Learning-by-doing in Data Markets*

Antoine Dubus,† Wilfried Sand-Zantman‡

December 17, 2024

Abstract

Data brokers learn from the large amounts of data they treat, lower their operational costs and gain efficiency, improving their ability to supply information to industries. When competing firms can buy consumer-related information from those brokers, we derive the implications of learning-by-doing from data on the firms' and brokers' profits. We show that learning effects can *harm* a data broker when firms anticipate how providing their data lowers the treatment cost of the broker, allowing it to sell information to rival firms. This strategic anticipation can deter firms from buying information from the broker, if doing so implies that competitors also remain uninformed, and all the more so when the broker enjoys a strong learning effect. When brokers compete, a broker with weak learning effects can dominate its competitors, and even achieve *monopoly* profits when competing with *more efficient* brokers.

Keywords: Data brokers, Information, Innovation, Competition, Exclusive deals.

JEL: D21; D43; D83; K21; L13; L40.

---

*We would like to thank Roman Dolecek, Flavio Pino, Maxence Valentin, as well as participants to the 2024 EARIE conference and to the Université Paris Cité seminar for useful remarks and suggestions.

†Department of Management, Technology and Economics, ETH Zurich Leonhardstrasse 21 Switzerland – 8092 Zurich, Switzerland; antoine1dubus@gmail.com.

‡ESSEC Business School and THEMA (CNRS); sand@essec.edu.

1 Introduction

The digitization of the economy has seen the rise of data markets. In industries such as online retail, banking, and real estate, companies now routinely acquire the services of data brokers such as Nielsen, Acxiom or Equifax. Firms provide their raw data to the brokers who use their powerful AI analytics to return fine-grained information, helping firms optimize interactions with their customers. For instance, Nielsen's Scantrack service enables retailers to share customer purchase and payment data directly with Nielsen (Koski, 2018). In turn, Nielsen provides refined information on these customers, allowing retailers to personalize their offers. Similarly, in the US, banks such as Bank of America provide Equifax, Acxiom and Transunion with customer financial data,[1] and then purchase the credit scores computed by these data brokers to offer consumers personalized loan conditions.[2] As pointed out by the FTC (2014), "data brokers report that they obtain data directly from their merchant and financial service company clients". Hence, by selling their analytics services, data brokers learn from the data of their clients and improve the quality of their data sets and algorithms. Data markets are characterized by effects of learning-by-doing (Bajari et al., 2019), and as data brokers accumulate data and experience, they can provide clients with better, more cost-effective analytics services.

A natural implication of these learning effects is that brokers should improve their performance as they treat more data, gain a strong competitive edge, and, in turn, increase their profits. This dynamic raises important questions: does a broker necessarily benefit from lower data treatment costs, or can stronger learning effects actually reduce its profits? In turn, when brokers compete, is an efficient broker that learns a lot from data necessarily more competitive than a less efficient broker?

Our main contribution is to show that data brokers can be worse off when they become more efficient at learning from data. Indeed, when a broker sells information as a monopolist, a stronger learning effect can reduce its profits by making firms reluctant to provide their data. Hence, a broker can achieve *lower profits* when it is able to lower its treatment cost. When many brokers compete, a strong learning effect can induce a competitive *disadvantage*, allowing a broker unable to learn from data to dominate its more efficient competitors. Perhaps even more surprisingly, we show that in this case, the least efficient broker can charge the highest price for information and make *monopoly profits*, while more efficient competitors make zero profits. When all brokers benefit from

[1] What You Need To Know About The Three Main Credit Bureaus; Forbes, January 15 2021.
[2] Which Credit Bureau Does Bank Of America Use? Forbes, January 31, 2024.

a strong learning effect, we show that firms may prefer not to acquire their services at all, while they would do so if brokers had a weaker learning effect.

To set up our analysis, we consider first a monopolist data broker and two firms – labeled A and B – competing on a product market à la Hotelling. These firms possess raw consumer data, which the broker can treat to return refined information on the willingness to pay of these customers, allowing firms to charge personalized prices. This requires the firms to provide its data to the broker, which reduces its further treatment cost according to an effect of learning-by-doing. To model the dynamic process, we assume (without loss of generality) that Firm A decides first whether to engage with the broker, followed by Firm B. Throughout the analysis, we say equivalently that the broker treats the data of the firms or sells them information.

To maximize its profit, the optimal strategy of the broker is to sell information exclusively to Firm A. Indeed, the total profits of the firms are reduced when they both buy information and, thus in this case, compete more fiercely. When the broker can commit initially not to contract with Firm B later, this strategy is implemented and allows the broker to earn its maximal level of profit. But when commitment is not possible, the broker may be tempted to sell its service both to Firm A but also to Firm B. Suppose first that learning effects are weak so potentially high data treatment cost. Then, despite the lack of commitment power, the broker can still sell its service to Firm A only and charge a premium price for exclusive information. Indeed, the additional profit Firm B could make by having his data treated after Firm A is unlikely to cover the additional treatment cost. So weak learning effects substitute for the commitment power, allowing the broker to make an exclusive sale to Firm A, and to maximize its profits.

With strong learning effects instead, Firm A anticipates that by sharing its data, it will reduce the broker's treatment costs, making it easier for the broker to sell information to Firm B as well. This would intensify competition in the product market, thereby reducing Firm A's willingness to pay for information, down to a point where treating the data of Firm A is not profitable anymore for the broker. In this case, the broker ends up selling information only to Firm B, and at a lower price than when selling only to Firm A. Indeed, when selling to Firm A only, the broker could leverage the threat to sell to Firm B if Firm A declined the offer. This threat is ineffective when selling only to Firm B, as the broker cannot later offer information to Firm A, resulting in a much lower price than Firm A would have paid for exclusive information. Hence, the ability to lower its treatment cost through a learning effect can *reduce* the profits of the broker.

When multiple rounds of negotiation (or renegotiation) are possible, allowing the broker to re-approach Firm A after selling to Firm B, the adverse impact of learning on profits is even stronger, and may result in neither firm purchasing information. If Firm A declined the initial offer of the broker, Firm B, in turn, faces a similar decision as Firm A before and also chooses not to buy information. By not purchasing information, a firm effectively prevents its competitor from acquiring information, and both firms remain uninformed. Hence, the broker can even *not sell information to firms at all* – and make zero profits – precisely because it learns from data.

When two brokers compete with potentially different abilities to learn from data, Firm A may prefer to buy information from the broker that learns the least from data. Indeed, by purchasing from the least efficient broker, Firm A is sure to be the only firm competing with information and to make more profit than Firm B downstream. In this case, the least efficient broker dominates its more efficient competitor, and for a broker, the ability to learn from data and improve its efficiency can turn into a competitive *disadvantage*.

Different abilities to learn can also enable brokers to escape Bertrand competition and achieve *monopoly* profits. This effect holds for a more efficient broker, but also for its less efficient competitor. Hence, introducing competition between buyers can reverse standard analysis, which typically finds how learning effects trigger price-undercutting strategies and predation practices among competing suppliers (Cabral and Riordan, 1994; Besanko et al., 2014). In our model with competing buyers, an information seller can achieve monopoly profits due to the different abilities of competing sellers to learn from data.

The learning effect can also induce an expansion of the data market. When the amount of data to treat is low, the broker may benefit from treating the data of both firms instead of only one firm. In particular, the potential losses from treating the data of Firm A at a high cost can be offset by the profit gain from reduced treatment costs when treating the data of Firm B. In this case, treating the data of Firm A has the value of an investment in cost-reducing innovation that benefits the broker.

Lastly, we show that a broker may be willing to share data (potentially in exchange for money) with another broker who is better at learning. Doing so allows the more efficient broker to sell to Firm B, an option that may not be profitable for the less efficient broker because of the high treatment cost. However, the less efficient broker then loses the competitive advantage provided by its limited learning capacity and the ability to guarantee exclusivity to Firm A. Introducing data property rights, granting firms control over the further use of their data by the brokers, can mitigate this negative effect of data sharing.

Our results have both managerial and policy implications. First, managers of the competing firms should consider their data as valuable assets benefiting brokers. With this in mind, managers should reconsider the value added from purchasing the analytics service of the brokers. If doing so generates a learning effect that allows their competitors to also benefit from these services, they may eventually lose from having their data treated.

Secondly, managers of data broker companies should account for these learning effects when dealing with their clients. The competitive nature of the markets in which they sell their services can induce subtle effects when a broker can learn from the data it treats. On the one hand, if a broker learns a lot from its data, it can benefit from treating the data of Firm A at a loss to reduce its further treatment cost. In this case, treating the data of Firm A has the value of an investment in a cost-reducing innovation. On the other hand, a broker does not always benefit from being better at learning from data. When brokers compete, the least efficient one may end up selling information as a monopolist while its more efficient competitor makes zero profits.

In terms of policy and welfare, our results contribute to policy debates where regulators are increasingly wary of the growing market power of data brokers, in particular regarding their ability to control the (exclusive) supply of competitive information.[3] By highlighting the importance of learning-by-doing from data, our results suggest that data markets could expand as data brokers become more efficient, selling information to all competing firms, thereby intensifying competition in the product market. This expansion could benefit consumers, but it will not always take place if firms consider the value of their data when engaging with brokers.

The rest of the article is organized as follows. We review the literature on learning-by-doing and on data markets in Section 2. We describe the model in Section 3, and we solve the product-market equilibrium in Section 4. We analyze the strategies of a monopolist data broker in Section 5, and we turn to competition in Section 6. We test the robustness of our results to several extensions in Section 7. We provide managerial and policy implications in Section 8. Section 9 concludes.

[3]See for instance the FTC (2014) report, and more recently, the Digital Platform Services Inquiry – March 2024 report on data brokers of the Office of the Australian Information Commissioner. Last accessed, 26.01.2024.

2 Literature

Learning-by-doing. Effects of learning-by-doing are a central topic in economics since the influential contribution of Arrow (1962). These effects have since then been identified across various industries, including aircraft manufacturing (Rosenberg, 1982; Miller and Chen, 1994; Von Hippel and Tyre, 1995), semiconductor production (Irwin and Klenow, 1994; Hatch and Mowery, 1998), service organizations (Darr et al., 1995), and automobile industries (Levitt et al., 2013).

With the increasing importance of data as a competitive factor, researchers have begun to explore whether digital firms experience learning effects with data. Bajari et al. (2019) provide striking empirical evidence for this in the case of Amazon's retail forecasting system. They show how, by developing new machine learning models to treat data, firms can significantly improve their future performance.

The theoretical literature has modeled learning effects in repeated games where sellers gain competitiveness with each sale. In models with infinite interactions between buyers and sellers, Cabral and Riordan (1994) and Besanko et al. (2014) show how learning-by-doing can lead to predation and dominance among competing suppliers, eventually lowering their profits.

In related contributions, Besanko et al. (2010) analyze a setting where experience may decay over time; Sweeting et al. (2022) consider strategic buyers who anticipate how their consumption will reduce further prices; Deng et al. (2023) look at varying bargaining power in the product market.

We contribute to this literature by introducing competition among buyers. We show that learning-by-doing can lower the profits of an upstream supplier, not only when suppliers compete but also when the supplier is a monopolist. When firms anticipate how purchasing the services of the broker impacts its ability to deal with competitors in the following stages, they may prefer not to buy information at all from the broker. In turn, we show that a weak learning effect can allow a broker to achieve monopoly profits even when it competes with more efficient brokers, such that introducing downstream competition can cancel out predation effects.

Sale of information. The literature has analyzed the competitive impacts of information and the selling strategies of data brokers. We model information allowing competing firms to locate consumers on a Hotelling line. This issue dates to Thisse and Vives (1988) who emphasize the pro-competitive effects of price discrimination. Recent papers have reconsidered this approach: Montes et al. (2019) analyze a broker selling information to firms; Bounie et al. (2021) show that firms benefit from identifying only consumers who are the closest to their location; Abrardi et al. (2024a) revisit this question in a circular city model, and show how a data broker can impact firm entry. Belleflamme et al. (2020) consider an alternative representation where competing firms can sell a good to a group of consumers, and these firms have information only on a share of the consumers in this group. They show that firms benefit from having access to partial information on consumers and that a broker would optimally sell only part of the accessible information, even when it has access to perfect information on consumers. Finally, Abrardi et al. (2024b) combine both types of approaches to analyze synergy gains resulting from the merger of information owned by competing brokers.[4]

We contribute to this literature by analyzing how brokers can learn from the data they treat, and how this changes their incentives to sell information to firms. While this literature usually considers one stage where the broker can sell information to all firms in the market, we consider a sequential sale of information, first to one firm and then to its competitor. Sequential interactions allow us to decompose the effects of learning-by-doing (Cabral and Riordan, 1994; Besanko et al., 2014; Sweeting et al., 2022): the broker can learn from its first interaction with a firm to lower its cost in the following period. Using this setting, we will show in particular that a broker may be unable to sell information to firms if it could learn from their data.

Commitment issues. A recent literature has revisited how the sale of data can generate externalities resulting from the Coase (1972) conjecture. Because data is non-rival, a seller cannot commit not to sell data to multiple buyers, which can generate negative externalities that may end up lowering the profits of the broker.

Jones and Tonetti (2020) consider commitment issues in a data-driven growth model. Liu et al. (2023) analyze how the lack of commitment ability on future sales dampens the market power of data sellers. They show that selling data under a subscription mode allows to escape the pro-competitive effect resulting from this lack of commitment ability.

Different from these approaches, our contribution is to show that data brokers can escape the Coase conjecture and achieve credible commitment when the cost to sell (in our case, the cost to treat data) is sufficiently high. While we do not consider non-rivalry in our model, our results easily generalize to previous settings as soon as selling data induces a cost, for instance from the transaction. This can even allow a broker to reach

[4]Other models consider data brokers selling information to competing platforms when consumers have privacy concerns (Zhang et al., 2024).

monopoly profits when competing with a broker having a different (higher or lower) future treatment cost. In turn, cost reductions resulting from effects of learning-by-doing can harm the profits of brokers by lowering their ability to credibly commit to exclusivity.

3 Model

We consider a market in which two firms, A and B, compete à la Hotelling, i.e., propose horizontally differentiated products to heterogeneous consumers. The two firms are located at the extremities of the Hotelling line, Firm A at 0 and Firm B at 1. Consumers are uniformly distributed on the [0,1] line, and we assume that consumers located at $x \in [0, 1]$ receive a utility V from purchasing the product but incur a cost $t > 0$ of consuming a product that does not perfectly fit their taste x. Therefore, buying from Firm A (resp. from Firm B) incurs a cost tx (resp. $t(1 - x)$). If p_A and p_B denote the prices set by the firms, a consumer's utility is given by

$$u(x) = \begin{cases} V - p_A - tx & \text{if buying from Firm A,} \\ V - p_B - t(1 - x) & \text{if buying from Firm B.} \end{cases}$$

On top of this classical Hotelling model, we assume that both firms have data on consumers located near them. Specifically, Firm A has data on consumers in the interval $[0, \delta_A]$, while Firm B has data on consumers in the interval $[1 - \delta_B, 1]$. We assume that δ_A and δ_B are between 0 and $\frac{1}{2}$. This ensures that $\delta_A \leq 1 - \delta_B$, so there is no overlap in the data each firm has on consumers.

Competition when firms have access to such information structures has been analyzed by Chen et al. (2020) and by Bounie et al. (2021) who show these information partitions allow firms to extract surplus from high-valuation consumers while maintaining a soft competition by leaving a large share of low-valuation consumers unidentified. Our assumption that partitions do not overlap can be motivated for instance if the data owned by each firm is generated by high-value customers sharing data as part of a loyalty program (Gabel and Guhl, 2022).[5] These information structures are represented in Figure 1, where the thick lines correspond to consumers on whom the closest firm has access to data, and the thin line represents consumers on whom firms have no data.

[5]Overlapping partitions are analyzed by Lee et al. (2011) in the case of a Hotelling line, and by Abrardi et al. (2024a) when multiple firms compete in a circular city model.

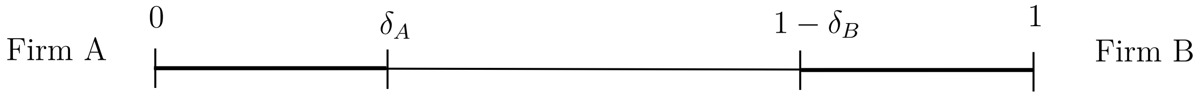


Figure 1: Information structures on the Hotelling line.

Firms cannot exploit this data on their own, but they can engage data brokers to treat this data and return perfect information about these consumers.[6] When brokers compete, it should be clear that a firm benefits from purchasing information from only one broker. Having its data treated by several brokers would be redundant and would not add any value to the firm.

To model the bargaining process between the brokers and the firms, we assume that brokers make sequential take-it-or-leave-it offers to the firms, starting with Firm A.[7] We delve into the details of these offers when solving the game in Sections 5 for a monopolist data broker, and in Section 6 when brokers compete.

Inferring consumer types from data incurs a cost proportional to the amount of data to treat. Moreover, to model the potential learning effects, we assume that this cost decreases with the amount of data previously treated. This representation aligns with the theoretical literature and is supported by empirical evidence (see e.g. Cabral and Riordan (1994), Hatch and Mowery (1998), and Besanko et al. (2019)).

Therefore, if a broker has treated the data of Firm A, it benefits from a cost reduction proportional to δ_A when treating Firm B's data. Hence, when it makes an offer to Firm B, a broker can treat Firm B's data at a lower cost if it has previously treated Firm A's data. Concretely, we assume that the cost function takes the following form

$$\begin{cases} c\delta_A \text{ for Firm A's data,} \\ c\delta_B \text{ for Firm B's data without learning effect,} \\ c\delta_B(1-\alpha\delta_A) \text{ for Firm B's data with learning effect when } \delta_A \text{ data has been treated,} \end{cases}$$

where α measures the strength of the effect of learning-by-doing. To ensure that costs are non-negative in the second stage, we focus on $\alpha \in [0,2]$.

The timing of the game is therefore as follows.[8]

[6] In Section 7.4 we consider the strategic choice of δ_A and δ_B by a broker.

[7] This is in line with the literature on learning-by-doing, as it allows to analyze separately the interactions between an upstream seller and each buyer (Besanko et al., 2010, 2019; Sweeting et al., 2022).

[8] Assuming that firms can observe whether their competitor has had its data treated by the broker is standard in the literature, as it enables the existence of equilibria in pure strategy at the price-competition stage (Thisse and Vives, 1988; Montes et al., 2019; Chen et al., 2020; Bounie et al., 2021). In practice, a firm can easily observe the baseline uniform price charged by its competitor and update its pricing strategy as a response to any change.

- Stage 1: Each broker makes an offer to sell information to Firm A who accepts or refuses the offers.
- Stage 2: Each broker makes an offer to sell information to Firm B who accepts or refuses this offer.
- Stage 3: Firms set a uniform price for consumers on whom they do not have information.
- Stage 4: If the firms have acquired information on δ_A and δ_B consumers, they then personalize prices for these identified consumers.
- Stage 5: Consumers choose whether to buy from the firms and which product to buy.

Our equilibrium concept is subgame perfection, and we will solve this model with backward induction. In particular, we will first analyze Stages 3, 4 and 5 of the game in Section 4, where we characterize the competitive equilibrium in the product market depending on firms' access to consumer information. We then analyze Stages 1 and 2 in Sections 5 and 6 respectively when the broker is a monopolist, and when brokers compete.

Note that we assume for simplicity that the broker's offers and the purchasing decisions of the firms are public.[9] Moreover, if a firm declines a broker's offer there is no possibility of further renegotiation. For instance, if Firm A refuses the offer of each broker in the first stage and Firm B purchases information, Firm A cannot reconsider the offers afterward. We analyze the model with renegotiation in Section 7.1.

4 Competitive Equilibrium in the Product Market

In this section, we analyze competition between Firm A and Firm B in the downstream market after they have acquired some information from a data broker. When Firm A knows the characteristics of consumers on $[0, \delta_A]$, it will set a personalized price $p_A(x)$ for each of them and charge a uniform price p_A to consumers on $[\delta_A, 1]$. Similarly, Firm B will price discriminate consumers on $[1 - \delta_B, 1]$ by setting a price $p_B(x)$ and charge a uniform price p_B to consumers on $[0, 1 - \delta_B]$.

In Stage 4, each firm knows the uniform price set by its competitor and set its personalized prices to match this uniform price – assuming that consumers will choose the

[9]This is a classical assumption in the literature on the sale of information that enables the existence of pure strategy equilibria (Montes et al., 2019; Chen et al., 2020; Bounie et al., 2021).

offer from the informed firms when indifferent. The personalized prices set by Firm A and Firm B satisfy:

$$V - tx - p_A(x) = V - t(1-x) - p_B \implies \quad p_A(x) = p_B + t - 2tx,$$
$$V - t(1-x) - p_B(x) = V - tx - p_A \implies \quad p_B(x) = p_A + 2tx - t.$$

Since we assume that $\delta_A \leq 1 - \delta_B$, there is a submarket consisting of consumers with $[\delta_A, 1-\delta_B]$ for which none of the firms has any precise information. In this submarket, firms compete as in the traditional Hotelling model using uniform prices. Therefore, to compute the demands – denoted d_A for Firm A and d_B for Firm B – we characterize the consumer $\tilde{x}$ who is indifferent between the two offers:

$$V - t\tilde{x} - p_A = V - t(1-\tilde{x}) - p_B \implies \tilde{x} = \frac{p_B - p_A + t}{2t}.$$

This leads to the demand functions $d_A = \tilde{x} - \delta_A = \frac{p_B - p_A + t}{2t} - \delta_A$ and $d_B = 1 - \delta_B - \frac{p_B - p_A + t}{2t}$.

Using these demands, we can compute the profits of the firms. Each firm makes profits by selling at a personalized price to the closest identified consumers, and by selling to consumers in the middle of the Hotelling line (with respective demands of unidentified consumers equal to d_A and d_B) at a uniform price:

$$\pi_A = \int_0^{\delta_A} p_A(x)dx + d_A p_A = \int_0^{\delta_A} (p_B + t - 2tx)dx + \left(\frac{p_B - p_A + t}{2t} - \delta_A\right) p_A,$$
$$\pi_B = \int_{1-\delta_B}^{1} p_B(x)dx + d_B p_B = \int_{1-\delta_B}^{1} (p_A + 2tx - t)dx + \left(\frac{p_A - p_B + t}{2t} - \delta_B\right) p_B.$$

We can now compute the optimal prices and demands, using first-order conditions on π_i with respect to p_i for $i \in \{A, B\}$. The optimality conditions lead to:[10]

$$p_A = t[1 - \frac{2}{3}\delta_B - \frac{4}{3}\delta_A], \quad p_B = t[1 - \frac{2}{3}\delta_A - \frac{4}{3}\delta_B].$$

One can see that uniform prices decrease with the shares of consumers identified δ_A and δ_B, and p_A (resp. p_B) decreases faster with an increase in δ_A (resp. δ_B) than with δ_B (resp. δ_A).

There are two effects at play when δ_A increases. First, the average valuation of the consumers that can purchase at price p_A decreases, shifting downward p_A. Secondly,

[10]We rule out negative prices from the analysis, and a price is taken equal to zero in case its expression below is negative.

Firm B's reaction to this increased competition is to lower its price p_B, lowering further price p_A in response. The price reduction in p_B is milder since Firm B only faces the increased competition when δ_A increases, and not the changes in the average valuation of consumers it sells to. For this reason, each firm's uniform price decreases more when it has more data treated than when its competitor has more of its data treated.

We can then derive the personalized prices and the location of the indifferent consumer:

$$p_A(x) = 2t - \frac{4t}{3}\delta_B - \frac{2t}{3}\delta_A - 2tx, \quad p_B(x) = 2tx - \frac{4t}{3}\delta_A - \frac{2t}{3}\delta_B \implies \tilde{x} = \frac{1}{2} + \frac{1}{3}\delta_A - \frac{1}{3}\delta_B,$$

whereas demands in equilibrium are as follows:

$$d_A = \frac{1}{2} - \frac{2}{3}\delta_A - \frac{1}{3}\delta_B, \qquad d_B = \frac{1}{2} + \frac{2}{3}\delta_B - \frac{1}{3}\delta_A.$$

This allows us to compute the firms' profits by replacing prices and demands with their equilibrium values:

$$\pi_i = \frac{t}{2} - \frac{7}{9}\delta_i^2 t + \frac{2}{9}\delta_{-i}^2 t - \frac{4}{9}\delta_i\delta_{-i}t + \frac{2}{3}\delta_i t - \frac{2}{3}\delta_{-i}t. \tag{1}$$

Similarly, we can compute consumer surplus:

$$\begin{aligned} CS(\delta_A, \delta_B) = V &+ \int_0^{\delta_A} -2t[1 - \frac{1}{3}\delta_A - \frac{2}{3}\delta_B] - tx\mathrm{d}x + \int_{\delta_A}^{\tilde{x}} -t[1 - \frac{4}{3}\delta_A - \frac{2}{3}\delta_B] - tx\mathrm{d}x \\ &+ \int_{\tilde{x}}^{1-\delta_B} -t[1 - \frac{2}{3}\delta_A - \frac{4}{3}\delta_B] - tx\mathrm{d}x + \int_{1-\delta_B}^{1} -2t[1 - \frac{1}{3}\delta_B - \frac{2}{3}\delta_A] - tx\mathrm{d}x \\ &= V + t[-\frac{5}{4} + \frac{17}{18}\delta_A^2 + \frac{17}{18}\delta_B^2 + \delta_A\delta_B]. \end{aligned} \tag{2}$$

In the remaining of the paper we will focus on the symmetric case where $\delta_A = \delta_B = \delta$.[11] Symmetry implies that, depending on whether firms have their data treated by the broker,

[11] This is for simplicity, and considering different amounts of information to treat for the firms would yield qualitatively similar results. We provide some intuitions in Section 5.2.

they make the following profits:

$$\begin{cases} \pi_A = \pi_B = \frac{t}{2} \quad \text{if firms do not have their data treated,} \\ \pi_A = \pi_B = \hat{\pi} = \frac{t}{2} - \delta^2 t \quad \text{if both firms have their data treated,} \\ \overline{\pi} = \frac{t}{2} + \frac{\delta t}{3}[2 - \frac{7\delta}{3}] \ \& \ \underline{\pi} = \frac{t}{2} - \frac{2\delta t}{3}[1 - \frac{\delta}{3}] \text{ if only one firm (with profits } \overline{\pi}\text{) treats its data.} \end{cases}$$

As discussed above, when a broker treats the data of a firm, it enhances the firm's ability to extract surplus from the closest consumers by personalizing prices. This increases the profits of the firm, but also intensifies competition, harming the competitor. Hence, a firm benefits from having its data treated at the expense of its competitor. This competitive effect of information is so strong that when both firms have their data treated, they achieve lower profits than if they had remained uninformed. Data is unilaterally pro-competitive for the industry overall (de Cornière and Taylor, 2024a), and we can rank the profits as follows:

$$\overline{\pi} > \frac{t}{2} > \hat{\pi} > \underline{\pi}$$

5 Monopoly Case

Now that we know how the downstream market equilibrium depends on the information firms can use, we can characterize the data broker's strategy. We will study two main elements: first, with whom a broker will contract, and second, at which price the information will be sold.

In this section, we analyze the optimal selling strategy of a monopolist broker. We will see how this strategy changes when the learning ability of the broker increases and when the broker can commit on its future actions. We then analyze competing brokers in Section 6.

The broker has three options (on top of doing nothing): selling to A only, selling to B only, and selling to both firms. As a benchmark, we consider a broker who can commit in the first period to sell only to Firm A, and not to Firm B later. We then detail how the lack of commitment power impacts the selling strategy of the broker.

5.1 Selling Strategies with Commitment

Selling to one firm. When the broker can commit to sell information only to Firm A, this firm knows that Firm B won't be able to purchase information if Firm A accepts the broker's offer. But what happens if Firm A turns down the offer? Then, the broker will have incentives to contract with Firm B, at least if there are some gains to do so. This means that Firm A's willingness to pay for information is the difference between its profits upon accepting the broker's offer and its profits if it declines the offer.

Suppose it is profitable for the broker to sell information to Firm B in case Firm A declines the offer. Using the notations introduced in the previous section, Firm A's profits are equal to $\overline{\pi}$ if it buys information, and to $\underline{\pi}$ if it stays uninformed and competes with Firm B having information. This implies that the broker can set a price equal to:

$$\overline{\rho} = \overline{\pi} - \underline{\pi},$$

and therefore, it makes the following profit net of the treatment cost:

$$\Pi_A = \overline{\rho} - c\delta.$$

This reasoning assumes that Firm B buys information if Firm A declines the offer of the broker. But is this always the case? The profits of Firm B in this configuration are equal to $\overline{\pi}$ if it buys information, and to the usual Hotelling profit equal to $\frac{t}{2}$ otherwise. This means the broker could charge Firm B a price equal to:

$$\hat{\rho} = \overline{\pi} - \frac{t}{2},$$

and therefore, make the following profit net of the treatment cost:

$$\Pi_B = \hat{\rho} - c\delta.$$

Hence, the broker can sell to Firm B only if Π_B is greater than zero, which, denoting $\tilde{c} = \frac{c}{t}$, is equivalent to:

$$\delta < \delta_1 = \frac{9}{7}\left(\frac{2}{3} - \tilde{c}\right) \tag{C1}$$

When this condition is not satisfied, the broker cannot sell to Firm B if Firm A declines the offer. This implies that the maximum price the broker can charge Firm A is also $\hat{\rho}$.

Hence, selling data to Firm A is not profitable either since (C1) is not satisfied, which implies $\hat{\rho} < c\delta$. This means that outside (C1), none of the firms will have their data treated.

Selling to both firms. In this case, each firm will earn the same profit, defined above as $\hat{\pi}$. The price the broker can charge for information equals the willingness to pay of each firm, i.e., the difference of profits if a firm buys information, $\hat{\pi}$, and if it remains uninformed while facing an informed competitor, $\underline{\pi}$. Hence, the broker can charge each firm a price equal to:

$$\underline{\rho} = \hat{\pi} - \underline{\pi}.$$

The broker incurs a cost $c\delta$ to treat the data of Firm A. It benefits from the learning effect and can then treat the data of Firm B at a cost $c\delta(1 - \alpha\delta)$. Overall, the broker makes the following total profits:

$$\Pi_{AB} = 2\underline{\rho} - c\delta - c\delta(1 - \alpha\delta).$$

Selling to only Firm B yields lower profits than selling to both firms when $\Pi_B < \Pi_{AB}$, which is equivalent to:

$$\delta < \delta_2(\alpha) = \frac{3}{5 - 3\alpha\tilde{c}}\left(\frac{2}{3} - \tilde{c}\right) \tag{C2}$$

Comparing profits in the different strategies gives us the ranking in Lemma 1.

Lemma 1. *For any α, $\Pi_A > \max\{\Pi_B, \Pi_{AB}\}$, and $\Pi_B < \Pi_{AB} \iff \delta < \delta_2(\alpha)$.*

To understand this result, we can compare prices in the different strategies:

$$\overline{\rho} > \max\{\hat{\rho}, 2\underline{\rho}\}.$$

Selling exclusively to Firm A yields the highest price of information according to two effects. First, $\overline{\rho}$ is greater than $\hat{\rho}$ because if Firm A declines the offer of the broker, Firm B will acquire information. This exerts a threat on Firm A, shifting upward its willingness to pay for information. On the contrary, if Firm A declines the offer, Firm B faces a situation where, if it does not buy information, both firms remain uninformed. Since there is no possibility for the broker to go back to Firm A after making an offer to Firm B, it cannot exert a threat on Firm B, and charges a lower price for information than when selling exclusively to Firm A.

Secondly, compared to a situation where a firm buys information in exclusivity, its gains from buying are sharply reduced when the competitor also buys information, according to a "discouragement effect". We can observe this effect by considering the difference of willingness to pay of a firm when it buys information in exclusivity and when its competitor also buys information:

$$\overline{\rho} - \hat{\rho} = \frac{2\delta t}{3} + \frac{2\delta^2 t}{9}.$$

The discouragement effect becomes stronger as δ increases. Because of this effect, it is optimal for the broker to sell information to only Firm A instead of both firms, even with a second-period cost equal to zero. It also means that the learning parameter is irrelevant since the broker only contracts with one firm.

Proposition 1.

When the broker can commit on its future actions, its optimal strategy is to sell exclusively to Firm A if $\delta < \delta_1$. The broker does not sell information to any firm otherwise.

5.2 Selling Strategies without Commitment

The broker may not be able to commit on its future behavior, for instance if exclusive contracts are prohibited by regulation. This lack of commitment impacts the broker's ability to implement its optimal selling strategy.

To understand the inefficiencies caused by the broker's lack of commitment, suppose the broker sells information to Firm A in the first period. In the second period, the broker has incentives to treat Firm B's data as long as the second-period surplus is positive.

To see when this is the case, we first derive the maximum price Firm B is ready to pay to have its data treated. This price is the difference between the profits of Firm B if it buys information and if only Firm A has its data treated $\underline{\rho} = \hat{\pi} - \underline{\pi}$. The second-period surplus is then given by $\underline{\rho} - c\delta(1 - \alpha\delta)$. Hence, the broker has incentives to sell information to Firm B if this surplus is positive, i.e., when:

$$\delta < \delta_{AB}(\alpha) = \frac{9}{11 - 9\alpha\tilde{c}} \left(\frac{2}{3} - \tilde{c} \right) \quad \text{(C3)}$$

Under this condition, the broker will have incentives to break its word and sell information to Firm B. This leads to the following result.

Lemma 2. *For any $\delta < \delta_{AB}(\alpha)$, the profits of the broker are lower in the no-commitment case than when it can commit.*

The lack of commitment prevents the broker from selling exclusively to Firm A, which is its profit-maximizing strategy. Indeed, Firm A can anticipate the incentives of the broker to sell to Firm B in the second stage when $\delta < \delta_{AB}(\alpha)$. This implies that the willingness to pay of Firm A for information is no longer $\overline{\rho}$, since Firm A knows that Firm B will buy information. In other words, the discouragement effect lowers the willingness to pay of Firm A, reducing the broker's potential profits.

In this case, the broker must choose between either selling to both firms – which happens when $\delta \in [0, \delta_2(\alpha)]$ – or to Firm B – which happens when $\delta \in [\delta_2(\alpha), \delta_{AB}(\alpha)]$.

On the contrary, when $\delta \in [\delta_{AB}(\alpha), \delta_1]$, the broker can still sell information exclusively to Firm A, even without an explicit exclusive contract, because then, selling information to Firm B would not cover the treatment cost and is therefore not profitable.[12] Hence, when the learning effect (measured by α) is not too high, there is still a way for the broker to sell (at least for some level of δ) its service exclusively.

Proposition 2 summarizes this discussion.

Proposition 2.

When the broker cannot commit to exclusive deals, the equilibrium is such that:

- *when $\delta \in [0, \delta_2(\alpha)]$, the broker sells to both firms at a common price $\underline{\rho} = \hat{\pi} - \underline{\pi}$;*
- *when $\delta \in [\delta_2(\alpha), \min\{\delta_1, \delta_{AB}(\alpha)\}]$, the broker sells only to Firm B at price $\hat{\rho} = \overline{\pi} - \frac{t}{2}$;*
- *when $\delta \in [\min\{\delta_{AB}(\alpha), \delta_1\}, \delta_1]$, the broker sells only to Firm A at a price $\overline{\rho} = \overline{\pi} - \underline{\pi}$;*
- *when $\delta \in [\delta_1, \frac{1}{2}]$, neither firm has its data treated.*

We represent the different areas of parameter value using Figure 2.[13]

[12] Note that this happens when $\alpha\tilde{c} < \frac{4}{9}$, since only then $\delta_{AB}(\alpha) < \delta_1$.

[13] We can see that if the broker sells different amounts of information to firms, the thresholds will be shifted upwards or downwards depending on the values of δ_A and δ_B, but the equilibria in the different zones we have characterized will remain the same.

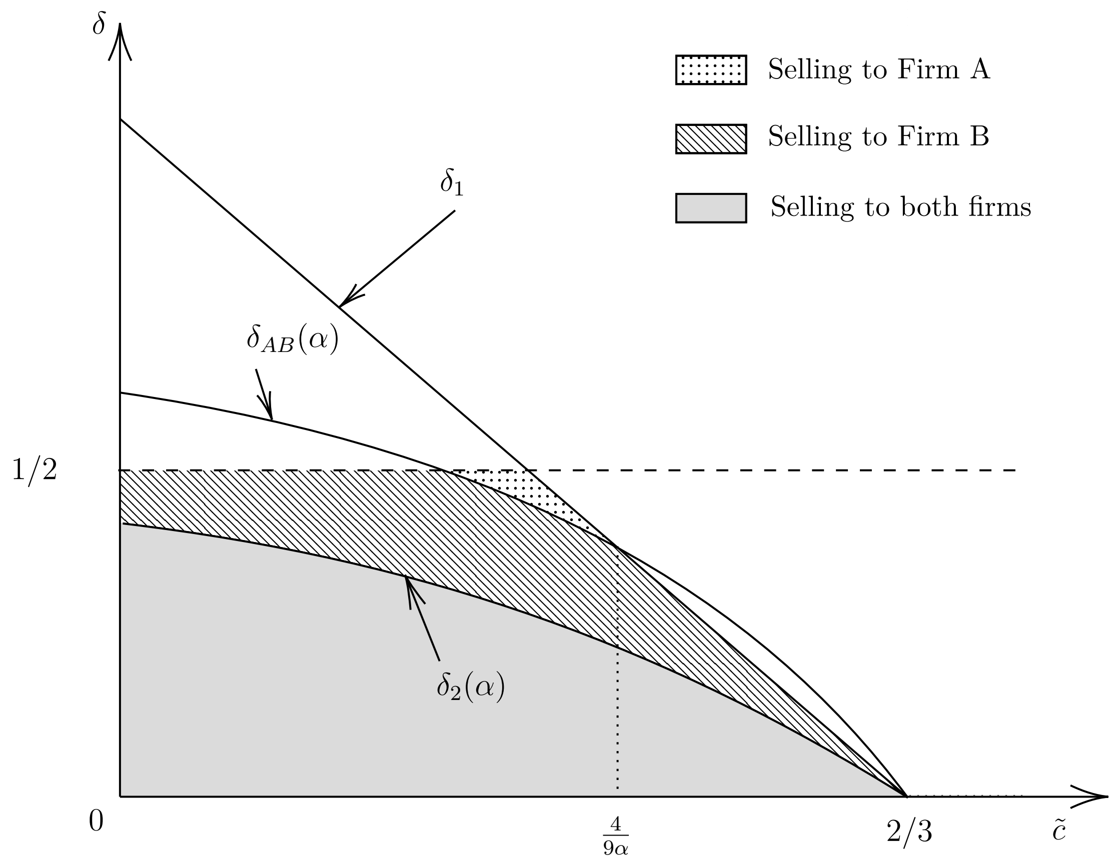


Figure 2: Selling strategies with $\alpha = 1$

The amount of data to treat δ captures different effects impacting jointly the ability of the broker to sell information to firms and the price of information. First, a large amount of data to treat induces a high treatment cost. Secondly, and different from previous models (Cabral and Riordan, 1994; Besanko et al., 2010; Sweeting et al., 2022), the amount of information sold by the broker also impacts the willingness to pay of the firms through the discouragement effect, which becomes stronger as the amount of data treated increases. This new mechanism arises from introducing competition between information buyers.

A reduction in the amount of data to treat (i.e. decreasing δ) has two opposite impacts on the ability of the broker to sell information to firms. First, it softens the discouragement effect; secondly, it also lowers the treatment cost. To understand how these effects impact the ability of the broker to sell information to firms, consider a case where $\delta \in [\delta_{AB}(\alpha), \delta_1]$ (so a point in the dotted area). There, the discouragement effect is strong, and the broker cannot sell information to Firm B as it already sells to Firm A.

As δ decreases the discouragement effect becomes weaker. Moreover, the learning effect tends to dominate the discouragement effect, allowing the broker to benefit from treating the data of Firm B if it has already treated Firm A's data. However, as long as δ is large enough (i.e. larger than $\delta_2(\alpha)$) the discouragement effect is still strong enough to prevent Firm A from having its data treated.

For small values of δ, i.e. when $\delta < \delta_2(\alpha)$ in the gray area, the reduction of the discouragement effect and of the treatment cost makes it then profitable for the broker to sell information to both firms.

Corollary 1. *A high treatment cost δc acts as a substitute for exclusivity commitments, enabling the broker to maximize profits by selling exclusively to Firm A. Conversely, a low treatment cost can prevent the broker from selling in exclusivity to Firm A.*

The impact of learning on the selling strategies and the profits of the broker depends on the amount of data to treat δ. Consider a value δ_0 located in the dotted area in Figure 2, where the broker sells information to Firm A.

An increase in the learning effect will change the identity of the firm buying information after a point, and the broker will sell information to Firm B instead of Firm A. This reduces the profits of the broker as they are maximized by selling exclusively to Firm A.

A further increase in the learning effect will then allow the broker to sell information to both firms instead of only Firm B. Hence, the learning effect expands the data market by allowing the broker to sell to more firms. When selling to both firms, the profits of the broker increase with α but remain below the first-best profits.

Lemma 3 provides conditions for these changes of selling strategies to take place.

Lemma 3.

- *For any $\delta_0 \in [\delta_{AB}(0), \delta_1]$, there exists $\hat{\alpha}$ such that the broker sells information to Firm A when $\alpha < \hat{\alpha}$ and to Firm B when $\alpha > \hat{\alpha}$.*
- *For any $\delta_0 \in [\delta_2(0), \max\{\delta_1, \frac{2}{5}\}]$, there exists $\hat{\hat{\alpha}} > \hat{\alpha}$ such that the broker sells information to one firm $\alpha < \hat{\hat{\alpha}}$ and to both firms when $\alpha > \hat{\hat{\alpha}}$.*

These changes of selling strategies lead to significant variations in the profits of the broker. In particular, our results highlight the potential loss when a stronger learning effect allows the broker to treat the data of more firms, but undermines its ability to ensure Firm A it will not sell data to Firm B. The profits of the broker when α increases are denoted by Π, and depicted in Figure 3.

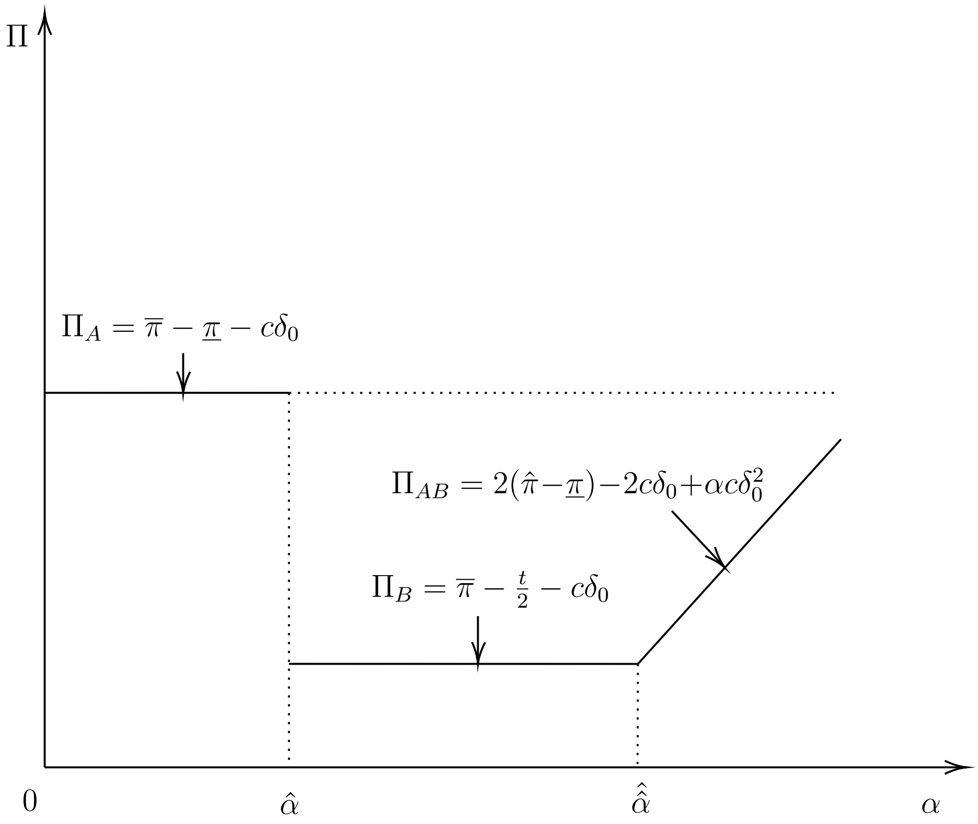


Figure 3: Profits of the broker as a function of α for $\delta_0 \in [\delta_2(0), \min\{\delta_1, \frac{2}{5}\}]$

We summarize the impacts of an increase in the strength of the learning effect on the profits of the firms in Proposition 3.

Proposition 3.

An increase in the learning effect:

- *may decrease the broker's profits when it sells to only one firm;*
- *increases its profits when it sells information to both firms;*
- *may expand the data market by allowing the broker to sell information to both firms.*

While previous literature has shown that learning-by-doing can lower the profits of competing suppliers (Cabral and Riordan, 1994), we show here that learning can also have adverse effects on the profits of an upstream monopolist. This profit loss occurs because buyers anticipate the future selling decision of the broker when the learning effect changes downstream competition, preventing the broker from selling information exclusively to Firm A.

6 Multiple Brokers

We now consider two competing brokers, denoted Broker $\overline{\alpha}$ and Broker $\underline{\alpha}$, differentiated only by their ability to learn from data: $\overline{\alpha} \geq \underline{\alpha}$. At each stage of the game, brokers compete in a (differentiated) Bertrand way to sell information to firms. We first analyze competition between brokers in Section 6.1. We then consider the possibility of data sharing between brokers in Section 6.2.

6.1 Competing Brokers

The impact of competition between brokers on market equilibrium depends on the firms' willingness to purchase information from a broker. There are three cases of interest depending on the value of δ. First, competition between brokers can induce a classical effect of downward pressure on prices. Secondly, a stronger learning effect can allow the most efficient broker to achieve monopoly profits, as in the absence of competition. Thirdly, having the strongest learning effect can also disadvantage a broker, yielding an equilibrium where the least efficient broker achieves monopoly profits while its more efficient competitor makes zero profits. The three following configurations showcase these effects, with Broker $\overline{\alpha}$ making positive profits in the first two cases, and zero profits in the third case.

Competition lowers the prices of information when both firms can purchase information from either one or the other broker, a scenario that occurs when δ is below $\delta_{AB}(\underline{\alpha})$. In this context, a broker has three strategic options: sell to Firm B only; sell to both firms; or do not sell. Broker $\tilde{\alpha}$ (with $\tilde{\alpha} \in \{\underline{\alpha}, \overline{\alpha}\}$) can achieve positive profits by selling to both firms if the profits net of the treatment costs are positive, $2[\hat{\pi} - \underline{\pi}] - c\delta - c\delta(1 - \tilde{\alpha}\delta) > 0$, which is equivalent to:

$$\delta < \delta_c(\tilde{\alpha}) = \frac{1}{\frac{11}{9} - \frac{\tilde{c}\alpha}{2}} \left(\frac{2}{3} - \tilde{c}\right) < \delta_{AB}(\tilde{\alpha})$$

Hence, we focus here on $\delta \in [0, \delta_c(\underline{\alpha})]$, such that both brokers can sell to both firms.

In the second stage, one of the brokers has treated the data of Firm A and would incur a treatment cost equal to $c\delta(1 - \tilde{\alpha}\delta)$ for treating Firm B's data. Let ρ_B be the price this broker charges Firm B for information.

If the competing broker can make positive profits by treating the data of Firm B at a cost equal to $c\delta$, it exerts a competitive pressure on Broker $\tilde{\alpha}$. Hence, Broker $\tilde{\alpha}$ sells information to Firm B at a price $\rho_B = c\delta$, and incurs a treatment cost equal to $c\delta(1-\alpha\delta)$. Competition in the second stage takes place when:

$$\delta < \delta_c(0) = \frac{9}{11}\left(\frac{2}{3} - \tilde{c}\right)$$

If the competing broker cannot make positive profits by treating the data of Firm B at a cost equal to $c\delta$, Broker $\tilde{\alpha}$ can sell to Firm B at a price $\rho_B = \hat{\pi} - \underline{\pi}$. There will be no competition in the second stage when $\delta \in [\delta_c(0), \delta_c(\underline{\alpha})]$.

In the first stage, both brokers face identical treatment costs. Anticipating further profits when selling to Firm B, each broker is ready to lower its price below the treatment cost, down to a value $\rho_A(\tilde{\alpha})$ for which losses in the first stage equal its anticipated gains in the second stage: $\rho_A(\tilde{\alpha}) - c\delta + \rho_B - c\delta(1-\tilde{\alpha}\delta) \geq 0 \implies \rho_A(\tilde{\alpha}) \geq c\delta - \rho_B + c\delta(1-\tilde{\alpha}\delta)$.

Hence, Broker $\underline{\alpha}$ sets a price of $\rho_A(\underline{\alpha}) = c\delta - \rho_B + c\delta(1-\tilde{\alpha}\delta)$ and does not sell information. In contrast, Broker $\overline{\alpha}$ sets a price just below $c\delta - \rho_B + c\delta(1-\underline{\alpha}\delta)$, sells information, and makes total profits equal to:

$$\Pi_{\overline{\alpha}} = \underbrace{c\delta - \rho_B + c\delta(1-\underline{\alpha}\delta)}_{\text{Profits in the first stage}} + \underbrace{\rho_B - c\delta(1-\overline{\alpha}\delta)}_{\text{Profits in the second stage}} = c\delta^2(\overline{\alpha} - \underline{\alpha}).$$

These profits depend on the difference in learning effects between the two brokers, $\overline{\alpha} - \underline{\alpha}$, which captures the intensity of upstream competition. In the special case where $\overline{\alpha} = \underline{\alpha}$, brokers are identical and make zero profits. Because of competition, these profits are naturally below what Broker $\overline{\alpha}$ would earn as a monopolist. Notably, $\Pi_{\overline{\alpha}}$ increases with c, as the relative competitive advantage of Broker $\overline{\alpha}$ over Broker $\underline{\alpha}$ grows when the treatment cost increases.

Broker $\overline{\alpha}$ can also achieve monopoly profits thanks to its strong learning effect when $\delta \in [\delta_c(\underline{\alpha}), \min\{\delta_c(\overline{\alpha}), \delta_{AB}(\underline{\alpha})\}]$. In this interval, since $\delta > \delta_c(\underline{\alpha})$, Broker $\underline{\alpha}$ cannot make positive profits by selling to both firms, nor can it sell information exclusively to Firm A as $\delta < \delta_{AB}(\underline{\alpha})$. Therefore, Broker $\underline{\alpha}$ either sells only to Firm B, or does not sell information to anyone.

However, selling exclusively to Firm A is profitable for Broker $\overline{\alpha}$, as it does not face competitive pressure from Broker $\underline{\alpha}$. In the second stage, once Firm A's data have been treated, Broker $\underline{\alpha}$ cannot make positive profits by selling to Firm B (since $\delta > \delta_c(\underline{\alpha}) > \delta_c(0)$). Hence, selling to Firm B is only profitable for Broker $\overline{\alpha}$ who does not face competitive pressure in the second stage either, and is able to make second-stage profits equal to $\hat{\pi} - \underline{\pi} - c\delta$. Overall, Broker $\overline{\alpha}$ makes monopoly profits, given by:

$$\Pi_{\overline{\alpha}} = 2(\hat{\pi} - \underline{\pi}) - c\delta - c\delta(1 - \alpha\delta).$$

Broker $\underline{\alpha}$ can make monopoly profits thanks to its *weaker* learning effect. This outcome occurs when $\delta \in [\max\{\delta_c(\overline{\alpha}), \delta_{AB}(\underline{\alpha})\}, \delta_{AB}(\overline{\alpha})]$. In this range, Firm A is willing to purchase information only from Broker $\underline{\alpha}$, since doing so prevents Firm B from purchasing data. On the contrary, Firm A does not buy information from Broker $\overline{\alpha}$ as doing so would allow Firm B to purchase information too, sharply intensifying competition in the product market.

Hence, Broker $\underline{\alpha}$ can sell information to Firm A and achieve *monopoly profits* equal to $\overline{\pi} - \underline{\pi} - c\delta$, as characterized in Section 5 where the broker sells information as a monopolist. This result holds even though both brokers incur the same treatment cost at this stage. Therefore, having a weaker learning effect provides Broker $\underline{\alpha}$ with a significant competitive advantage over Broker $\overline{\alpha}$.

For the remaining values of the amount of data to treat δ,[14] competition between brokers lowers the price they set, but does not affect which firm purchases information.

Proposition 4. *In the case of competing brokers:*

- *for intermediate values of δ, one of the brokers – either Broker $\overline{\alpha}$ or Broker $\underline{\alpha}$ – does not face competitive pressure and can sell information as a monopolist;*
- *for small or high values of δ, introducing competition between brokers lowers the price of information but does not change which firm buys information.*

Previous literature has shown how learning-by-doing can give competing suppliers incentives to engage in predation practices and price-undercutting strategies (Cabral and Riordan, 1994; Besanko et al., 2014). In contrast, by introducing competition between buyers, we show that learning can enable an upstream monopolist to escape Bertrand competition and achieve monopoly profits. This effect can benefit both types of brokers, including the broker with the weakest learning effect.

[14]For $\delta \in [\delta_{AB}(\underline{\alpha}), \delta_c(\overline{\alpha})]$ where Broker $\overline{\alpha}$ can to sell information to both firms, while Broker $\underline{\alpha}$ can sell information to only Firm A; for $\delta \in [\delta_{AB}(\overline{\alpha}), \delta_1]$ where both brokers can sell only to Firm A; and for $\delta \in [\delta_c(\overline{\alpha}), \delta_{AB}(\underline{\alpha})]$ where both brokers can sell only to Firm B. We solve these cases in Appendix A.2.

6.2 Data Sharing among Brokers

Consider now a broker that has treated the data of Firm A and assume this broker can share this data with its competitor, potentially in exchange for a monetary transfer. Once data is shared, the broker that has received the data can use it to improve its technology, i.e., reduce its treatment cost through a learning effect. For simplicity, we assume that this learning effect occurs at no additional cost.[15] Following data sharing, both brokers compete to sell information to Firm B, with both brokers now benefiting from the learning effect.

Data sharing is profitable if the broker receiving the data is better at learning than its competitor, allowing it to achieve higher profits when selling information to Firm B thanks to this stronger learning effect. Profitable data-sharing practices may take place in two cases: when both firms purchase information and when only Firm A purchases information.

First, consider the case where $\delta \in [0, \delta_2(\underline{\alpha})]$, so that both firms purchase information. In the second stage, one broker faces a treatment cost equal to $c\delta$ and the other one to $c\delta(1-\tilde{\alpha}\delta)$ with $\tilde{\alpha} \in \{\underline{\alpha}, \overline{\alpha}\}$. If both brokers compete à la Bertrand, the profits of the broker benefiting from the learning effect are equal to:

$$\Pi = c\tilde{\alpha}\delta^2$$

This broker may also choose to sell its data to its competitor, who would then benefit from the learning effect. There are two cases to consider, depending on whether Broker $\underline{\alpha}$ or $\overline{\alpha}$ has access to data they might share. Let us also denote $\tilde{\tilde{\alpha}} \in \{\underline{\alpha}, \overline{\alpha}\}$ with $\tilde{\alpha} \neq \tilde{\tilde{\alpha}}$.

Now, suppose Broker $\tilde{\alpha}$ decides to share data with Broker $\tilde{\tilde{\alpha}}$. By leveraging this data, Broker $\tilde{\tilde{\alpha}}$ can reduce its treatment cost, allowing it to charge a price for information equal to $c\delta(1-\tilde{\alpha}\delta)$, and make profits:

$$\Pi_{\tilde{\tilde{\alpha}}} = c\delta(1-\tilde{\alpha}\delta) - c\delta(1-\tilde{\tilde{\alpha}}\delta) = c\delta^2(\tilde{\tilde{\alpha}} - \tilde{\alpha})$$

These profits are positive only if $\tilde{\tilde{\alpha}} > \tilde{\alpha}$. Therefore, there is room for data sharing only if Broker $\underline{\alpha}$ shares its data with Broker $\overline{\alpha}$.

Otherwise, if Broker $\overline{\alpha}$ has treated the data of Firm A and benefits from a cost reduction due to the learning effect, it gains nothing from sharing the data with Broker

[15] Of course, learning can be costly for the receiver of the data. Introducing a learning cost would not change our qualitative insights.

$\underline{\alpha}$. Sharing would allow Broker $\underline{\alpha}$ to lower its price and increase the competitive pressure on Broker $\overline{\alpha}$ when selling to Firm B. This would lower the profits of the brokers, making data sharing unprofitable in this scenario.

We thus focus on Broker $\underline{\alpha}$ that has treated the data of Firm A, and we consider its incentives to share data with Broker $\overline{\alpha}$ in exchange for a monetary transfer. Broker $\underline{\alpha}$ has two options: it can either choose not to share data and make profits of $c\underline{\alpha}\delta^2$; or it can share the data with Broker $\overline{\alpha}$ in exchange for a transfer equal to $c\delta^2(\overline{\alpha} - \underline{\alpha})$, in which case Broker $\underline{\alpha}$ does not sell information to Firm B and makes no further profits. Hence, Broker $\underline{\alpha}$ sells its data if: $c\delta^2(\overline{\alpha} - \underline{\alpha}) > c\underline{\alpha}\delta^2$, which holds if and only if $\overline{\alpha} > 2\underline{\alpha}$.

Anticipating potential gains from selling its data in the second period, Broker $\underline{\alpha}$ will compete more aggressively in the first period, potentially lowering its first-period price $\rho_A(\underline{\alpha})$ down to a point where its loss in the first stage equals its anticipated gains from data sharing in the second stage:

$$\rho_A(\underline{\alpha}) - c\delta + c\delta^2(\overline{\alpha} - \underline{\alpha}) \geq 0 \implies \rho_A(\underline{\alpha}) \geq c\delta(1 - \delta(\overline{\alpha} - \underline{\alpha}))$$

In turn, Broker $\overline{\alpha}$ can sell information to Firm A by setting a price just below $\rho_A(\underline{\alpha})$. In this case, Broker $\overline{\alpha}$ makes (negative) profits equal to $c\delta^2(\underline{\alpha} - \overline{\alpha})$ when selling to Firm A.

We have shown that Broker $\overline{\alpha}$ makes second-period profits equal to $c\overline{\alpha}\delta^2$ so its total profits over both periods are equal to

$$\Pi^S_{\overline{\alpha}} = \underbrace{c\delta^2(\underline{\alpha} - \overline{\alpha})}_{\substack{\text{Profits in the} \\ \text{first stage}}} + \underbrace{c\overline{\alpha}\delta^2}_{\substack{\text{Profits in the} \\ \text{second stage}}} = c\underline{\alpha}\delta^2 < \Pi_{\overline{\alpha}}$$

Hence, the ability for brokers to share data intensifies competition in the first stage, by increasing the value of Firm A's data for Broker $\underline{\alpha}$. This reduces the profits of Broker $\overline{\alpha}$.

Now, consider the case where only one firm purchases information. Here, data sharing between brokers becomes relevant only if Broker $\underline{\alpha}$ sells information to Firm A, which occurs when $\delta \in [\max\{\delta_{AB}(\underline{\alpha}), \delta_2(\overline{\alpha})\}, \delta_{AB}(\overline{\alpha})]$. In this interval, selling information to Firm B is not profitable for Broker $\underline{\alpha}$, but it would be profitable for Broker $\overline{\alpha}$ if it could benefit from the learning effect.

By selling its data to Broker $\overline{\alpha}$, Broker $\underline{\alpha}$ enables the sale of information to Firm B. However, anticipating this effect, Firm A will not agree to engage with Broker $\underline{\alpha}$ (and provide its data) to prevent data sharing with Broker $\overline{\alpha}$, and Firm A will remain uninformed. Consequently, brokers end up competing à la Bertrand with identical treatment costs when selling to Firm B, and make zero profits.

These negative effects of data sharing on the brokers' profits can be offset if Firm A holds property rights over its data (see Dosis and Sand-Zantman (2023) on data ownership). When brokers sell to both firms, data sharing lowers the price Firm A pays for information. In this case, Firm A benefits from data sharing among brokers and does not enforce its data property rights. On the contrary, Firm A will enforce its property rights and prevent data sharing between brokers when doing so allows Broker $\underline{\alpha}$ to sell information exclusively to Firm A.[16]

Proposition 5.

When brokers can share data:

- *For low values of δ, brokers compete more fiercely than when data sharing is not feasible. As Firm A benefits from this enhanced competition, it does not enforce property rights over its data.*
- *For intermediate values of δ, Broker $\underline{\alpha}$ sells information to Firm A who enforces its property rights to prevent Broker $\overline{\alpha}$ from selling information to Firm B.*
- *Otherwise, brokers do not share data and the possibility of data sharing does not change the equilibrium outcomes.*

7 Extensions

We analyze several extensions to probe the robustness of our baseline model. Namely, we allow for renegotiation in Section 7.1, we consider Firm A that is unaware of the learning effect in Section 7.2, and we analyze uncertainty over the type of the broker in Section 7.3. Throughout this section, we consider a monopolist broker, but the main qualitative insights that we derive also apply when brokers compete.

7.1 Alternative Timing with Renegotiation

The baseline model focuses on two stages where the broker makes an offer successively to Firm A and Firm B. Here, we allow for renegotiation if Firm A initially declines the offer of the broker. This setting is indeed used by recent articles on learning-by-doing

[16]This echoes the results of Hunold and Shekhar (2022) who show that a buyer will not share cost-reducing innovation with a monopolist supplier if doing so benefits competing buyers in the future.

(Besanko et al., 2010, 2014; Sweeting et al., 2022). In this section, the broker can always make follow-up offers to a firm that has previously declined, even multiple times.

Renegotiation changes the equilibrium derived in the baseline model only if Firm A has declined the offer and Firm B has accepted it. Indeed, in this case, renegotiation means that the initial decision of Firm A remains unaffected. However, at the time it makes an offer to Firm B, the broker now has the possibility to make a subsequent offer to Firm A. Hence, Firm B faces the same decision as Firm A initially, leading Firm B to also find it unprofitable to have its data treated.

Hence, renegotiation generates a situation where each firm is willing to wait for its competitor to buy information before buying itself. As a result, none of the firms have their data treated with renegotiation. This outcome arises for values in $\delta \in [\delta_2(\alpha), \delta_{AB}(\alpha)]$. Hence, an increase in the learning effect induces a contraction of the data market when renegotiation is possible.

Proposition 6.

When renegotiation is possible, the presence of learning effects may decrease the amount of information treated by the broker.

Proposition 6 has important implications, as stronger learning effects can entirely suppress the incentives of the firms to have their data treated. Hence, when the broker benefits from a learning effect, it may become unable to sell information to firms, and achieves zero profits.

7.2 Myopic Firms

Firm A may not be aware that the broker can learn from data and reduce its future treatment cost. The outcome in this case will be different from the one derived in Section 5 only for $\delta \in [\delta_{AB}(0), \delta_{AB}(\alpha)]$. There, when Firm A is aware of the learning effect, it anticipates that Firm B will buy information thanks to the learning effect. This leads to a low willingness to pay of Firm A for information because of the discouragement effect, and it is not profitable for the broker to treat the data of Firm A.

When Firm A is not aware of the learning effect, it believes that the broker will not treat the data of Firm B if it buys information since $\delta \geq \delta_{AB}(0)$. Firm A's willingness to pay for information is then the difference between its profits with exclusive information, $\overline{\pi}$, and those when it remains uninformed and faces Firm B informed, $\underline{\pi}$. Hence, the profits of the broker when selling to Firm A are:

$$\overline{\pi} - \underline{\pi} - c\delta$$

Since Firm A believes that Firm B will not buy information, Firm A is not subject to the discouragement effect, leading to a higher willingness to pay for information than when it is aware of the learning effect. In turn, these profits are positive since $\delta < \delta_1$, and the broker finds it profitable to treat the data of Firm A. Additionally, as $\delta < \delta_{AB}(\alpha)$, the broker also treats the data of Firm B. Overall, for $\delta \in [\delta_{AB}(0), \delta_{AB}(\alpha)]$, Firm A has its data treated because it does not anticipate that the broker, benefiting from lower treatment costs thanks to the learning effect, will also sell information to Firm B.

Lemma 4. *For $\delta \in [\delta_{AB}(0), \delta_{AB}(\alpha)]$, the broker treats the data of both firms when Firm A is not aware of the learning effect.*

Comparative statics. Let us look at the impact of an increase in the learning effect on the strategies of the broker. We have shown that an increase in α increases the threshold $\delta_{AB}(\alpha)$. When Firm A does not anticipate the learning effect, the broker treats the data of Firm A if $\delta > \delta_{AB}(\alpha)$ and of both firms if $\delta_{AB}(\alpha) > \delta$. Therefore, a stronger learning effect expands the data market by relaxing the conditions under which the broker can treat the data of Firm B.

Proposition 7.

When Firm A is myopic, an increase in the value of α expands the data market by allowing the broker to sell to both firms instead of only Firm B when $\delta \in [\delta_{AB}(0), \delta_{AB}(\alpha)]$.

7.3 Asymmetric Information on the Learning Effect

The baseline model assumes that Firm A has complete information over the broker's type α when purchasing information, and can therefore perfectly anticipate whether the broker will later sell to Firm B. Yet, a firm may not perfectly know to which extent a broker can learn from data, even if it can have a good idea of the value of these improvements.

In previous section, we have considered a scenario where Firm A is unaware of the ability of the broker to learn from data. We now shift our focus to a situation where firms recognize that the broker can learn from data, but have incomplete information regarding the value of the learning effect of the broker. To streamline the analysis, we assume that the data broker can be of two types, each characterized by different learning effects: $\alpha \in \{\underline{\alpha}, \overline{\alpha}\}$, $\overline{\alpha} \geq \underline{\alpha}$, with λ being the probability that the broker is of type $\underline{\alpha}$.

Since the price the broker can charge to treat Firm B's data does not depend on its type, we focus here on the sale of information to Firm A. Uncertainty over the type of the broker becomes significant only if Firm A is willing to buy information from one type of broker but not from the other. As before, $\delta_{AB}(\overline{\alpha})$ and $\delta_{AB}(\underline{\alpha})$ denote the values of threshold $\delta_{AB}(\alpha)$ for brokers of types $\overline{\alpha}$ and $\underline{\alpha}$. Similarly, $\delta_2(\underline{\alpha})$ and $\delta_2(\overline{\alpha})$ defined earlier characterize the value of threshold $\delta_2(\alpha)$ respectively for brokers of type $\overline{\alpha}$ and $\underline{\alpha}$.

Hence, there are three cases to consider, depending on the relative positions of the different thresholds.

Consider first the case where $\delta \in [\max\{\delta_{AB}(\underline{\alpha}), \delta_2(\overline{\alpha})\}, \delta_{AB}(\overline{\alpha})]$. In this scenario, if Firm A believes the broker is $\underline{\alpha}$-type, then Firm A is willing to pay $\overline{\pi} - \underline{\pi}$ for information, as it knows that purchasing this information will prevent the broker to sell to Firm B in the second stage. If Firm A believes instead that the broker is $\overline{\alpha}$-type, then it accepts to pay at most $\hat{\pi} - \underline{\pi}$ for information, anticipating the broker will sell to Firm B in the second period.

When Firm A does not know the broker's type, we can show that either (i) both types of brokers sell to Firm A at the same (pooling) price, with only the most efficient broker selling to Firm B, (ii) or Broker $\overline{\alpha}$ sells to both firms while Broker $\underline{\alpha}$ sells to only Firm B.

To see that, let us denote by ρ^P the pooling price offered in the first period by both brokers. This offer will be accepted by Firm A if it generates as least as much average profits than refusing it, in which case the broker, regardless of its type, will sell to Firm B. The condition for accepting this offer is then

$$\lambda[\overline{\pi} - \rho^P] + (1-\lambda)[\hat{\pi} - \rho^P] \geq \underline{\pi} \iff \rho^P \leq \lambda[\overline{\pi} - \underline{\pi}] + (1-\lambda)[\hat{\pi} - \underline{\pi}]. \tag{3}$$

Let us now look at the incentives of the brokers. First, the least efficient Broker $\underline{\alpha}$ can either sell to only Firm A at the pooling price ρ^P, or to only Firm B at price $\overline{\pi} - \frac{t}{2}$. This means Broker $\underline{\alpha}$ will propose a pooling offer to Firm A if and only if

$$\rho^P \geq \overline{\pi} - \frac{t}{2}. \tag{4}$$

Secondly, the most efficient Broker $\overline{\alpha}$ can either sell to Firm A at the pooling price ρ^P and then to Firm B at price $\hat{\pi} - \underline{\pi}$, or it can sell to only Firm B at a price $\overline{\pi} - \frac{t}{2}$. Therefore, Broker $\overline{\alpha}$ will sell to Firm A at the pooling price if and only if

$$\rho^P - c\delta + \hat{\pi} - \underline{\pi} - c\delta(1-\overline{\alpha}\delta) \geq \overline{\pi} - \frac{t}{2} - c\delta \iff \rho^P \geq \overline{\pi} - \frac{t}{2} - (\hat{\pi} - \underline{\pi} - c\delta(1-\overline{\alpha}\delta))\,. \quad (5)$$

Since $\delta \geq \delta_{AB}(\overline{\alpha})$, profits $\hat{\pi} - \underline{\pi} - c\delta(1-\overline{\alpha}\delta)$ are positive, and conditions (3), (4), and (5) are compatible if and only if

$$\lambda[\overline{\pi} - \underline{\pi}] + (1-\lambda)[\hat{\pi} - \underline{\pi}] \geq \overline{\pi} - \frac{t}{2} \iff \lambda \geq \underline{\lambda} = 1 - \frac{\frac{t}{2} - \underline{\pi}}{\overline{\pi} - \hat{\pi}} \quad (6)$$

- For $\lambda \geq \underline{\lambda}$, the two types of brokers will propose the same price ρ^P to Firm A and only the $\overline{\alpha}$-type broker will treat Firm B's data (at a price $\hat{\pi} - \underline{\pi}$).
- For $\lambda < \underline{\lambda}$, Broker $\underline{\alpha}$ will sell to only Firm B, at a price $\overline{\pi} - \frac{t}{2}$. Because of this choice, Broker $\overline{\alpha}$ cannot propose a price that will be accepted by Firm A, and its strategy is the same as the other broker.

Consider now the case where $\delta_{AB}(\underline{\alpha}) < \delta_2(\overline{\alpha})$, and $\delta \in [\delta_{AB}(\underline{\alpha}), \delta_2(\overline{\alpha})]$. When the brokers' types are public, Broker $\overline{\alpha}$ will treat Firm A and Firm B's data at price $\hat{\pi} - \underline{\pi}$, whereas Broker $\underline{\alpha}$ will treat only Firm B's data a price $\overline{\pi} - \frac{t}{2}$.

When the brokers' types are no longer public, the same reasoning as above applies. The only difference is that condition (5) now writes as

$$\rho^P - c\delta + (\hat{\pi} - \underline{\pi} - c\delta(1-\overline{\alpha}\delta)) \geq 2(\hat{\pi} - \underline{\pi}) - c\delta - c\delta(1-\overline{\alpha}\delta)) \iff \rho^P \geq \hat{\pi} - \underline{\pi} \quad (7)$$

Since $\overline{\pi} - \frac{t}{2} \geq \hat{\pi} - \underline{\pi}$, conditions (3), (4) and (7) are compatible again if condition (6) is satisfied.

Therefore, if $\lambda \geq \underline{\lambda}$, both brokers will propose a pooling price ρ^P to Firm A and Broker $\overline{\alpha}$ will also sell information to Firm B. However, if $\lambda < \underline{\lambda}$, Broker $\underline{\alpha}$ will offer $\overline{\pi} - \frac{t}{2}$ to Firm B whereas Broker $\overline{\alpha}$ will offer a low price $\hat{\pi} - \underline{\pi}$, accepted by both firms. In this last case, the equilibrium is separating, allowing Firm A to infer the type of the broker from the price it charges for information.

Thirdly, consider $\delta \in [\delta_2(\underline{\alpha}), \min\{\delta_{AB}(\underline{\alpha}), \delta_2(\overline{\alpha})\}]$. There, Firm A knows that both brokers will sell information to Firm B if Firm A has its data treated. Hence, Firm A is willing to buy information at a maximal price equal to $\hat{\pi} - \underline{\pi}$. In this area, this strategy is profitable only for Broker $\overline{\alpha}$. Thus, this broker will make an offer to Firm A at price $\hat{\pi} - \underline{\pi}$ while Broker $\underline{\alpha}$ will propose the higher price $\overline{\pi} - \frac{t}{2}$ that Firm A will decline. In this case, the equilibrium is separating.

We can summarize these different configurations:

- When $\delta \in [\delta_2(\underline{\alpha}), \min\{\delta_{AB}(\underline{\alpha}), \delta_2(\overline{\alpha})\}]$, Broker $\underline{\alpha}$ offers the high price $\overline{\pi} - \frac{t}{2}$, which is only accepted by Firm B, and Broker $\overline{\alpha}$ charges price $\hat{\pi} - \underline{\pi}$ accepted by both firms.

- When $\delta_{AB}(\underline{\alpha}) < \delta_2(\overline{\alpha})$: For $\delta \in [\delta_{AB}(\underline{\alpha}), \delta_2(\overline{\alpha})]$, both brokers charge a pooling price ρ^P if λ is high. If λ is low, the equilibrium is separating with Broker $\overline{\alpha}$ selling at price $\hat{\pi} - \underline{\pi}$ to each firm, while Broker $\underline{\alpha}$ offers a high price $\overline{\pi} - \frac{t}{2}$ for both firms, which is only accepted by Firm B.

- When $\delta \in [\max\{\delta_{AB}(\underline{\alpha}), \delta_2(\overline{\alpha})\}, \delta_{AB}(\overline{\alpha})]$, both brokers charge price ρ^P if λ is high or a high price for both if λ is low.

In any other cases, the selling strategy of a broker does not depend on its type, and uncertainty over the broker's type does not change its pricing decision.

There are three main conclusions from this analysis. First, when the equilibrium is pooling, Broker $\overline{\alpha}$ sells to both firms while Broker $\underline{\alpha}$ sells to only Firm A. This implies that Broker $\overline{\alpha}$ makes higher profits than Broker $\underline{\alpha}$ due to uncertainty, even though the opposite is true when firms have perfect information on the type of the broker.

Secondly, the fact that Broker $\overline{\alpha}$ makes higher profits than Broker $\underline{\alpha}$ also differs from standard models of contracting with incomplete information where both types typically earn the same profits in a pooling equilibrium. This difference stems from the sequentiality of the game we consider, since Broker $\overline{\alpha}$ can make additional profits by selling to Firm B in the second stage.

Thirdly, when its type is private information a broker would like Firm A to believe that its learning effect is weak, which would prevent it from selling to Firm B at the following stage. Hence, uncertainty decreases the profits of Broker $\underline{\alpha}$ compared to the full information case, while Broker $\overline{\alpha}$ can benefit from uncertainty.

Overall, even when firms may not perfectly know to what extent the broker can learn from their data, the possibility of a strong learning effect will significantly shape their willingness to pay for information and the ability of the broker to sell to the firms. Because of uncertainty, Firm A will be cautious and may consider that a (weak) broker can have a strong learning effect. This suggests that the results we derive under perfect information hold when firms have incomplete information regarding the learning effect of the broker.

7.4 Strategic Choice of Information Sales by the Broker

Our baseline analysis focuses on firms willing to obtain information on exogenous shares of consumer demand. This type of interaction routinely occurs in the data industry where firms request data processing from a broker by supplying a specific dataset, without granting the broker control over which part of the demand is covered by this data.

Another way to model the relationship between a broker and the firms is to assume that firms have information on a large share of the demand (say, each firm has information on the closest half of the demand). The broker can then strategically select the portion of consumers $\delta_A \in [0, \frac{1}{2}]$ and $\delta_B \in [\frac{1}{2}, 1]$ on whom it sells information to each firm. Recent literature addresses similar issues by examining the optimal values of δ_A and δ_B when information is sold to both firms simultaneously (Bounie et al., 2021; Abrardi et al., 2024a). In practice, this type of interaction also represents the functioning of the US credit bureaus for instance.

We discuss the implications of this strategic choice in this section but, to keep our discussion within reasonable length, we do not include the details of the formal analysis, and focus here on the main intuitions.

To fix ideas, let $\hat{\delta}_A(\alpha)$ and $\hat{\delta}_B(\alpha)$ represent the optimal values of information chosen by the broker to maximize its profits when selling to the firms with a learning effect of strength α. Intuitively, as the ability of the broker to learn from data increases, we should expect two effects: first, that both $\hat{\delta}_A(\alpha)$ and $\hat{\delta}_B(\alpha)$ increase, and secondly, that the broker makes higher profits.

On the one hand, the value of $\hat{\delta}_A(\alpha)$ increases since it allows the broker to lower its marginal cost in the following stage. On the other hand, the broker can also sell more information $\hat{\delta}_B(\alpha)$ to Firm B since its treatment cost is lower with a greater α.

Contrary to this intuition, we can show that $\hat{\delta}_A(\alpha)$ *decreases* with α. The rationale behind this result is similar to the main mechanism highlighted in our baseline model. As α increases, Firm A anticipates that providing its data to the broker will increase its ability to lower its treatment cost, allowing it to sell more information to Firm B. In turn, this intensifies the competitive pressure on Firm A which loses profits when α increases. Hence, a higher α reduces the marginal value of information for Firm A. This gives the broker incentives to sell *less information* to Firm A when its learning effect increases. This reduction of $\hat{\delta}_A(\alpha)$ is so strong the broker can sell less information to firms overall when α increases, i.e. $\hat{\delta}_A(\alpha) + \hat{\delta}_B(\alpha)$ can decrease with α. In turn, an increase in the strength of the learning effect α can decrease the profits of the broker.

Note that when the broker can strategically choose the value of $\hat{\delta}_B$ it is always able to sell information to Firm B. This implies that exclusivity is never achievable, even though it can be profit maximizing for the broker. In other words, its ability to choose δ_A and δ_B can make the broker lose profits compared to our baseline scenario. Conversely, because the broker cannot commit to sell few information to Firm B, its lack of commitment ability harms Firm A, and lowers its profits, down to a point where selling to Firm A can be unprofitable for the broker. This effect is similar to the main model, and we can show that the core mechanism is robust to a strategic choice of the amounts of data to treat. Here again, this loss results from the inability of the broker to commit to exclusivity once it can choose δ_B freely. This implies that a broker would in fact prefer *not to be able* to choose the amounts of data to treat, even if the technology to do so was available for free.

8 Implications and Discussion

8.1 Strategic Implications

Our results have implications for managers of data broker companies and firms competing in the product market. First, a broker enjoying a strong learning effect can lose profits and competitiveness due to this ability, and even reach a point where it is unable to sell information to firms.

Hence, investments in enhanced learning capacities, for instance through R&D programs and the development of sophisticated algorithmic techniques can make a broker *lose* profits. This sends a cautionary note to brokers striving to improve their data analytics capabilities, as being less proficient at learning from data may, paradoxically, represent a significant competitive advantage. By not overly optimizing their learning capabilities, brokers may maintain a favorable market position and avoid triggering adverse competitive dynamics that could reduce their profitability.

Managerial Insight 1. *Brokers can lose profits if they can better learn from data.*

Moreover, we have shown that firms can be reluctant to deal with brokers when they are uncertain about their types. Hence, brokers can make higher profits if they can convince a firm of their limited ability to learn from data. This can be achieved for instance through transparency regarding the analytical techniques brokers employ for data processing. In an extreme scenario where Firm A is unaware of the learning effect, the broker can more frequently sell data, leveraging the learning effect to enhance its market position and profitability.

Managerial Insight 2. *Data brokers can achieve higher profits if they can convince firms that they have a limited ability to learn from data, or if they can commit not to use these data to improve their services in the future.*

Conversely, firms make higher profits if they anticipate that providing their data to the broker will reduce its treatment cost and allow it to treat the data of their competitor. In this case, a firm can be better off by not acquiring information, even if this means that its competitor will buy information afterward. Overall, managers of companies that collect valuable data should anticipate how these data can be leveraged by data brokers. This awareness could significantly change the bargaining power dynamics between these firms and the brokers.

Managerial Insight 3. *When brokers benefit from a learning effect, a firm can achieve higher profits by not buying information.*

Finally, we have shown that a broker may achieve higher profits by treating the data of Firm A at a loss, in particular when competing fiercely with another broker. Doing so has the flavor of an investment in innovation reducing its treatment cost when dealing with Firm B.

Managerial Insight 4. *Treating the data of a firm can have the value of an investment in innovation, reducing the brokers' future treatment costs.*

8.2 Policy Implications

Turning to the implications of our results for policymakers, we first discuss how learning effects can impact consumer surplus by changing the ability of a broker to sell information to the firms, and in turn their pricing decisions. We then consider recent policies on data sharing and data property rights.

8.2.1 Consumer Surplus

A rapid look at Equation (2) (in Section 4) shows that consumer surplus increases when a firm has access to information, and is maximized when both firms have this access. More information allows firms to charge targeted prices to consumers and better extract their surplus, shifting downward aggregate surplus. But information also sharply intensifies competition between firms. The net effect on consumers is positive, therefore consumer surplus is the highest when both firms are informed and the lowest when they are not.

In the baseline model, the ability of the broker to learn from data does not change aggregate consumer surplus when it induces a change in the identity of the firm buying information. Indeed, in this case, as firms can buy information on the same number of consumers ($\delta_A = \delta_B$), it does not matter whether Firm A or Firm B buys information, the resulting surplus remains the same. When the broker sells information to both firms instead of only Firm B due to the learning effect, aggregate consumer surplus is increased compared to a situation where the learning effect is weaker and only Firm B buys information.

When renegotiation is introduced, we have shown in Proposition 6 that the learning effect can yield an equilibrium where the broker does not sell information to firms. In this case, consumer surplus is reduced because of the learning effect, as without learning, the broker would sell information to Firm B, which would benefit consumers.

Finally, when Firm A is unaware of the learning effect, we have shown in Proposition 7 that the broker can sometimes sell information to more firms due to the learning effect. This, in turn, increases consumer surplus.

This ambiguous impact of learning on consumer surplus is important for regulators enforcing policies aiming at encouraging data-driven innovation. For instance, subsidies fostering public research in data mining and the processing of large data sets using AI can enhance the ability of brokers to learn from data. This will critically shape the provision of information in the market, with ambiguous impacts for consumers.

Policy Insight 1. *Policies promoting investments in innovation allowing brokers to better learn from data will benefit consumers by expanding the data market, but can lower their surplus by discouraging firms from buying information.*

8.2.2 Data Sharing

There is also a growing debate among academics and policymakers on the regulation of information-sharing practices among firms and how they can impact market competition (Jullien and Sand-Zantman, 2021; Bhargava et al., 2024).

By considering the possibility for brokers to share data after the first stage, we provide useful elements to these debates. Indeed, we show that when brokers compete head to head, they compete more fiercely when they can share information. This benefits the firms, but does not impact consumer surplus.

When the least efficient broker achieves monopoly profits by selling to Firm A, its ability to share data with a more efficient broker may undermine its dominant position and restore fierce competition between brokers, resulting in equilibrium where only Firm B acquires data. In turn, this reduces sharply the brokers' industry profits. Overall, the ability for brokers to share their data intensifies upstream competition and lowers their profits.

Policy Insight 2. *Allowing brokers to share their data reduces their market power with respect to the firms competing in the product market.*

These competitive effects of information sharing can be offset if a regulator gives Firm A property rights over its data. In this case, data property rights restore the ability of the broker to credibly commit not to share Firm A's data with other brokers, which also benefits Firm A.

Policy Insight 3. *Data property rights giving Firm A control over the reuse of its data by a broker can benefit both Firm A and data brokers.*

9 Conclusions

Learning-by-doing resulting from data usage, AI training and the development of algorithms can reduce the competitiveness of data brokers and their ability to sell data, reducing in turn consumer surplus. The implications that we have derived for managers and policymakers are central in today's digital economy, where data brokers have become key players by supplying information to any type of business.

These results also have important implications for the literature that usually assumes increasing returns to scale in data. We have shown that, anticipating these potential returns to scale, firms may prefer to hoard their data rather than provide them to a broker. This suggests that, although the effects of learning-by-doing may be theoretically viable, their actual realization is contingent on specific market characteristics that are required for firms to be willing to provide their data to a broker. While we cast our analysis on data markets, the core effects that we identify are broadly applicable to any situation involving learning-by-doing with a downstream competitive market.

We hope this work can be used and built on to further explore how the effects of learning-by-doing impact market outcomes. Further research could analyze the decision of brokers to invest in innovation allowing them to better learn from data – in our model, to choose the value of α. Our results suggest that brokers might use investments to differentiate, such that one broker or the other may behave as a monopolist for a share of the market, to the detriment of downstream firms and of consumers. Also for further research, effects of learning by doing can influence the decision of a broker to integrate vertically with one of the firms, and to further sell to the remaining firm after integration (see de Cornière and Taylor (2024b) for an analysis of such data-driven mergers when there is no learning effect).

References

Abrardi, Laura, Carlo Cambini, and Flavio Pino, "Data Brokers Competition, Synergic Datasets, and Endogenous Information Value," 2024b.

_ , _ , Raffaele Congiu, and Flavio Pino, "User data and endogenous entry in online markets," *The Journal of Industrial Economics*, 2024a.

Arrow, Kenneth J, "The economic implications of learning by doing," *The Review of Economic Studies*, 1962, *29* (3), 155–173.

Bajari, Patrick, Victor Chernozhukov, Ali Hortaçsu, and Junichi Suzuki, "The impact of big data on firm performance: An empirical investigation," in "AEA papers and proceedings," Vol. 109 American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203 2019, pp. 33–37.

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

Besanko, David, Ulrich Doraszelski, and Yaroslav Kryukov, "The economics of predation: What drives pricing when there is learning-by-doing?," *American Economic Review*, 2014, *104* (3), 868–897.

_ , _ , and _ , "How efficient is dynamic competition? the case of price as investment," *American Economic Review*, 2019, *109* (9), 3339–3364.

_ , _ , _ , and Mark Satterthwaite, "Learning-by-doing, organizational forgetting, and industry dynamics," *Econometrica*, 2010, *78* (2), 453–508.

Bhargava, Hemant, Antoine Dubus, David Ronayne, and Shiva Shekhar, "The Strategic Value of Data Sharing in Interdependent Markets," Technical Report, CRC TRR 190 Rationality and Competition 2024.

Bounie, David, Antoine Dubus, and Patrick Waelbroeck, "Selling strategic information in digital competitive markets," *The RAND Journal of Economics*, 2021, *52* (2), 283–313.

Cabral, Luis MB and Michael H Riordan, "The learning curve, market dominance, and predatory pricing," *Econometrica: Journal of the Econometric Society*, 1994, pp. 1115–1140.

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

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

Darr, Eric D, Linda Argote, and Dennis Epple, "The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises," *Management science*, 1995, *41* (11), 1750–1762.

de Cornière, Alexandre and Greg Taylor, "Data and competition: A simple framework," *RAND Journal of Economics, forthcoming*, 2024a.

_ **and** _ , "Data-Driven Mergers," *Management Science*, 2024b.

Deng, Shanglyu, Dun Jia, Mario Leccese, and Andrew Sweeting, "Bargaining and Dynamic Competition," 2023.

Dosis, Anastasios and Wilfried Sand-Zantman, "The ownership of data," *The Journal of Law, Economics, and Organization*, 2023, *39* (3), 615–641.

FTC, "Data brokers: A call for transparency and accountability," *Ramirez, Edith and Brill, Julie and Ohlhausen, Maureen K and Wright, Joshua D and McSweeny, Terrell*, 2014, (s 47).

Gabel, Sebastian and Daniel Guhl, "Comparing the effectiveness of rewards and individually targeted coupons in loyalty programs," *Journal of Retailing*, 2022, *98* (3), 395–411.

Hatch, Nile W and David C Mowery, "Process innovation and learning by doing in semiconductor manufacturing," *Management Science*, 1998, *44* (11-part-1), 1461–1477.

Hippel, Eric Von and Marcie J Tyre, "How learning by doing is done: problem identification in novel process equipment," *Research policy*, 1995, *24* (1), 1–12.

Hunold, Matthias and Shiva Shekhar, "Supply chain innovations and partial ownership," *Review of Industrial Organization*, 2022, *60* (1), 109–145.

Irwin, Douglas A and Peter J Klenow, "Learning-by-doing spillovers in the semiconductor industry," *Journal of political Economy*, 1994, *102* (6), 1200–1227.

Jones, Charles I and Christopher Tonetti, "Nonrivalry and the Economics of Data," *American Economic Review*, 2020, *110* (9), 2819–2858.

Jullien, Bruno and Wilfried Sand-Zantman, "The economics of platforms: A theory guide for competition policy," *Information Economics and Policy*, 2021, *54*, 100880.

Koski, Heli, "How Do Competition Policy and Data Brokers Shape Product Market Competition?," Technical Report, ETLA Working Papers 2018.

Lee, Dong-Joo, Jae-Hyeon Ahn, and Youngsok Bang, "Managing consumer privacy concerns in personalization: a strategic analysis of privacy protection," *Mis Quarterly*, 2011, pp. 423–444.

Levitt, Steven D, John A List, and Chad Syverson, "Toward an understanding of learning by doing: Evidence from an automobile assembly plant," *Journal of political Economy*, 2013, *121* (4), 643–681.

Liu, Ernest, Song Ma, and Laura Veldkamp, "Data sales and data dilution," 2023.

Miller, Danny and Ming-Jer Chen, "Sources and consequences of competitive inertia: A study of the US airline industry," *Administrative science quarterly*, 1994, pp. 1–23.

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

Rosenberg, Nathan, *Inside the black box: technology and economics*, cambridge university press, 1982.

Sweeting, Andrew, Dun Jia, Shen Hui, and Xinlu Yao, "Dynamic Price Competition, Learning-by-Doing, and Strategic Buyers," *American Economic Review*, 2022, *112* (4), 1311–1333.

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

Zhang, Xin, Lihong Cheng, Yugang Yu, and Yong Tan, "To Partner or Not to Partner? The Partnership Between Platforms and Data Brokers in Two-Sided Markets," *Information Systems Research*, 2024.

A Appendix

A.1 Proof of Lemma 3

To understand how an increase in α impacts the selling strategy of the broker and its profits, we distinguish two cases.

Case 1: $\delta_0 \leq \frac{2}{5}$

Without learning ($\alpha = 0$) and for a value δ_0 between $\delta_{AB}(0)$ and δ_1, the broker sells information to Firm A and makes profits $\overline{\pi} - \underline{\pi} - c\delta_0$. As the learning effect becomes stronger, the threshold $\delta_{AB}(\alpha)$ below which the broker sells information to Firm B instead of Firm A increases, and there exists a value $\hat{\alpha} = \frac{11}{9\tilde{c}} - \frac{1}{\tilde{c}\delta_0}(\frac{2}{3} - \tilde{c})$ for which $\delta_{AB}(\alpha)$ will be greater than δ_0 only if α is above $\hat{\alpha}$. Hence, if the learning effect becomes stronger, the broker will sell information more often to Firm B and less often to Firm A.

When the learning effect increases furthermore, the broker will sell information to both firms instead of only to Firm B, in which case its profits increase with α. Indeed, there is a value denoted $\hat{\hat{\alpha}} = \frac{5}{3\tilde{c}} - \frac{1}{\tilde{c}\delta_0}(\frac{2}{3} - \tilde{c})$[17] for which $\delta_2(\alpha)$ will be greater than δ_0 only if α is above $\hat{\hat{\alpha}}$. Thus, the broker sells information to only Firm B if the learning effect is below $\hat{\hat{\alpha}}$ making profits equal to $\overline{\pi} - \underline{\pi} - c\delta_0$ whereas it sells information to both firms when the learning effect is above $\hat{\hat{\alpha}}$, making then a profit equal to $2(\hat{\pi} - \underline{\pi}) - 2c\delta_0 + \alpha c\delta_0^2$.

Case 2: $\delta_0 > \frac{2}{5}$

The threshold $\delta_2(\alpha)$ below which the broker sells information to both firms instead of only Firm B increases when α is below $\frac{5}{2}$ and is then equal to $\frac{2}{5}$ for $\alpha \geq \frac{5}{2}$. This implies that for any value $\delta_0 > \frac{2}{5}$, even when the learning effect is very strong, the broker will not sell information to both firms since $\delta_0 > \delta_2(\alpha)$ for any value of α. In turn, the resulting profits of the broker as a function of α are depicted in Figure 4 when $\delta_0 > \delta_2(\alpha)$.

[17] Note that $\hat{\hat{\alpha}} > \hat{\alpha}$.

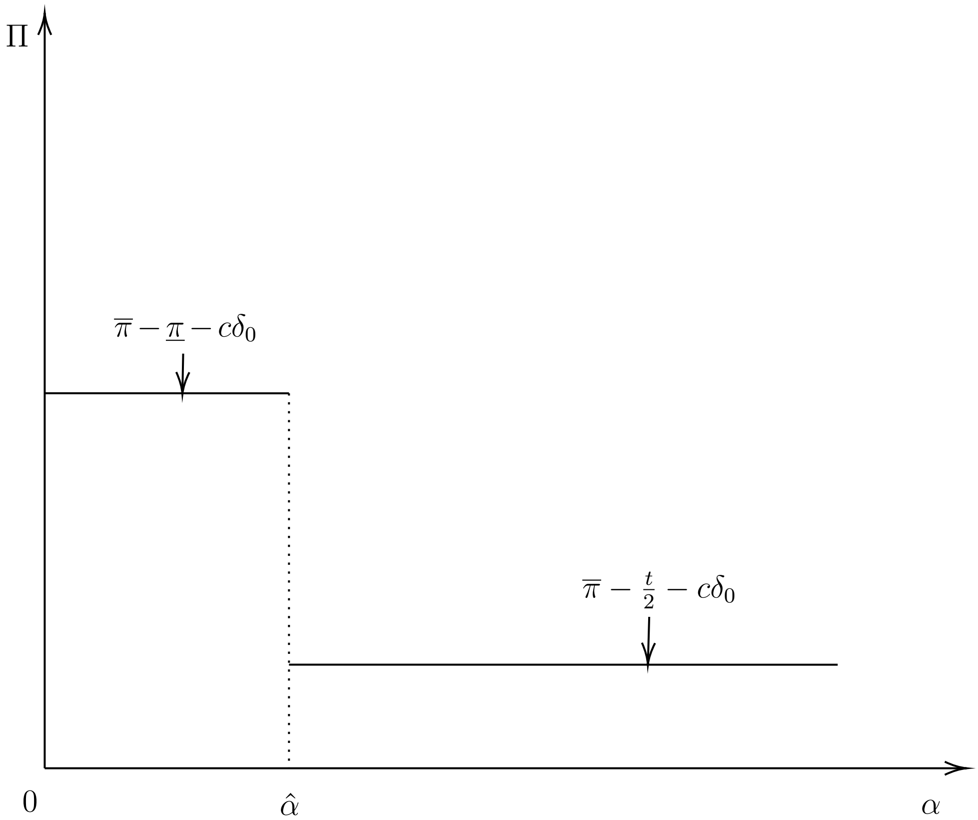


Figure 4: Profits of the broker when $\delta \geq \frac{2}{5}$

A.2 Proof of Proposition 4

We consider the two remaining configurations, and we show that there, competition between brokers reduces their ability to charge the highest price for information, but does not change the identity of the firm that buys information.

First, consider $\delta_{AB}(\underline{\alpha}) < \delta_c(\overline{\alpha})$ and $\delta \in [\delta_{AB}(\underline{\alpha}), \delta_c(\overline{\alpha})]$. In this case, Broker $\overline{\alpha}$ is willing to sell information to both firms, while Broker $\underline{\alpha}$ can sell information to only Firm A.

Hence, Firm A can buy information from Broker $\overline{\alpha}$ at price $\rho_A(\overline{\alpha})$, and, anticipating that Firm B will buy information later, Firm A would make profits equal to

$$\hat{\pi} - \underline{\pi} - \rho_A(\overline{\alpha})$$

Or Firm A can buy information from Broker $\underline{\alpha}$ at a price $\rho_A(\underline{\alpha})$. Because Broker $\underline{\alpha}$ has a low learning effect, a broker can sell to Firm B only if Firm A declines the offer of Broker $\underline{\alpha}$. Hence, buying from Broker $\underline{\alpha}$, Firm A would make profits equal to

$$\overline{\pi} - \underline{\pi} - \rho_A(\underline{\alpha})$$

In the second stage, it is not profitable for Broker $\underline{\alpha}$ to sell information to Firm B, and Broker $\overline{\alpha}$ sells information to Firm B only if it has also sold to Firm A, in which case it can charge a price equal to the willingness to pay of Firm B

$$\rho_B(\overline{\alpha}) = \hat{\pi} - \underline{\pi}$$

Thus, the profits of Broker $\overline{\alpha}$ selling to Firm A and then to Firm B are equal to

$$\Pi_{\overline{\alpha}} = \rho_A(\overline{\alpha}) + \hat{\pi} - \underline{\pi} - c\delta - c\delta(1 - \overline{\alpha}\delta)$$

The profits of Broker $\underline{\alpha}$ selling only to Firm A are equal to

$$\Pi_{\underline{\alpha}} = \rho_A(\underline{\alpha}) - c\delta$$

We can now compare the willingness to pay of Firm A for each of the lowest bound of prices, $c\delta$ for Broker $\underline{\alpha}$ and $-\hat{\pi} + \underline{\pi} + c\delta + c\delta(1 - \overline{\alpha}\delta)$ for Broker $\overline{\alpha}$.

$$\hat{\pi} - \underline{\pi} - \rho_A(\overline{\alpha}) \leq 2[\hat{\pi} - \underline{\pi}] - c\delta - c\delta(1 - \overline{\alpha}\delta)$$

if Firm A buys from Broker $\overline{\alpha}$ at its lowest possible price.

$$\overline{\pi} - \underline{\pi} - \rho_A(\underline{\alpha}) \leq \overline{\pi} - \underline{\pi} - c\delta$$

if Firm A buys from Broker $\underline{\alpha}$ at its lowest possible price. Comparing both maximal willingness to pay, we find that:

$$\overline{\pi} > 2\hat{\pi} - \underline{\pi} - c\delta(1 - \overline{\alpha}\delta)$$

This inequality always holds and Firm A has a higher willingness to pay for the information of Broker $\underline{\alpha}$ than for that of Broker $\overline{\alpha}$, even accounting for a potentially higher price.

Hence, Broker $\overline{\alpha}$ will exert a threat on Broker $\underline{\alpha}$ by charging the lowest possible price for its information, implying that Firm A can make profits equal to $2[\hat{\pi} - \underline{\pi}] - c\delta - c\delta(1 - \overline{\alpha}\delta)$. In turn, Broker $\underline{\alpha}$ can charge a price to Firm A that guarantees this competitive level of profits, which verifies:

$$\overline{\pi} - \underline{\pi} - \rho_A(\underline{\alpha}) = 2[\hat{\pi} - \underline{\pi}] - c\delta - c\delta(1 - \overline{\alpha}\delta)$$

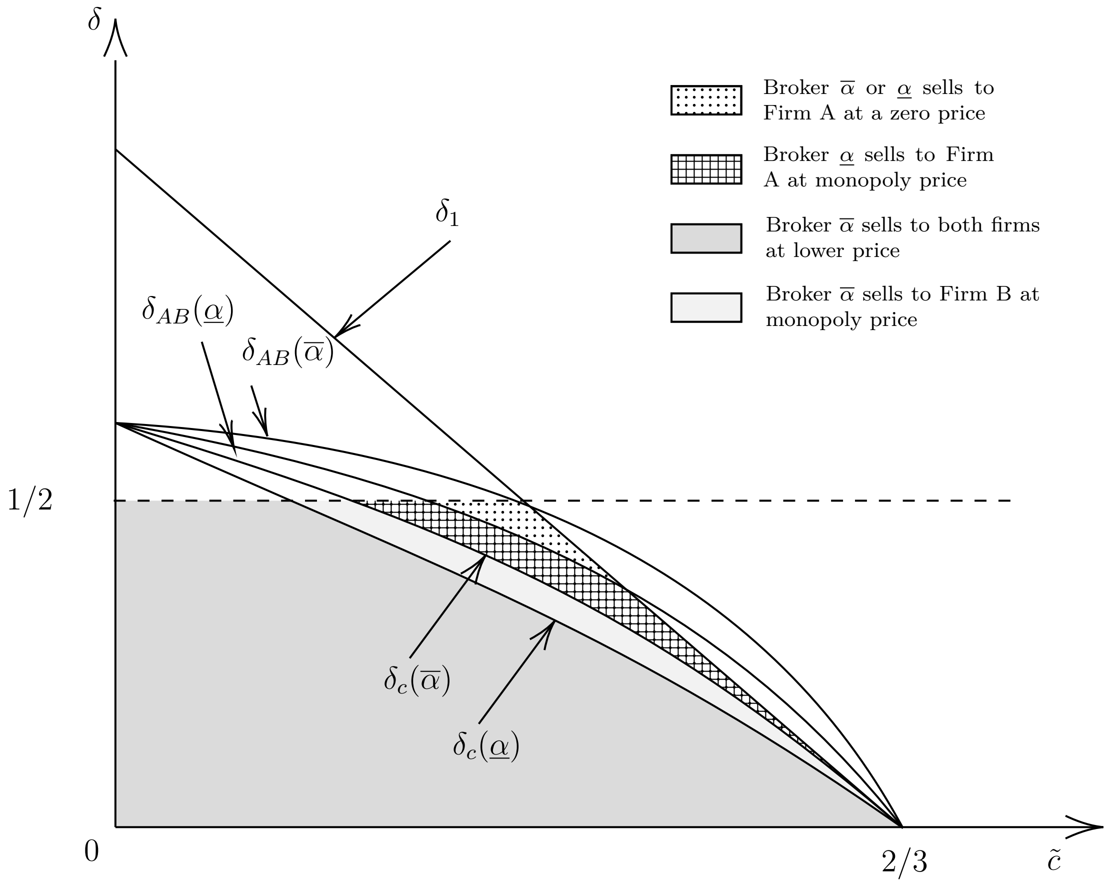
$$\rho_A(\underline{\alpha}) = \overline{\pi} - 2\hat{\pi} + \underline{\pi} + c\delta + c\delta(1 - \overline{\alpha}\delta)$$

In turn, Broker $\underline{\alpha}$ makes profits equal to

$$\Pi_{\underline{\alpha}} = \overline{\pi} - 2\hat{\pi} + \underline{\pi} + c\delta(1 - \overline{\alpha}\delta)$$

Secondly, consider the situation where both brokers can sell information only to the same firm: either to Firm A when $\delta \in [\delta_{AB}(\overline{\alpha}), \delta_1]$, or only to Firm B when $\delta \in [\delta_c(\overline{\alpha}), \delta_{AB}(\underline{\alpha})]$. As brokers sell information only at one stage, they do not benefit from the learning effect and incur a treatment cost equal to $c\delta$. Hence, brokers are perfectly identical in this case, and they compete à la Bertrand by lowering their price to the treatment cost. In turn, they make zero profits.

The different thresholds are represented in the following figure.



δ
Broker $\overline{\alpha}$ or $\underline{\alpha}$ sells to Firm A at a zero price
Broker $\underline{\alpha}$ sells to Firm A at monopoly price
Broker $\overline{\alpha}$ sells to both firms at lower price
Broker $\overline{\alpha}$ sells to Firm B at monopoly price
δ_1
$\delta_{AB}(\underline{\alpha})$
$\delta_{AB}(\overline{\alpha})$
1/2
$\delta_c(\overline{\alpha})$
$\delta_c(\underline{\alpha})$
0
2/3
$\tilde{c}$