69*(4), 967–1008.

Raith, M. (1996). A general model of information sharing in oligopoly. *Journal of economic theory*, *71*(1), 260–288.

Rhodes, A., and Zhou, J. (2022). Personalized pricing and competition. *SSRN Working Paper 4103763.*

Thépot, J. (2007). Prices as strategic substitutes in a spatial oligopoly. *Available at SSRN 963411.*

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

Wen, W., and Zhu, F. (2019). Threat of platform-owner entry and complementor responses: Evidence from the mobile app market. *Strategic Management Journal*, *40*(1).

Zhu, F., and Liu, Q. (2018). Competing with complementors: An empirical look at amazon.com. *Strategic Management Journal*, *39*(10), 2618–2642.

Appendix

Proof of Proposition 1. To solve the model, we first focus on a given market k. Depending on the value of f, the market can either be fully or partially covered. Moreover, the seller can choose to either set his price following standard Hotelling competition or to set it as a local monopolist. When useful, the superscript indicates the level of market coverage (hd - hotelling duopoly, dm - duopolistic monopoly lm - local monopoly respectively), while the subscript indicates the seller's pricing strategy (c - competitive and m - monopoly respectively).

First, suppose that the seller sets his price as a local monopolist and that f is so high that the market is partially covered. The platform extracts all surplus from consumers through tailored prices. FOCs of the seller profits with respect to its price leads to

$$p_{s_m}^{lm} = \frac{c_s + f + u}{2}; \qquad p_{p_m}^{lm} = u - t + tx.$$

The seller opts to only pass half of the fee to consumers, as a way to obtain a higher market share. Indifferent consumers' locations are

$$\tilde{x}_{s,no} = \frac{u - c_s - f}{2t}; \qquad \tilde{x}_{p,no} = \frac{c_p + t - u}{t},$$

while profits are

$$\pi_{s_m}^{lm} = \frac{(c_s + f - u)^2}{4t}; \qquad \pi_{p_m}^{lm} = \frac{c_p^2 - f(c_s + f) - 2c_p u + u(f + u)}{2t}.$$

These results hold as long as the seller can make positive profits and market shares do not overlap. The seller obtains positive profits as long as it can at least profitably serve the consumer located in $x = 0$. Thus, if $f > u - c_s$, the seller would not enter the market and the platform could not enter. Instead, by equating the indifferent consumers' locations, we find that if $f \leq 3u - 2t - c_s - 2c_p$, then market shares overlap.

When market shares overlap, the seller's pricing strategy does not change, as he still prices as a local monopolist. On the other hand, the platform adjusts his pricing strategy: while the platform can still extract all surplus from consumers located in $(\tilde{x}_{s,no}, 1]$, he instead must beat the seller's offer for consumers located in $[\tilde{x}_{p,no}, \tilde{x}_{s,no}]$. Thus, on this last segment, the platform sets a tailored price equal to

$$p_{p_m}^{dm} = \frac{c_s + f - 2t + u + 4tx}{2},$$

which results in the indifferent consumer being located in

$$\tilde{x}_{s,p} = \frac{2c_p - c_s - f + 2t - u}{4t}.$$

This results in profit being

$$\pi^{dm}_{s_m} = \frac{(c_s + f - u)(-2c_p + c_s + f - 2t + u)}{8t};$$
$$\pi^{dm}_{p_m} = \frac{4c_p^2 - c_s^2 - 6c_s f - 5f^2 - 4tc_s + 4tf - 4t^2 + 2u(3c_s + f + 6t) - u^2 - 4c_p(c_s - f + 2t + u)}{16t}.$$

Next, suppose instead that the seller adopts competitive pricing, and that f is low enough that we have full market coverage. In this scenario we have standard Hotelling competition, where both the seller and the platform simultaneously set their uniform prices. Then, the platform will set the tailored prices. By standard computations, we obtain equilibrium prices

$$p^{hd}_{s_c} = \frac{c_s + c_p + 2f + t}{2}; \qquad p^{hd}_{p_c} = cp + f.$$

As we can see, the platform sets his uniform price as low as possible, since in equilibrium all the consumers he serves will purchase through tailored prices, which are equal to

$$p^{hd}_{p_c}(x) = \frac{c_s + c_p + 2f - t + 4tx}{2}.$$

These prices lead to the indifferent consumer being located in

$$\tilde{x}_{s,p} = \frac{c_p - c_s + t}{4t},$$

while equilibrium profits are equal to

$$\pi^{hd}_{s_c} = \frac{(c_p - c_s + t)^2}{8t}; \qquad \pi^{hd}_{p_c} = \frac{(c_p - c_s)^2 + 2t(8f + 3(c_s - c_p)) + 9t^2}{16t}.$$

From the platform's profits function, it is clear that his profits are increasing in f. In turn, the fee is directly passed to the consumers, both by the seller and by the platform. By analysing consumer utility, we find that the consumer with the lowest utility after purchase is the one located in $x = 1$, as the platform can extract most of her surplus. Thus, these results hold as long as the net utility of the consumer located in $x = 1$ is ≥ 0, which translates to $f \leq \frac{2u-3t-c_p-c_s}{2}$.

When $f > \frac{2u-3t-c_p-c_s}{2}$, the strategies regarding uniform pricing do not change; however, the platform must change the tailored prices he proposes to consumers located on the segment $[\frac{2u-t-2f-c_p-c_s}{2t}, 1]$ in order to allow consumers to maintain non-negative utility. To those consumers, the platform offers a tailored price equal to

$$p_{p_c}^{hd}(x) = u - t(1-x) - c_p,$$

leading to profits equal to

$$\pi_{p_c}^{hd} = \frac{-c_p^2 - 6c_pc_s - c_s^2 - 8fc_p - 8fc_s - 8f^2 - 18tc_p - 6tc_s - 8ft - 9t^2 + 8u(c_p + c_s + 2f + 3t) - 8u^2}{16t}.$$

As f increases, consumers' net utility decreases. In particular, the next threshold is reached when the indifferent consumer $\tilde{x}_{s,p}$ net utility is equal to 0, which translates to $f = \frac{4u-3t-c_s-3c_p}{4}$. When $f \geq \frac{4u-3t-c_s-3c_p}{4}$, duopolistic monopoly ensues. While the seller loses market share due to the increased fee, the platform can poach these consumers through tailored prices, leading to profits equal to

$$\pi_{s_c}^{dm} = \frac{(c_p - c_s + t)(2u - t - 2f - c_p - c_s)}{4t}$$

$$\pi_{p_c}^{dm} = \frac{-5c_p^2 - c_s^2 - 8fc_s - 12f^2 - 6tc_s - 16tf - 9t^2 - 2c_p(3c_s + 8f + 9t - 8u) + 8u(c_s + 3f + 3t) - 12u}{8t}$$

Finally, as f increases further, the consumers lost by the seller are too far from the platform to be poached. This happens when $\tilde{x}_{s,no} = \tilde{x}_{p,no}$, which results in $f = \frac{4u-3t-c_s-3c_p}{2}$. Above this threshold, the platform and seller become local monopolists: the platform extracts all available surplus from his consumers, while the seller continues to price competitevely by construction. The seller's profits maintain the same form as above, while platform profits are

$$\pi_{p_c}^{lm} = \frac{c_p^2 - f(c_s + 2f + t) + 2fu + u^2 - c_p(f + 2u)}{2t}$$

Finally, when $f \geq \frac{4u-3t-c_s-3c_p}{2}$, the seller cannot profitably serve any consumer, and thus he does not enter the market, and neither can the platform.

Having analysed all cases, we now focus on the seller's pricing strategy as a function of f. By comparing the profits function, we find that if $f < \bar{f}$ the seller obtains higher profits with competitive pricing, while otherwise he opts for monopolistic pricing. $\bar{f}$ is found by equating the seller's profits function, and is equal to

$$\bar{f} = c_s - c_p - t + \sqrt{2c_s^2 - c_p^2 - 2tc_p - t^2 + 2c_pu - 4c_su + 2tu + u^2}$$

Proof of Proposition 2. Seller's and platform's profits as a function of f are

$$\pi_s = \begin{cases} \frac{(c_p-c_s+t)^2}{8t} \quad \text{for} \quad 0 \le f < \frac{4u-3t-c_s-3c_p}{4} \\ \frac{(c_p-c_s+t)(2u-t-2f-c_p-c_s)}{4t} \quad \text{for} \quad \frac{4u-3t-c_s-3c_p}{4} \le f < \bar{f} \\ \frac{(c_s+f-u)(-2c_p+c_s+f-2t+u)}{8t} \quad \text{for} \quad \bar{f} \le f < 3u-2t-c_s-2c_p \\ \frac{(c_s+f-u)^2}{4t} \quad \text{for} \quad 3u-2t-c_s-2c_p \le f < u-c_s \\ 0 \quad \text{for} \quad f \ge u-c_s \end{cases}$$

$$\pi_p = \begin{cases} \frac{(c_p-c_s)^2+2t(8f+3(c_s-c_p))+9t^2}{16t} \quad \text{for} \quad 0 \le f < \frac{2u-3t-c_p-c_s}{2} \\ \frac{-c_p^2-6c_pc_s-c_s^2-8fc_p-8fc_s-8f^2-18tc_p-6tc_s-8ft-9t^2+8u(c_p+c_s+2f+3t)-8u^2}{16t} \quad \text{for} \quad \frac{2u-3t-c_p-c_s}{2} \le f < \frac{4u-3t-3c_p-c_s}{4} \\ \frac{-5c_p^2-c_s^2-8fc_s-12f^2-6tc_s-16tf-9t^2-2c_p(3c_s+8f+9t-8u)+8u(c_s+3f+3t)-12u^2)}{8t} \quad \text{for} \quad \le f < \bar{f} \\ \frac{4c_p^2-c_s^2-6c_sf-5f^2-4tc_s+4tf-4t^2+2u(3c_s+f+6t)-u^2-4c_p(c_s-f+2t+u)}{16t} \quad \text{for} \quad \bar{f} \le f < 3u-2t-c_s-2c_p \\ \frac{c_p^2-f(c_s+f)-2c_pu+u(f+u)}{2t} \quad \text{for} \quad 3u-2t-c_s-2c_p \le f < u-c_s \\ 0 \quad \text{for} \quad f \ge u-c_s \end{cases}$$

To find the equilibrium fee, we convert all the thresholds on f to thresholds on c_p, and maximize platform's profits across all markets:

$$\max_f \int_0^1 \pi_p \, dc_s$$

Standard calculations yield to $f^* = \frac{2u+4t+4c_p-3}{10}$, which corresponds to all markets being under duopolistic monopoly with the sellers pricing as a local monopolist. Replacing f^* in firms' prices and profits gives the results described in the Proposition.

Proof of Proposition 3. When both the platform and sellers have data, they price discriminate all consumers they serve. Thus, their uniform prices do not influence their strategies.

First, suppose that f is so high that the market is partially covered: then, as no consumer can be reached by both the platform and the seller, each of them will set their tailored prices to extract all surplus from each consumer. This leads to

$$p_s^{TM}(x) = u - tx; \qquad p_p^{TM}(x) = u - t + tx.$$

the location of the last consumer buying from the seller and the platform are respectively

$$\tilde{x}_{s,no} = \frac{u - c_s - f}{t}; \qquad \tilde{x}_{p,no} = \frac{c_p + t - u}{t},$$

which results in profits being

$$\pi_s^{lm} = \frac{(c_s + f - u)^2}{t}; \qquad \pi_p^{lm} = \frac{c_p^2 - 2f(c_s + f) - 2uc_p + 2fu + u^2}{2t}.$$

When $f \geq u - c_s$, the seller cannot profitably serve any consumer and leaves the market, thus also impeding entry to the platform. Instead, when $f < 2u - t - c_p - c_s$ we have $\tilde{x}_{s,no} > \tilde{x}_{p,no}$, and the consumers in $[\tilde{x}_{p,no}, \tilde{x}_{s,no}]$ become contestable.

Let us focus on one of these contestable consumers. If she is served by the seller, the platform obtains profits equal to f. Instead, if she is served by the platform, the platform obtains $p_p^{TM}(x) - c_p$. By standard calculations, we find that the platform is better off by leaving to the seller all the consumers located in $[\tilde{x}_{p,no}, \tilde{x}_{p2,no}]$, where $\tilde{x}_{p2,no} = \frac{c_p+f+t-u}{t}$. Intuitevely, $\tilde{x}_{p2,no}$ is the last consumer from which the platform can extract revenues equal to $c_p + f$. Thus, it is more profitable for the platform to leave consumers in $[\tilde{x}_{p,no}, \tilde{x}_{p2,no}]$ to the seller, as he his more efficient in extracting surplus. Thus, the seller's profits are the integral of the monopolistic tailored price from 0 to $\tilde{x}_{s,no}$, while the platform's profits are the integral of his monopolistic tailored price from $\tilde{x}_{s,no}$ to 1 plus the agency profits $f\tilde{x}_{s,no}$, resulting in

$$\pi_s^{dm} = \frac{(c_s + f - u)^2}{t}$$

$$\pi_p^{dm} = \frac{-c_s^2 - 3f^2 - 2ft - t^2 - 2c_s(2f + t - 2u) - 2c_p(c_s + f + t - u) + 6fu + 4tu - 3u^2}{2t}.$$

Finally, when $\tilde{x}_{s,no} > \tilde{x}_{p2,no}$, the platform starts contesting consumers located in $[\tilde{x}_{p2,no}, \tilde{x}_{s,no}]$ as it can extract from them revenues higher than f. This scenario holds whenever $f < \frac{2u-t-c_p-c_s}{2}$. Price competition follows the following strategies:

- Consumers located in $[0, \tilde{x}_{p2,no})$ are poached by the seller through his monopolistic tailored price $p_s^{TM}(x)$;
- Consumers located in $[\tilde{x}_{p2,no}, \tilde{x}_{s,no})$ are contested;
- Consumers located in $[\tilde{x}_{s,no}, 1]$ are poached by the platform through his monopolistic tailored price $p_p^{TM}(x)$.

Let us focus on the second segment: both the platform and the seller will set their competitive tailored prices to beat the opponent's best offer, as Bertrand competition ensues. The seller will not price any lower than $c_s + f$, while the platform will not price lower than $c_p + f$, leading to

$$p_s^{TC}(x) = c_p + f + t - 2tx; \qquad p_p^{TC}(x) = c_s + f - t + 2tx.$$

The indifferent consumer in the second segment is thus located in

$$\tilde{x}_{s,p} = \frac{c_p - c_s + t}{2t},$$

and profits are equal to

$$\pi_s^{hd} = \frac{-c_p^2 + c_s^2 - 2f^2 - 2c_s t - 4ft - t^2 - 2c_p(c_s + 2f + t - 2u) + 4u(f+t) - 2u^2}{4t}$$

$$\pi_p^{hd} = \frac{c_p^2 - c_s^2 - 4c_s f - 2f^2 - 2tc_s - t^2 - 2c_p(c_s + t) + 4u(c_s + f + t) - 2u^2}{4t}.$$

Platform's profits as a function of f are

$$\pi_p = \begin{cases} \frac{c_p^2 - c_s^2 - 4c_s f - 2f^2 - 2tc_s - t^2 - 2c_p(c_s+t) + 4u(c_s+f+t) - 2u^2}{4t} & \text{for } 0 \leq f < \frac{2u - t - c_p - c_s}{2} \\ \frac{-c_s^2 - 3f^2 - 2ft - t^2 - 2c_s(2f+t-2u) - 2c_p(c_s+f+t-u) + 6fu + 4tu - 3u^2}{2t} & \text{for } \frac{2u-t-c_p-c_s}{2} \leq f < 2u - t - c_p - c_s \\ \frac{c_p^2 - 2f(c_s+f) - 2uc_p + 2fu + u^2}{2t} & \text{for } 2u - t - c_p - c_s \leq f < u - c_s \\ 0 \quad \text{for } f \geq u - c_s \end{cases}$$

Having computed profits for all levels of f, we now find the equilibrium fee by maximizing platform total profits with respect to f. FOC w.r.t. to f result in $f_{ds}^* = \frac{3u - t - c_p - 1}{3}$. When plugged in the functions above, we find that this fee results in all markets being under duopolistic monopoly, with both the seller and the platform only pricing through their monopolistic tailored prices.

Proof of Proposition 4. With regards to platform and sellers' profits, direct comparisons between the results described in Propositions 2 and 3 show that all firms' profits increase with data sharing. With regards to consumer surplus, recall that under no data sharing all markets are under duopolistic monopoly with the seller opting for monopolistic pricing. Consumer surplus in a given market k is thus equal to

$$CS^k = \int_0^{\tilde{x}_{p,no}} u - tx - p_{s_m}^{lm}\, dx + \int_{\tilde{x}_{p,no}}^{\tilde{x}_{s,no}} u - t(1-x) - p_{p_m}^{dm}\, dx + \int_{\tilde{x}_{s,no}}^{1} u - t(1-x) - p_{p_m}^{lm}\, dx =$$

$$= \frac{(4c_p - 3 + 10c_s + 4t - 8u)^2}{800t},$$

which, once summed across all markets leads to

$$CS = \frac{(7 + 4c_p + 4t - 8u)^3 - (4c_p - 3 + 4t - 8u)^3}{24000t}.$$

Instead, under data sharing, the seller and the platform never compete head to head for any consumer: thus, they can extract all surplus from consumers, and $CS = 0$.

By adding firms' profits and CS in the two cases, we find that total welfare increases under data sharing.