Data and competitive markets: some notes on competition, concentration and welfare

António Osório

Universitat Rovira i Virgili (Dept. Economics) and ECO-SOS
(antonio.osoriodacosta@urv.cat)

Abstract

Companies are increasingly using data to predict behavior and improve the relation with their customers. In this context, data exchange raises important concerns regarding competition, concentration and welfare. This paper presents a novel linear demand approach that captures data and information effects in competitive markets, which are conveniently summarized in a precision parameter. Subsequently, the proposed approach is applied to study the firm's incentives to exchange data and their impact in fundamental market variables, welfare and market concentration measures. We found that the incentives for data exchange between competitor firms emerge when the individual information gains are strong enough to compensate for the competitor's information gains, and the associated strategic correlation effect between varieties. The results also suggest that market concentration tends to increase after data exchange, but both consumers and producers benefit from it. The reason is that better data allows firms to positioning closer to consumers' needs.

Keywords: Data; Data Exchange; Linear Demand; Consumer Targeting; Competitive Markets.
JEL classification: D11, L11, L40.

---

1. Introduction

Companies are increasingly adopting Artificial Intelligence (AI) and big data in the relation with their customers to predict behaviour, and automate and improve decision-making. Companies are collecting and processing large amounts of data, which translates into products with designs and characteristics that better fit consumers' needs, but also in more accurate consumers targeting and strategic decisions

(Fudenberg and Villas-Boas, 2012; Shen and Villas-Boas, 2018). In this process, some companies are becoming increasingly dominant, transforming businesses and the way markets structure themselves. However, researchers, policy-makers and competition authorities are far from understanding the competitive implications of data technologies. There is a sense that data may improve the firms market position, increase market concentration and affect consumer welfare, but even that is not clear (Belleflamme et al., 2020; Taylor and Wagman, 2014). There are also concerns on how data is used, in particular, the possibility that firms may exchange or trade data about consumers (Acquisti et al., 2016; Bergemann et al., 2020; Bonatti and Cisternas, 2020; Liu and Serfes, 2006; Shy and Stenbacka, 2016). In this context, price discrimination and data trading are also rising important concerns (Clavorà Braulin and Valletti, 2016; Montes et al., 2019; Shiller, 2014).

In general terms, it is difficult to study the effects of data on competition, market structure and concentration, mergers and acquisitions, antitrust policy or even welfare, because data is intangible, and appears in many different forms and uses (Acquisti et al., 2016; De Corniere and Taylor, 2019). However, general knowledge about the effects of data in markets and the economy is crucial for policy and decision makers to implement policies that enhance markets and protect consumers.

In this paper, we present an approach that can help academics and researchers to better understand the effects of data in competitive markets. In this context, the objective and contribution of this paper are twofold. First, to present two linear demand models that capture data and information processing effects in competitive markets. Second, to study the implications of data exchange between firms on the main market variables (prices, quantities, profits), and welfare and concentration measures.

In this paper, firms exchange information in a bilateral way. In other words, firms have information to share/exchange and the question is whether this exchange is beneficial for the involved firms, i.e., whether their profits after the information exchange are higher or lower.

The first challenge is to model how firms process information about consumers. In this context, we follow the Gaussian signals literature (Chamley, 2004, Vives, 1999), in which the data dimension is summarized by a single precision parameter, the inverse of the variance. Consequently, every time firms improve their data sets, data precision improves linearly, which is a convenient property.

The second challenge is to integrate the associated data effects into the representative consumer utility, and consequently into the demand function, in a meaningful way. In this context, we want an intuitive and rich framework that can be easily applied by researchers and practitioners, that renders tractable expressions, that extends the existent linear demand models, and that can be applied into a large variety of problems, which become increasingly complex as more features are added. In this context, this paper proposes two approaches to model data in competitive markets: the *additive* and the *multiplicative* models.

In the additive model, the data component enters into the demand function in a simple and additive way. Consequently, better data increases sales, prices and profits. This approach offers a benchmark.

In the multiplicative model, in addition to the effect of data on the market variables, the model delivers a strategic correlation effect between the firms' targeting/positioning. For instance, when two firms exchange data they improve their data sets, but they also become more correlated in strategic terms, because they have the same data about the consumers' preferences (e.g., browsing history, location, gender, consumption habits, economic conditions, etc.). Consequently, their varieties, which represent their strategic targeting/positioning (e.g., products/services, strategies, objectives, etc.), become more correlated, which increases the competition between them.[1]

In terms of model selection, the multiplicative model has the advantage of generating richer trade-offs, but the additive model has the advantage of being more tractable.

In addition, this paper also studies the firms' incentives to exchange data. Data exchange raises important concerns regarding market competition, concentration and welfare. In this context, we must distinguish three different types of data exchange scenarios:

1. *Horizontal data exchange* when two firms competing in the same market exchange data between them.[2]
2. *Independent data exchange* when two firms competing in different and non-integrated markets exchange data between them.[3]

[1]The correlation effect in the multiplicative model can also be related with the idea of horizontal substitution between product varieties (e.g., Dixit, 1979; Singh et al., 1984; Shapley and Shubik, 1969).

[2]Examples of horizontal data exchange would be the data exchange between Lending Institutions or between Insurance companies about the rating and characteristics of their customers. Or hypothetically we could imagine for example a data exchange between Tesco and Carrefour or between Bank of America and Citigroup, which operate and compete in the same industry.

[3]Example of independent data exchange would be the data exchange between Facebook and

3. *Vertical data exchange* when two firms vertically integrated in customer/supplier relations exchange data.[4]

Even though the model in this paper is not suitable to capture data exchange effects in the vertical and independent cases, which are more related with production and cost efficiency, this distinction is important because data exchange has different implications in terms of strategic correlation and market competition in each of these three different scenarios. This paper focuses on horizontal data exchange, but some comments regarding the other types of data exchange are made because of their importance and novelty.

We found that the incentives for horizontal data exchange between competitor firms emerge under very mild conditions. Firms have incentives to exchange data providing that the information gains are strong enough, and the associated correlation effect is not too strong. Otherwise, the firm with better data has no incentives to exchange data. In both models, data exchange increases the sales and the profits of the involved firms. However, in the multiplicative model, prices may fall because of the increase in competition due to the increasing correlation.

We also found that market concentration increases after data exchange, which raises concerns regarding market competition, but both consumers and producers seem to benefit from it. These results are robust in both models, and are in some sense positively puzzling, as we would expect some negative effects for consumers. The increase in consumers' surplus is driven by the fact that producers' targeting/positioning become closer to the representative consumer preferences. The strength of these results depends on the magnitude of the information and correlation effects, as both effects seem to benefit consumers, either through better targeting/positioning or through more competition between firms.

These results are in line with the argument that supports the existence of benefits for consumers and the society from data sharing and exchange, and the creation of trustee platforms to centralize, manage and anonymize data.

We also found that the great losers from this process are the firms excluded from the data exchange. They do not benefit from any information improvement, and they

Data Brokers, which then sell this data to other companies operating in industries different from Facebook. Or hypothetically we could imagine for example a data exchange between Aldi and Banco Santander or between Inditex and Facebook (either through a data broker or not), which operate in completely unrelated industries.

[4]Examples could be the data exchange between auto manufacturers and BOSCH, or between Airline Companies and Airbus/Boeing, regarding the reliability and parameters of their components, which has been a common practice in those industries.

are affected by the increase in competition and the improved data of the competing firms involved in the data exchange.

We also note that in independent data exchange between firms operating in independent markets the effect on market power and concentration may be even stronger. In those cases, the data benefits are not diminished by strategic correlation effects because the involved firms operate in independent markets. Moreover, those firms' data sets are more likely to have great complementarity between them, which would strengthen even more the information effect.[5] In this context, independent data exchange may pose a higher threat in terms of market concentration and competition than horizontal data exchange. These observations may open a completely new perspective about the impact and the risks of data exchange in competitive markets.

Related literature

This paper is related with the theoretical research on information sharing in oligopoly that focus on the firms' incentives to share private information about a stochastic demand or costs parameter. Information sharing has an information improvement effect, which favors profits, and a correlation effect on the strategies that depends on the type of competition model and uncertainty structure (Raith, 1996; Vives, 1990; Vives, 1999). The correlation effect increases in common value (e.g., Clarke, 1983; Gal-Or, 1985; Li, 1985; Vives, 1984) and decreases in private value problems (e.g., Fried, 1984; Li, 1985; Shapiro, 1986). In the case of common values and strategic substitutes (e.g., Cournot competition with product substitution), information sharing decreases profits if product substitution is sufficiently high, because in this case the correlation of strategies has a strong effect, but also because information sharing improves the information precision of the competitors (Vives, 1999).

In this paper, firms observe the other firms' actions through prices. The difference between firms is in terms of targeting/positioning. Firms are uncertain about consumers' preferences, and for that reason unable to target consumers perfectly. Strategic correlation emerges when firms are targeting consumers using similar data sets. For that reason, in the multiplicative model, substitution effects depend on the covariance between the targeting/positioning strategies. In this sense, competition

[5] At this stage, it is important to stress the distinction between data sets complementarity, which is related with the idea of data overlapping and similarity between data sets, and correlation between positioning, which is related with the idea of firms being targeting the representative consumer preferences in the same way.

increases the more the firms' strategies become correlated, which depend on how similar their data sets are. A similar effect also appears in Belleflamme et al. (2020). In our setting, there is also an information improvement effect from data exchange that benefits profits, independently on whether firms' strategies become correlated. Moreover, consumers tend to benefit from targeting/positioning closer to their needs.

Bergemann et al. (2020) also models data in a linear demand framework with Gaussian signals and uncertainty about the demand intercept of each consumer. They study the data intermediary optimal strategy in terms of pricing and data granularity when consumers' preferences are correlated and there is a monopolist firm. They found that the intermediary optimal strategy is to supply aggregate data, but consumers lose because they are not compensated for aggregate externalities in their individual data. Our focus and approach are different. The firms' uncertainty is about the aggregate consumers' preferences in terms of targeting/positioning, and a specific loss function captures this intangible uncertainty.

This paper presents a novel approach to model data effects in linear demands models, but a large number of papers have been capturing data effects, price discrimination and data trading in Hotelling and vertical differentiation type models.

For instance, Montes et al. (2019) consider a Hotelling duopoly model. They found that the optimal strategy for a seller of data is exclusive selling, thereby maximizing the value of data by eliminating competition between firms competing with the same information. This outcome is bad for consumers. A similar exclusivity equilibrium arises under vertical product differentiation (Clavorà Braulin and Valletti, 2016). More recently, several studies have complemented the literature on information and competition. For instance, Gu et al. (2019) discusses conditions for data brokers to share their data, Bounie et al. (2021) shows that information sellers can balance the competitive and rent extraction effects of information, while Clavorà Braulin (2021) derives conditions for price discrimination to be pro or anti-competitive.

Liu and Serfes (2006) consider perfect price discrimination under horizontal and vertical differentiation to investigate the firms' incentives to share customer information and their welfare implications. Under horizontal differentiation, firms have no incentives to trade information. However, under vertical differentiation, the firm with the smaller loyal customer base sells its information to the firm with the larger loyal customer base. Information sharing benefits firms, but not consumers.

Fudenberg and Villas-Boas (2012) offer a series of different approaches to price discrimination based on purchase history. They found that in competitive settings information might actually lead to more intense competition because of less product differentiation.

Note two aspects relating those papers with the present paper. First, strategic alignment effects associated with the fact that firms are targeting the same consumers using the same data, which also emerges in the multiplicative model. Second, better data produces a positive effect, which improves the firms' general perception about the consumers' preferences, but does not allow discrimination among consumers. In this context, our approach seems more adequate to tackle general/aggregate market problems. Nonetheless, as pointed in Bergemann et al. (2020), individual and aggregate data are extremely correlated, as consumers might not be so much different.[6]

The rest of the paper is organized as follows: Section 2 describes the process of information processing and data aggregation, Section 3 models the consumers preferences, derives the market demand and the equilibrium in the additive and multiplicative models, Section 4 study data exchange between competing firms, Section 5 briefly extends the analysis to data exchange between independent firms and discusses vertical data exchange, and Section 6 concludes.

2. Data processing and aggregation

This section describes how firms process information about the representative consumer preferences. The objective is to offer a simple, but sufficiently rich framework. We borrow our approach from the Gaussian signals literature (Chamley, 2004).

2.1. ***The data processing model***

In our model, there is a representative consumer, whose preferences are not perfectly known by firms. Let $v \in \mathbb{R}$ be the abstract representation of these preferences.

[6]Advertising plays an important role in the targeting process. Shen and Villas-Boas (2018) considers the effect of advertising based on consumers past purchasing behavior and on firms pricing strategies. They found that if advertising is not annoying for consumers, firms end up charging higher prices and having higher demand because of the better match between their products and the consumers' preferences, which can be better for consumers and firms. Otherwise, if advertising is annoying for consumers, the consumers depend more on prices. Similarly, some consumers rely on third-party recommendations, in this context, De Corniere and Taylor (2019) study the incentives to provide biased information. In their setting, sellers are differentiated, but some consumers do not observe their characteristics and offers. However, an intermediary, which is integrated in one seller, is able to identify the best match for each consumer. They found that when consumers prioritize price, the bias favored firm offers lower utility, and bias harms consumers. However, when consumers prioritize quality, the bias favored firm offers a better deal than its rival does, and consumers are better-off than with no bias.

From the firm's perspective, $v \sim N(\overline{v}, \sigma_v^2)$ is normal distributed with mean $\overline{v}$ and variance σ_v^2, where $\rho_v = 1/\sigma_v^2$ is the inverse of the variance of v, and denotes the initial information precision about the unknown true value v, which we assume common to all firms.

However, in order to improve the knowledge about the value v, firms spend resources collecting, maintaining and analyzing data about the preferences of the representative consumer. At this stage, we ignore any potential costs and focus on information processing. We model the process of information acquisitions of firm j by an unbiased normal distributed private signal $s_j = v + \epsilon_j$, where ϵ_j is a noise term that is independent of v, i.e., $\epsilon_j \sim N(0, \sigma_{\epsilon_j}^2)$ and $s_j \sim N(v, \sigma_{\epsilon_j}^2)$, where $\rho_{\epsilon_j} = 1/\sigma_{\epsilon_j}^2$ is the inverse of the variance of ϵ_j and captures the precision of the signal of firm j about the unknown true value v. In our model, the signal represents the data of firm j, and the precision summarizes the information quality of the data.

Therefore, after obtaining the signal s_j, by Bayes' rule, the updated distribution of v is again normal distributed with mean:

$$\overline{v}_j' = \frac{\rho_{\epsilon_j}}{\rho_v + \rho_{\epsilon_j}} s_j + \frac{\rho_v}{\rho_v + \rho_{\epsilon_j}} \overline{v},$$

for $j = 1, ..., n$, where $\rho_j = \rho_v + \rho_{\epsilon_j}$ is the new information precision of firm j about the unknown true value v, and $\sigma_j^2 = 1/\rho_j$ is the new variance, i.e., the inverse of the precision, and measures the amount of uncertainty in the data of firm j about the value v.[7]

2.2. ***Data aggregation - no information overlapping and independent signals***

An interesting possibility of the present analysis is the study of the firms' gains and losses from data exchange and sharing. In this context, consider the case in which

[7]At this stage, we must note the advantages of the Gaussian signals framework in terms of modelling data, which are extremely convenient for research and applied work (Chamley, 2004; Vives, 1999). First, the simplicity of the Bayesian updating rule of the Gaussian model. Second, the fact that the information process can be summarized by a single parameter, i.e., the precision or its reciprocal (the variance), which simplifies the analysis enormously. Third, after the observation of a signal, the precision of the updated distribution is augmented linearly, which makes computations easier and intuitive. Fourth, the approach is fully consistent. For instance, when the variance of the signal tends to zero, the precision tends to infinity, the signal's weight tends to one, and the mean tends to the true value. Fifth, the increase in precision can be anticipated and a cost/benefit analysis on the decision to spend costly resources on acquiring and processing information can be done.

two firms join their data sets about the representative consumer preferences, i.e., they join their individual signals into a more informative joint signal. Therefore, after adding the signals s_j and s_k, both firms will have exactly the same data (assuming no strategic behavior like hiding, misrepresenting or lying), and the precision of the joint signal will be the same for both firms. In this context, firms' j and k updated distribution of v is again normal distributed with mean:

$$\overline{v}_{j+k} = \frac{\rho_{\epsilon_k}}{\rho_v + \rho_{\epsilon_j} + \rho_{\epsilon_k}} s_k + \frac{\rho_{\epsilon_j}}{\rho_v + \rho_{\epsilon_j} + \rho_{\epsilon_k}} s_j + \frac{\rho_v}{\rho_v + \rho_{\epsilon_j} + \rho_{\epsilon_k}} \overline{v},$$

precision $\rho_{j+k} = \rho_v + \rho_{\epsilon_j} + \rho_{\epsilon_k}$, and variance $\sigma^2_{j+k} = 1/\rho_{j+k}$. The new precision of the joint signal is the sum of the individual precisions. Since $\rho_{\epsilon_j} \geqslant 0$ and $\rho_{\epsilon_k} \geqslant 0$, the joint signal must be more informative than any of the previous signals alone, i.e., $\rho_{j+k} \geqslant \rho_j$ and $\rho_{j+k} \geqslant \rho_k$ (or $\sigma^2_{j+k} \leqslant \sigma^2_j$ and $\sigma^2_{j+k} \leqslant \sigma^2_k$).[8] The process of data aggregation is summarized into a single parameter, the joint precision (or its reciprocal, the variance), which can be decomposed in the precision of the individual signals.

In order to reduce the number of parameters and simplify expressions, it is convenient to let $\rho_v = 0$.

2.3. ***Data aggregation - information overlapping***

The observation made in the previous subsection that the precision of the new signal adds linearly to the existing precision is true if the signals are independent and there is no information overlapping. However, if firms j and k decide to join their data sets, we may have to remove some information overlapping that might exist in the data sets. For instance, the firm j data set may contain information about location and income, while the firm k data set may contain information about gender and income. Therefore, both data sets overlap on income.[9]

The objective of this section is not to develop a theory of information overlapping, but to introduce the reader to this important issue. This subsection can be skipped in a first read.

Information overlapping might be related with the idea of correlation, but it is a different concept. Information overlapping is closely related with the idea of comple-

[8]The new variance becomes: $\sigma^2_{j+k} = \sigma^2_v \sigma^2_{\epsilon_j} \sigma^2_{\epsilon_k} / (\sigma^2_v \sigma^2_{\epsilon_j} + \sigma^2_v \sigma^2_{\epsilon_k} + \sigma^2_{\epsilon_j} \sigma^2_{\epsilon_k})$, with $\sigma^2_{j+k} \to \sigma^2_{\epsilon_j} \sigma^2_{\epsilon_k} / (\sigma^2_{\epsilon_j} + \sigma^2_{\epsilon_k})$ for $\sigma^2_v \to \infty$ or $\rho_v \to 0$.

[9]See Westenbroek et al. (2020) for a study on the strategic implications of data overlapping and externalities among competing data aggregators and data providers.

mentarity between data sets. Moreover, independent signals do not guarantee that there is no information overlapping. Firms may collect data through independent means, but they may end up having similar data. For instance, two independent data sets may have information about the age and gender of consumers.

In the big data and AI context, the information overlapping issue may be even trickier. For instance, if two data sets show that the consumer prefers the blue color that may not necessarily be information overlapping. This information may be confirming and increasing the precision with which the algorithm is sure that the consumer prefers the blue color. Nonetheless, in general, we must expect some degree of information overlapping between two data sets.

In the case of information overlapping, the precision of the joint signals can be represented/modeled by the sum of the individual precisions minus the overlapping of the information in both signals, i.e., $\rho_{j+k} = \rho_v + \rho_{\epsilon_j} + \rho_{\epsilon_k} - h_{jk}$, where $h_{jk} \in [0, \min\{\rho_{\epsilon_j}, \rho_{\epsilon_k}\}]$ captures the amount of data overlapping in terms of precision associated with the merger of the two data sets. The maximum overlapping occurs when the relevant information contained in the smaller data set is also contained in the larger data set, i.e., $h_{jk} = \min\{\rho_{\epsilon_j}, \rho_{\epsilon_k}\}$. In this case, it is natural to expect that the firm with the better data has no incentives to share information, as there is nothing new in the other firm's data set. On the other hand, the minimum overlapping occurs when the information in both firms' data sets is completely new to each other, i.e., $h_{jk} = 0$, as we assume in Section 2.2. Nonetheless, in general, we must expect h_{jk} to take intermediate values.

Information overlapping, as presented in this paper, has some analogy with the concept of joint entropy of two correlated random variables in information theory. The joint entropy is the sum of the individual entropies minus the mutual information (e.g., see MacKay, 2003). However, the true function and distribution of h_{jk} might be unknown.[10] Nonetheless, we must note that the results and arguments in this paper do not depend on the knowledge about the function h_{jk}. We can obtain similar insights with no information overlapping by setting $h_{jk} = 0$, using the framework of Section 2.2, and exploring data asymmetries between firms in terms of the signals precision, i.e., by varying ρ_{ϵ_j} for fixed ρ_{ϵ_k}, and vice versa.

[10] h_{jk} should have the following properties. If $\rho_{\epsilon_j} \to 0$ or $\sigma^2_{\epsilon_j} \to \infty$, then $h_{jk} \to 0$ and $\rho_{j+k} \to \rho_v + \rho_{\epsilon_k}$, i.e., no information implies no overlap. If $\rho_{\epsilon_j} \to \infty$ or $\sigma^2_{\epsilon_j} \to 0$, then $h_{jk} \to \rho_{\epsilon_k}$ and $\rho_{j+k} \to \infty$, i.e., the information in a fully informative data set must contains the information in the less informative data set.

3. Data, preferences, demand and equilibrium

This section models how that information transforms into decisions, equilibrium quantities, prices and profits by proposing two approaches to model data in competitive markets.

We consider a representative consumer that represents an aggregate of myopic consumers, which derives utility from the consumption of a representative good Q, where $Q = q_1 + ... + q_n$ denotes the aggregate consumption of the representative good, and q_j denotes the consumption of the quantity of firm $j = 1, ..., n$. The utility derived from the consumption of the representative good increases with the quantity in a decreasing way, i.e., $u'(Q) \geqslant 0$ and $u''(Q) \leqslant 0$. The representative consumer also derives utility from the consumption of a numeraire good Q_0.

In the AI and big data setting in this paper, the meaning of variety is related with the firm's strategic targeting or positioning given the data available about the representative consumer preferences (e.g., browsing history, location, gender, habits, economic conditions, etc.). In this context, targeting/positioning is characterized by a parameter $v_j \in \mathbb{R}$. The representative consumer has an optimal/preferred targeting/positioning $v \in \mathbb{R}$, which is not perfectly known by firms, and incurs in an utility loss that increases with the distance between the preferred and the consumed targeting/positioning. In this context, it is natural to consider the quadratic loss function $(v_j - v)^2$ (see Subsection 3.1), or the absolute value loss function $|v_j - v|$ (see Subsection 3.2), which is then integrated into the conditional indirect utility function.

In our context, the conditional indirect utility formulation must capture big data and information processing effects and, simultaneously, should:

1. Be able to extend the existing and commonly used demand models. The objective is not to build a new theory, but to integrate data in the existing literature.
2. Be intuitive and render tractable and simple expressions. The objective is to offer a framework that can be easily applied by researchers and practitioners into a large variety of problems.[11]
3. Be such that the process of information and beliefs update described in Section 2 can be embedded in a linear and simple framework. Otherwise, we have to

[11] For instance, to study the implications of information and big data on competition, market structure, mergers and acquisitions, antitrust policy, welfare, among other problems. The study of these problems tends to become increasingly complex as more features are added. For that reason, it is important to start simple.

deal with complex non-linear filtering problems, which add no value in terms of ideas and intuition.

In this context, the indirect utility formulation that delivers linear demand functions seems appropriate. In what follows, this paper proposes two approaches to model data in competitive markets in the linear demand framework.

3.1. ***The additive model of data***

We start by considering the case in which the information component enters into the utility function in an additive way. In this context, consider the following functional form for the utility function:

$$u_i(Q, Q_0, v_k, v) = aQ - bQ^2/2 - \alpha \sum_{k=1}^{n} (v_k - v)^2 q_k + Q_0, \tag{1}$$

where the parameter a controls the rate at which the utility increases with the aggregate consumption, the parameter b controls the rate at which the marginal utility decays with the aggregate consumption, and α captures the importance of the targeting/positioning match for the representative consumer.

The utility loss is quadratic with respect to the distance between the preferred and the consumed targeting/positioning, and increases with the consumed quantity, i.e., $(v_j - v)^2 q_j$. In other words, the dis-utility from the mismatch is more severe the more units are consumed.

This utility formulation (1) delivers linear demand functions. In this context, let $\sum_{k=1}^{n} p_k q_k + Q_0 = m$ be the representative consumer budget constraint, where m denotes the income of the representative consumer, and p_j denotes the price of the targeting/positioning offered by firm j. Since the representative consumer maximizes the utility subject to the budget constraint, the solution to this constrained optimization problem returns the following system of inverse linear demand functions:

$$p_j = a - bQ - \alpha(v_j - v)^2, \tag{2}$$

for $j = 1, ..., n$, with $p_j \geqslant 0$ for $j = 1, ..., n$.[12]

[12]Following the discussion in Section 2, from the firms' perspective, v is a random variable and each firm has a private distribution of beliefs about the consumers preferred variety/positioning v. In this context, each firm will offer a variety/positioning v_j that according to their beliefs would minimize the consumer utility loss, i.e., each firm will offer the variety/positioning $v_j = E[v|I_j]$, where I_j denotes firm $j's$ information about the representative consumer, and $(v_j - v)^2$ captures the firm $j's$ uncertainty about v.

In expected terms, $E[(v_j - v)^2 | I_j] = \sigma_j^2 = 1/\rho_j$ for $j = 1, ..., n$. Then, the system of inverse demand functions (2) is a system of expected inverse demand functions:

$$p_j = a - bQ - \alpha/\rho_j, \tag{3}$$

for $j = 1, ..., n$, with $p_j \geqslant 0$ for $j = 1, ..., n$ (with some abuse of notation $p_j = E[p_j | I_j]$). In this model, each firm sells its product at different prices, because each firm holds different information about the representative consumer preferred targeting/positioning and is targeting/positioning the representative consumer in different ways. The more precise information about the representative consumers' preferences each firm has, the closer its targeting/positioning gets from the consumer preferences and higher its demand.[13,14]

We can now compute the equilibrium prices and quantities. In the Cournot model, each firm maximizes the profit function $\pi_j = (p_j - c)q_j$ with respect to quantities q_j for $j = 1, ..., n$, with p_j replaced by expression (3), where c is the constant marginal cost per unit.[15] The solution of the associated system of n first order conditions delivers the following result.

Proposition 1. *In the additive n-firms model, suppose that $a - c \geqslant \alpha(n/\rho_j - \sum_{k \neq j}^{n} 1/\rho_k)$ for all j, then each firm j equilibrium quantities, prices and profits are given by the following expressions:*

$$q_j = \frac{a - c - \alpha(n/\rho_j - \sum_{k \neq j}^{n} 1/\rho_k)}{b(n+1)}, \tag{4}$$

$$p_j = \frac{a + nc - \alpha(n/\rho_j - \sum_{k \neq j}^{n} 1/\rho_k)}{(n+1)}, \tag{5}$$

[13] Since firms compete with each other, we should be precise about how much information each firm knows about the other firms. If we assume that firm j has imperfect information about how much the other firm k knows about the representative consumer preferences, i.e., the value of ρ_k for all $k \neq j$, and vice versa. Then, we must specify an additional system of beliefs about how much each firm knows about the other firms and so on. In this paper, we abstract from these considerations and assume that each firm j knows the competitors ρ_k from observing their market prices, which is not the same as knowing their data sets.

[14] Briefly, the timing would be as if: First, the representative consumer maximizes the utility subject to the budget constraint. Second, each firm variety/positioning is determined by the quality of their data (precision of their signal). Third, firms compete a la Cournot.

[15] In this paper, we focus on the Cournot model. However, the impact of information may depend on whether we consider Cournot or Bertrand competition (Vives, 1984).

and,

$$\pi_j = \frac{(a - c - \alpha(n/\rho_j - \sum_{k\neq j}^{n} 1/\rho_k))^2}{(b(n+1))^2}, \tag{6}$$

for $j = 1, ..., n,$ *respectively.*

The proof follows from the discussion preceding the statement of the Proposition. The condition $a - c \geqslant \alpha(n/\rho_j - \sum_{k\neq j}^{n} 1/\rho_k)$ for all j guarantee that quantities (and prices) are non-negative, and that there are no corner solutions or equilibrium existence problems.

Consequently, we have the following result regarding an increase in the information precision.

Corollary 1. *An information quality improvement in the data of firm j in Proposition 1, increases firm j equilibrium quantities, prices and profits, and decreases all other firms $k \neq j$ equilibrium quantities, prices and profits.*

In other words, an exogenous unilateral information quality improvement in the data of firm j, of any magnitude, has always a positive effect in terms of quantities, prices and profits of firm j, and a negative effect in terms of quantities, prices and profits of the other firms $k \neq j$.

In our setting, firms always benefit from better data. However, if gathering and processing data would be costly, there could be some trade-off. Nonetheless, the message is clear regarding the impact of data on the firms' key variables. The above result also implies that data has value and can be priced, which is linked with the associated benefits.

3.2. ***The multiplicative model of data***

We now consider the case in which the information component enters the utility function through the rate at which the marginal utility decays with the quantity consumed. Intuitively, the multiplicative model adds interaction effects between competing targeting/positioning approaches.

In this context, we consider the following functional form for the utility function:

$$u_i(Q, Q_0, v_k, v) = aQ - b(\sum_{k=1}^{n} |v_k - v| q_k)^2/2 + Q_0, \tag{7}$$

where, as before, the parameter a controls the rate at which the utility increases with the aggregate consumption and the parameter b controls the rate at which the marginal utility decays with the aggregate consumption.

The utility loss is now the absolute distance between the preferred and the consumed targeting/positioning. This term interacts with the quantity consumed, i.e., $|v_j - v|q_j$. This effect was already present in the additive formulation, but now the mismatch term is placed inside the square, which opens the possibility for interaction effects.

The utility formulation (7) delivers linear demand functions, but now the information component enters in a multiplicative way in the slope of the inverse linear demand function. In this context, let $\sum_{k=1}^{n} p_k q_k + Q_0 = m$ be the representative consumer budget constraint, where m denotes the income of the representative consumer and p_j denotes the price per unit of the product offered by firm j.

The representative consumer solution of this constrained optimization problem returns the following system of inverse demand functions:

$$p_j = a - b|v_j - v| \sum_{k=1}^{n} |v_k - v|q_k, \tag{8}$$

for $j = 1, ..., n$, with $p_j \geqslant 0$ for $j = 1, ..., n$.[16]

Since firms have different data about the representative consumer preferences, the varieties/position v_j and v_k are different, in pairwise terms, we can write:

$$|v_j - v||v_k - v| = |(v_j - v)(v_k - v)|,$$

which is the absolute value covariance between targeting/positioning j and k, i.e., $|\sigma_{jk}|$. Therefore, the system of inverse demand functions (8) is a system of expected inverse demand functions that can be written in terms of the absolute values of the covariances as follows:

$$p_j = a - b \sum_{k=1}^{n} |\sigma_{jk}|q_k, \tag{9}$$

for $j = 1, ..., n$, with $p_j \geqslant 0$ for $j = 1, ..., n$ (with some abuse of notation $p_j = E[p_j|I_j]$). Therefore, if the firms' targeting/positioning j and k have high covariance, the impact that each has on the other is high.

The covariance absolute value means that it is irrelevant whether the correla-

[16] Following the discussion in Section 2, since from the firms' perspective v is a random variable and each firm has a private distribution of beliefs about the consumers preferred variety/positiong v, each firm will offer a variety/positioning v_j that according to their beliefs would minimizes the consumer utility loss, i.e., each firm will offer the variety $v_j = median = E(v|I_j)$, where I_j denotes the firm j information about the representative consumer. In order to understand this observation, just note that the absolute deviation is not minimized at the mean, but at the median. However, since beliefs are normally distributed, the mean equals the median.

tion between the firms' strategic targeting/positioning is negative or positive. The covariance always increases when two firms' approach each other, independently on whether they are approaching the representative consumer from the opposite direction (negative covariance case) or from the same direction (positive covariance case).

Therefore, the multiplicative model delivers two effects associated with better data. The first effect is an increase in the information quality, which is materialized in better precision. The second effect is a correlation effect between the firms' targeting/positioning.[17]

In order to obtain information in terms of correlations and data improvements, we apply the Cauchy-Schwarz Inequality $|\sigma_{jk}| \leqslant \sigma_j \sigma_k$ for $j \neq k$ and $|\sigma_{jk}| = \sigma_j^2$ for $j = k$. In this context, we write $|\sigma_{jk}| = \sigma_j \sigma_k s_{jk}$ or $|\sigma_{jk}| = s_{jk}/\sqrt{\rho_j \rho_k}$, where $0 \leqslant s_{jk} \leqslant 1$ for $j \neq k$, with $s_{jk} = s_{kj}$ and $s_{jj} = 1$ for $j = k$, is the absolute value of the Pearson product-moment correlation coefficient (henceforth, correlation coefficient) between firms' j and k varieties. This parameter captures the degree of linear relation between varieties j and k, which may also be related with the idea of horizontal substitution between the product varieties of firm j and k (Dixit, 1979; Singh et al., 1984; Shapley and Shubik, 1969). Consequently, the system of inverse demand functions (9) can be written in terms of the firms' precision and the absolute value correlation coefficients as follows:

$$p_j = a - b\sigma_j \sum\nolimits_{k=1}^{n} s_{jk}\sigma_k q_k = a - b/\sqrt{\rho_j} \sum\nolimits_{k=1}^{n} s_{jk} q_k/\sqrt{\rho_k}, \tag{10}$$

for $j = 1, ..., n$. Therefore, the more two firms are correlated, the higher the impact that each firm has on the other firm, and the opposite otherwise.

In order to simplify the analysis, we derive the equilibrium prices and quantities for the case of two firms, i.e., $n = 2$. A general number of firms involves complex expressions. We present a numerical example with three firms in Section 4.3 below. In the Cournot model, each firm maximizes the profit function $\pi_j = (p_j - c)q_j$ with respect to quantities q_j for $j = 1, ..., n$, with p_j replaced by expression (10), where c is the constant marginal cost per unit. The solution of the associated system of two first order conditions delivers the following result.

Proposition 2. *In the multiplicative* 2*-firms model, suppose that* $2\sqrt{\rho_j} \geqslant s\sqrt{\rho_k}$ *for all* j*, then each firm* j *equilibrium quantities, prices and profits are given by the*

[17]For instance, when two firms exchange data, they are improving their precisions, but they are also becoming more correlated in strategic terms. In our model, they become perfectly correlated because they have the same information.

following expressions:

$$q_j = \frac{(a-c)\sqrt{\rho_j}(2\sqrt{\rho_j} - s\sqrt{\rho_k})}{b(4-s^2)}, \quad (11)$$

$$p_j = \frac{2(a+c) - (a-c)s\sqrt{\rho_k/\rho_j} - cs^2}{(4-s^2)}, \quad (12)$$

and,

$$\pi_j = \frac{(a-c)^2(2\sqrt{\rho_j} - s\sqrt{\rho_k})^2}{b(4-s^2)}, \quad (13)$$

respectively, for $j = 1, 2$, *where* $s = s_{12} = s_{21} \in [0, 1]$ *is the degree of correlation between the firm's* 1 *and* 2 *varieties.*[18]

The proof follows from the discussion preceding the statement of the Proposition. In the case of $n = 2$ firms, the conditions $2\sqrt{\rho_1} \geqslant s\sqrt{\rho_2}$ and $2\sqrt{\rho_2} \geqslant s\sqrt{\rho_1}$ guarantee that quantities (and prices) are non-negative, and that there are no corner solutions or equilibrium existence problems.

The following result summarizes the effect of information in the basic market variables of firm j and its opponents.

Corollary 2. *An information quality improvement in the data of firm j in Proposition 2, increases firm j equilibrium quantities, prices and profits, and decreases the other firm $k \neq j$ equilibrium quantities, prices and profits.*

This result is similar to Corollary 1 in the additive model. An unilateral information quality improvement in the data of firm j, of any magnitude, benefits firms j and penalizes its opponents. A Firm always benefits from better data.

[18] The same equilibrium written in terms of variances and covariances, with $n = 2$, would be the firm j equilibrium quantities, prices and profits given by:

$$q_j = \frac{(a-c)|\sigma_{jj}|(2|\sigma_{kk}| - |\sigma_{jk}|)}{b(4|\sigma_{jj}||\sigma_{kk}| - |\sigma_{jk}|^2)}, \ p_j = \frac{2(a+c)|\sigma_{jj}||\sigma_{kk}| - (a-c)|\sigma_{jj}||\sigma_{jk}| - c|\sigma_{jk}|^2}{b(4|\sigma_{jj}||\sigma_{kk}| - |\sigma_{jk}|^2)},$$

and,

$$\pi_j = \frac{(a-c)^2|\sigma_{jj}|(2|\sigma_{kk}| - |\sigma_{jk}|)^2}{b(4|\sigma_{jj}||\sigma_{kk}| - |\sigma_{jk}|^2)^2},$$

respectively, for $2|\sigma_{jj}| \geqslant |\sigma_{jk}|$ and $j = 1, 2$, where $|\sigma_{jk}| = |\sigma_{kj}|$ and $|\sigma_{jj}| = \sigma_j^2$.

The result has implicit that the correlation effect remains constant after an information improvement. In reality, in the case that firms j and k compete in the same market, an unilateral information improvement may lead to an increase or a decrease in correlation, depending on whether both firms targeting/positioning become closer or more distant from each other, respectively.

4. Horizontal data exchange - Data exchange between competing firms

In this section, we consider the case in which two competing firms exchange data about the representative consumer preferences, without side payment.[19] Following the discussion in Section 2, if firm j and k join their data sets, then both firms will have the same data and information precision, i.e., the same ρ_{j+k}. In this context, we analyze how differences in the firms' information precision affect the incentives to exchange data among competing firms. This is an important aspect because firms with different data sets contribute differently to the joint data set, which results in differences in terms of benefits. In the multiplicative model, there are also strategic correlation effects.

*4.1. **The Additive Model - Data exchange between competing firms***

In the additive model, it is mutually profitable for two or more firms to exchange data with each other providing the information gains are sufficiently large and the firms data sets are of similar quality. Otherwise, if the two competing firms are sufficiently asymmetric in terms of data, the firm with better data may have no incentives in exchange data with the firm with weaker data, because the firm with better data would be giving away its information advantage to a competitor.

In this context, we must compare the firms' profits before and after the data exchange.

Proposition 3. *In the additive n-firms model, for $\rho_j \geqslant \rho_k$, firm j has incentives to exchange data with firm k if:*

$$n\frac{\rho_{j+k} - \rho_j}{\rho_j} \geqslant \frac{\rho_{j+k} - \rho_k}{\rho_k}. \tag{14}$$

[19] Since mergers and acquisitions between competing firms are usually subject to permission by the market governing bodies and competition authorities, data exchange between competing firms seems an appealing possibility. Despite the fact that these practices are illegal according to the data protection laws, e.g., the EU General Data Protection Regulation 2016/679 (GDPR), they are difficult to monitor by the market governing bodies and competition authorities because evidence can be easily erased and it is difficult to identify which data has been used to teach and calibrate an AI algorithm.

For $\rho_j \leqslant \rho_k$, firm j has always incentives to exchange data with firm k.

In other words, the firm with better data has incentives (i.e., obtains higher profits) to exchange data with the firm with worst data when inequality (14) is satisfied, i.e., providing that the information gains are relatively stronger than the competitor information gains. The firm with the worst data has always incentives to exchange data with the firm with better data.

Note also that the larger the number of firms in the market, i.e., higher the competition, the higher the incentive for two firms to exchange data among them.

A particular case of Proposition 3 is the symmetric case.

Corollary 3. *In the symmetric case, i.e., $\rho_j = \rho_k$, firms always have incentives to exchange data with each other.*

In the symmetric case, inequality (14) is always satisfied because $n \geqslant 1$ and $\rho_{j+k} \geqslant \rho_j$.

Regarding the other market variables, we have the following result.

Proposition 4. *In the additive n-firms model, if firms have incentives to exchange data, then both firms' equilibrium quantities and prices increase.*

The same condition (14) in Proposition 3 applies to Proposition 4. Therefore, if two firms exchange data, they will increase sales and charge higher prices. These results show that the incentives to exchange data among firms are quite strong.

4.2. ***The Multiplicative Model - Data exchange between competing firms***

In addition to data improvement effects, the multiplicative model adds correlation effects between firms targeting/positioning, which leads to higher market competition between the firms involved in the data exchange. If two firms improve their information by joining their data sets, they are simultaneously improving their precision, but they are also becoming more correlated in strategic terms because they have exactly the same data about the consumers' preferences (e.g., browsing history, location, gender, consumption habits, economic conditions, etc.). Consequently, they are most likely to become aligned in strategic terms. Their strategic targeting/positioning (e.g., product/service, strategies, objectives, etc.), become more similar, which increases the competition between them. Intuitively, the correlation effect is similar to a substitution effect (Dixit, 1979; Singh et al., 1984; Shapley and Shubik, 1969).

In this paper, the correlation effect is represented by an increase in the parameter s, from s to $s' \geqslant s$, after firms have exchanged data.

Proposition 5. *In the multiplicative 2-firms model, firm j has incentives to exchange data with firm k if:*

$$\frac{\rho_{j+k}}{(2+s')^2} \geqslant \frac{(2\sqrt{\rho_j} - s\sqrt{\rho_k})^2}{(2+s)^2(2-s)^2}. \tag{15}$$

In other words, firms have incentives to exchange information when inequality (15) is satisfied, i.e., providing that the information gains are enough relatively to the competitor information gains (i.e., the differences between ρ_{j+k} and ρ_j, and between ρ_{j+k} and ρ_k, respectively), and the correlation effect is not too strong (i.e., the difference between s' and s). Note also that in the multiplicative model, contrary to the additive model, if the correlation effect is too strong, it is not always true that the firm with weaker data has incentives to exchange data with the firm with better data.

In the multiplicative model, in order for data exchange to occur, it is important that both firms are similar in terms of information/data quality and that the correlation effect after the data exchange does not become too large. The symmetric case illustrates this point.

Corollary 4. *In the symmetric case, i.e., $\rho_j = \rho_k$, firms have incentives to exchange data if the information gains compensate the correlation effect.*

In the symmetric case, i.e., $\rho_j = \rho_k$, inequality (15) becomes $\rho_{j+k}/(2+s')^2 \geqslant \rho_j/(2+s)^2$, which is satisfied only when ρ_{j+k} is sufficiently larger than ρ_j (i.e., the information gains), and s' is not too large with respect to s (i.e., the correlation effect).[20]

If there is no correlation effect, i.e., if $s' = s$, the additive and multiplicative models deliver similar results.

In the multiplicative model, when firms join their data sets, quantities depend on the information quality and on the correlation effect, but prices become independent of the information quality and depend only on the correlation effect.

Regarding the basic market variables, we have the following result.

[20] In the symmetric case, the condition that guarantees that the consumer surplus improves after data exchange is $\rho_{j+k}(1+s')/(2+s')^2 \geqslant \rho_j(1+s)/(2+s)^2$, which is always satisfied if there exist data exchange incentives because $s' \geqslant s$. Intuitively, the correlation effect affects the consumer surplus less than the profits, which is not surprising, as consumers benefit from higher competitions. See Section 4.3 below.

Proposition 6. *In the multiplicative* 2*-firms model, if firms have incentives to exchange data, then both firms' equilibrium quantities increase. The equilibrium price of the firm with better data decreases after the data exchange.*

In other words, the firm with better data increases quantities and decreases prices because of the increasing competition effect, while the firm with worse data increases quantities and may even increase prices if the obtained information gains are large enough relative to the correlation effect. This observation highlights that the firm with the worst data is the one that most benefits from data exchange.

To summarize, the results suggest that firms with similar information quality have strong incentives to exchange data between them, but not between firms with significantly different levels of information quality. This is an important contribution of this paper, that has potentially strong consequences for the regulation of data driven markets, regarding data dominance and competitiveness. The results are similar across both models, but the multiplicative model adds a correlation effect that increases competition between the firms involved in the data exchange.

4.3. ***Numerical Analysis***

This section presents a series of numerical examples to show how data exchange between firms competing in the same market impact on:

1. market variables like equilibrium prices, quantities and profits - Table 1.
2. welfare measures like traded quantities, consumer surplus and total profits (producers' surplus) - Table 2.
3. market concentration variables like the market shares and the HHI - Table 3.

In this context, consider three equal firms, but with potential different data sets about the representative consumer preferences. The three-firm case is interesting for observing how data exchange between firms 1 and 2 affect each other, but also how it affects firm 3, the firm not involved in the data exchange.

In order to make the analysis clearer and to have a common denominator, firms start with the same information precision $\rho_1 = \rho_2 = \rho_3 = 0.5$ (with $\rho_v = 0$ implying that $\rho_j = \rho_{\epsilon_j}$ for all j). Firms have the same information quality, but different data sets.

Subsequently, we vary firm 1 information precision from 0.5 to 0.75 and 1.00, with all other parameters fixed at $\rho_2 = \rho_3 = 0.5$, $a = 1$, $c = 0$, $b = 1$ and $s_{12} = s_{13} = s_{23} = 0.5$.[21] In order to not be repetitive, we focus on the multiplicative model

[21]The parameters are chosen to avoid corner solutions or equilibrium existence problems. The obtained results are sufficiently representative and are robust to the choice of parameters.

while commenting on the additive model. With no data overlapping, after the data exchange, firms' 1 and 2 joint precision becomes $\rho_{1+2} = \rho_1 + \rho_2$ and their correlation coefficient increases from $s_{12} = 0.5$ to $s'_{12} = 1$, because they have exactly the same information, while firm's 3 information precision and correlation coefficients remain constant.

In the additive model, as shown in Propositions 3 and 4, providing that firms have data sets of similar quality, firms 1 and 2 information quality improvements allow them to charge higher prices, increase sales and obtain higher profits. Things only change when the quality differences between data sets are sufficiently large. In this case, the asymmetry in terms of information quality is so large that firm 1 has no incentives to exchange information with firm 2 because it will obtain lower profits. The same happens in the multiplicative model, as shown in Propositions 5 and 6, with the difference that firms tend to charge lower prices (compare row (1) with (2) and row (3) with (4) in Table 1). Competition increases because the firms' 1 and 2 varieties become perfectly correlated. Nonetheless, both firms' profits increase. Things only change when the information quality differences between data sets is so large that firm 1 has no incentives to exchange data with firm 2, in which case the firm 1 profits fall after the data exchange (compare row (5) with (6) in Table 1).[22]

In both models, the firm with better information obtains higher profits, and the loser is firm 3, the firm that is not participating in the data exchange. Firm's 3 sales, prices and profits fall as the quality of information of firms 1 and 2 improves.

In this context, a crucial question is the impact of data exchange on welfare (see Footnote 20). Table 2 shows an increase in consumers surplus, which is driven by the fact that producers are delivering varieties closer to the representative consumer preferences, which is leading to higher consumption.[23] This result is particularly strong in both models and shows the existence of positive effects for consumers associated data exchange (compare row (1) with (2), compare row (3) with (4), and row (5) with (6) in Table 2). Similarly, the producers' surplus also increases because of the higher profits.

[22]However, recall that in the multiplicative model, if the benefits from data exchange are small relative to the correlation effect, data exchange is not beneficial, even when firms are similar in terms of information quality (not show in Table 1). This situation is more likely to occur when the overlapping between data sets is large.

[23]The representative consumer surplus is given by expressions (1) and (7) in the additive and multiplicative models, respectively, with $Q_0 = m - \sum_{j=1}^{n} p_j q_j$ and the normalization $m = 0$. This approach allows us to recover the effect of price variations in the consumers' surplus. The producers

	ρ_1	ρ_2	ρ_{1+2}	q_1	p_1	π_1	q_2	p_2	π_2	q_3	p_3	π_3
(1)	0.50	0.50		0.167	0.333	0.056	0.167	0.333	0.056	0.167	0.333	0.056
(2)			1.00	0.299	0.299	0.090	0.299	0.299	0.090	0.144	0.288	0.042
(3)	0.75	0.50		0.281	0.374	0.105	0.154	0.308	0.048	0.154	0.308	0.048
(4)			1.25	0.383	0.306	0.117	0.383	0.306	0.117	0.129	0.258	0.033
(5)	1.00	0.50		0.398	0.398	0.159	0.144	0.287	0.041	0.144	0.287	0.041
(6)			1.50	0.467	0.311	0.145	0.467	0.311	0.145	0.115	0.231	0.027

Table 1: **Data exchange effects on the market variables - multiplicative model:** for varying $\rho_1 = 0.5$, 0.75 and 1.00, with fixed $\rho_2 = \rho_2 = 0.5$, $\rho_v = 0$, $a - c = 1$, $b = 1$ and $s_{12} = s_{13} = s_{23} = 0.50$. In all cases, after firms' 1 and 2 data exchange, their information quality parameter increases to $\rho_{1+2} = \rho_1 + \rho_2$, and the correlation coefficient between these two firms increases to $s'_{12} = 1.00$.

	ρ_1	ρ_2	ρ_{1+2}	Q	CS	PS
(1)	0.50	0.50		0.500	0.167	0.167
(2)			1.00	0.743	0.261	0.221
(3)	0.75	0.50		0.589	0.194	0.200
(4)			1.25	0.894	0.313	0.268
(5)	1.00	0.50		0.686	0.222	0.241
(6)			1.50	1.049	0.366	0.317

Table 2: **Data exchange effects on welfare measures - multiplicative model:** for varying $\rho_1 = 0.5$, 0.75 and 1.00, with fixed $\rho_2 = \rho_2 = 0.5$, $\rho_v = 0$, $a - c = 1$, $b = 1$ and $s_{12} = s_{13} = s_{23} = 0.50$. In all cases, after firms' 1 and 2 data exchange, their information quality parameter increases to $\rho_{1+2} = \rho_1 + \rho_2$, and the correlation coefficient between these two firms increases to $s'_{12} = 1.00$.

In this context, another interesting question is the impact of data exchange in terms of market concentration. Table 3) shows that the improvement in the information quality of firms 1 and 2 tends to increase their market shares, but not always. For instance, rows (3) and (4) in Table 3 show a fall in the market share of firm 1 after a profitable data exchange with firm 2. Nonetheless, data exchange tends to increase market power and market concentration, as measured by the Herfindahl–Hirschman Index (HHI).[24] These observations are also true for the additive model.

	ρ_1	ρ_2	ρ_{1+2}	s_1	s_2	s_3	HHI
(1)	0.50	0.50		0.333	0.333	0.333	0.333
(2)			1.00	0.403	0.403	0.194	0.362
(3)	0.75	0.50		0.476	0.262	0.262	0.364
(4)			1.25	0.428	0.428	0.144	0.387
(5)	1.00	0.50		0.581	0.210	0.210	0.425
(6)			1.50	0.445	0.445	0.110	0.408

Table 3: **Data exchange effects on market concentration measures - multiplicative model:** for varying $\rho_1 = 0.5$, 0.75 and 1.00, with fixed $\rho_2 = \rho_2 = 0.5$, $\rho_v = 0$, $a - c = 1$, $b = 1$ and $s_{12} = s_{13} = s_{23} = 0.50$. In all cases, after firms' 1 and 2 data exchange, their information quality parameter increases to $\rho_{1+2} = \rho_1 + \rho_2$, and the correlation coefficient between these two firms increases to $s'_{12} = 1.00$.

The observations made regarding prices, quantities, profits, consumers and producers' surplus, market shares and the HHI are also true when there are more than two firms involved in the data exchange.

Note that the examples presented exaggerate the information effects in order to highlight the impact of data. In more realistic settings, with data overlapping, the information effect might not be so strong, and consequently the incentives to

surplus is the sum of the firms' profits, i.e., $PS = \sum_{j=1}^{n} \pi_j$.

[24]The HHI is an indicator of market competition. It weighs the firms' market shares in relation to the industry. The HHI is the sum of the squares of the market shares of the firms within the industry, i.e., each firm's market share is weighted by its own market share. Formally, $HHI = 10,000 \sum_{j=1}^{n} s_j^2$ where $s_j = q_j/Q$ for $j = 1, ..., n$. An increase in the HHI indicates a decrease in competition, and vice versa.

exchange data weaker, because the correlation effect may be the dominant factor. Nonetheless, the direction of the data exchange effects is clear.

5. Data exchange between non-competing firms

The previous results show that two (or more) firms competing in the same market have incentives to share information when the mutual information gains are of similar magnitude and the correlation effects are not too strong. Data exchange seems to be beneficial for firms under very reasonable conditions. Nonetheless, one may argue that firms in the same industry may have similar data about consumers, and consequently the overlap between data sets is very large, and for that reason the incentives and benefits of data exchange are weaker. This argument may make sense.

5.1. ***Independent data exchange***

Along this line of reasoning, we question what may happen if two companies in different industries would merger their data sets. The answer is that the effect is likely to be even stronger. The reason is that the complementarity between data sets is likely to be high, which would lead to low overlapping between data sets. Moreover, since firms operate in different markets there may be no strategic correlation effects. Therefore, larger incentives to data exchange. This effect could be mimicked in our model by an increase in the information quality of the involved firms relative to their specific market, without a similar improvement in any of its competitors.

The results in Proposition 4 for the additive model and the results in Proposition 6 for the multiplicative model may formalize this intuition. With no correlation effects, unilateral information improvements of any magnitude, benefit that firm and penalize its competitors. A firm always benefits from better data. In this context, we should expect data to become a crucial strategic variable in competitive markets, like quality, technology or price.

This intuition can be seen in Tables 1, 2 and 3 by comparing the improvement in firm's 1 position relatively to its competitors from rows (1), (3) and (5). For instance, if firm 1 doubles the other firm's information precision, then it is able to obtain about 60% of the market share instead of the initial 33%.

5.2. ***Vertical data exchange***

Finally, the last possibility regarding data exchange concerns the incentives of vertical integrated firms to exchange data. We call vertical data exchange when two firms vertically integrated in customer/supplier relations exchange data.

The model in this paper is not suitable to capture the associated data exchange effects. However, in those cases, data exchange is similar to data integration. In addition to information effects, the model must be able to generate data integration effects that operate directly in the efficiency of the firms' production function or in the cost function, in line with the concept of "Industry 4.0"and "Internet of Things (IoT)".

Nonetheless, we can consider a Gaussian signals model similar to the one in this paper, which conveniently reduces the data component into a single parameter. The key aspect in a vertical data exchange is the reduction of uncertainty over the demanded/supplied quantities, which is achieved either by improving the information about the quality, durability and maintenance of the produced goods or by improving the information about the customers/suppliers needs and uses.

6. Conclusion

In order to study the AI and big data implications on a large variety of problems concerning competitive markets (e.g., market structure, mergers and acquisitions, antitrust policy, welfare, among others), we need models able to capture data and information effects. In this context, this paper presents a linear demand approach to model data, and studies how data and data exchange affects the equilibrium variables, welfare and concentration in competitive markets.

Data and information processing are modelled as Gaussian signals, in which a single precision parameter summarizes all information in the data set. This framework is then integrated into the representative consumer utility, and consequently into the demand function. In this context, we propose an additive model, in which the data component enters additively in the intercept of the demand function, and a multiplicative model, in which the data component enters multiplicatively into the slope of the demand function.[25] The multiplicative model has the advantage of generating richer trade-offs, but the additive model has the advantage of being more tractable. In general terms, the multiplicative model is superior for applied work and policymaking because those trade-offs, i.e., strategic correlation effects between the firms' targeting/positioning, are real.

In addition to this methodological contribution, the paper also studies the firms' incentives to data exchange among competing firms (horizontal data exchange), and

[25] The Gaussian approach in this paper can also be extended to horizontal and vertical differentiation models, like the Hotelling (1929) and Salop (1979) type models, and their multiple variations (see also Bos and Vermeulen, 2020). The same approach may also be applied to study vertical data exchange and model cost reductions due to better data.

discusses data exchange among independent firms not competing in the same market (independent data exchange). We found that firms have incentives to exchange data when the information gains are strong enough relative to the competitor's information gains, and the strategic correlation effect is not too strong. Firms with good data have lower incentives to exchange data with other firms. We also found that market concentration tends to increase after data exchange, but both consumers and producers seem to benefit because the producers' targeting/positioning become closer to the representative consumer preferences. The losers from this process are the firms excluded from the data exchange.

These observations have potentially strong consequences for the regulation of data driven markets, in terms of data dominance, concentration, market competitiveness and welfare.

The obtained results are in line with the argument that supports the existence of potential benefits for consumers and the society from data exchange, and the creation of trustee platforms to centralize, manage and anonymize data.[26] However, the benefits of data exchange depend crucially on how data is used by firms. In this paper, data is used to better targeting/positioning towards consumer preferences. However, if data is used by firms to exploit consumers or price discriminate then results may change. Further research should aim for a more granular distinction of the market implications and characteristics of the different types of data exchange.

We also note that independent data exchanges between firms operating in independent markets may pose a larger threat in terms of competition and antitrust laws than horizontal data exchanges because of the great complementarity between data sets and the absence of strategic correlation effects. This intuition should be the subject of further research and may open a new perspective about the impact and risks associated with data exchange in competitive markets. Data exchange/sharing is also likely to raise concerns regarding facilitating subsequent collusion in the product market, in a similar way as the sharing/exchange of research and development or technology do.[27]

Finally, data sharing seems to offer potential benefits for institutions, firms and the society, but it is a very sensitive issue. In this context, policy and decision makers

[26] An example of such a platform is the European Union data infrastructure called Gaia-X, in which data and services can be collated and shared in an open, transparent and secure environment.

[27] Information overlapping is closely related with the idea of complementarity between data sets. However, the same data may repeat in two data sets and not necessarily be overlapping but reinforcing. For instance, the gender of the consumer is overlapping, but the preference for the blue color may be reinforcing. These comments are left as suggestions for further research.

need to implement policies that enhance markets and protect consumers, but in order to regulate and incentivize data technologies they must understand the limits and the impact of their decisions. This paper is a step forward in better understanding the role of data in competitive markets.

Acknowledgments: Support from the GRODE, Universitat Rovira i Virgili and the Spanish Ministry of Science and Innovation Project RTI2018-094733-B-100 (AEI/FEDER, UE) and PID2019-105982GB-100 is gratefully acknowledged. I would like to thank Juan Pablo Rincón-Zapatero, as well as several seminars and congress participants for helpful comments and discussions. The usual caveat applies.

Appendix: Proofs of the Results

Proof of Corollary 1. The proof is obtained by verifying the sign of the derivative of q_j, p_j, π_j, q_k, p_k and π_k with respect to ρ_j and $k \neq j$, under the assumption that $a - c \geqslant \alpha(n/\rho_j - \sum_{k \neq j}^{n} 1/\rho_k)$ for all j. ■

Proof of Corollary 2. The proof is obtained by verifying the sign of the derivative of $q_1, p_1, \pi_1, q_2, p_2$ and π_2 with respect to $\sqrt{\rho_1}$ (or ρ_1), under the assumption that $2\sqrt{\rho_1} \geqslant s\sqrt{\rho_2}$ and $2\sqrt{\rho_2} \geqslant s\sqrt{\rho_1}$. ■

Proof of Proposition 3 and Corollary 3. In the additive n-firms model, the profit of firm j in Expression (6) after having exchanged data with firm k becomes:

$$\pi'_j = (a - c - (n-1)\alpha/\rho_{j+k} + \alpha \sum\nolimits_{l \neq k,j}^{n} 1/\rho_l)^2/((n+1)b)^2,$$

for $j = 1, ..., n$. After some algebra, for $\rho_j \geqslant \rho_k$, this profit is higher than the profit before exchanging data, which is given by Expression (6). In other words, $\pi'_j \geqslant \pi_j$ if inequality (14) is satisfied. Otherwise, if $\rho_j \leqslant \rho_k$, inequality (14) is always satisfied. In the symmetric case, i.e., $\rho_j = \rho_k$, inequality (14) is always satisfied because $n \geqslant 1$ and $\rho_{j+k} \geqslant \rho_j$. ■

Proof of Proposition 4. Simply compare $q'_j \geqslant q_j$ and $p'_j \geqslant p_j$ to obtain again inequality (14). ■

Proof of Proposition 5 and Corollary 4. In the multiplicative 2-firms model, the profit of firm j in Expression (13) after exchanging data with firm k becomes:

$$\pi'_j = (a-c)^2 \rho_{j+k}(2 - s')^2/(b(4 - s'^2)^2),$$

for $j = 1, 2$, where $s' = s'_{jk} = s'_{kj} \geqslant s$ is the degree of substitution after the data exchange. After some algebra, this profit is higher than the profit before exchanging data, which is given by Expression (13), i.e., $\pi'_j \geqslant \pi_j$, if inequality (15) is satisfied. In the symmetric case, i.e., $\rho_j = \rho_k$, inequality (15) becomes $\rho_{j+k}/(2+s')^2 \geqslant \rho_j/(2+s)^2$, which is satisfied if ρ_{j+k} is sufficiently larger than ρ_j and/or s' is not too larger than s. ■

Proof of Proposition 6. In the multiplicative 2-firms model, the equilibrium quantity and price of firm j in Expressions (11) and (12), respectively, after having exchanged data with firm k become:

$$q'_j = (a - c)\rho_{j+k}/b(2 + s'),$$

and,

$$p'_j = a(2 - s') + c(2 + s' - s'^2)/(4 - s'^2),$$

respectively.

The firm j quantity after the data exchange is larger than before, which is given by Expression (11), i.e., $q'_j \geqslant q_j$, if the following inequality is satisfied:

$$\rho_{j+k} \geqslant \frac{(2 + s')\sqrt{\rho_j}(2\sqrt{\rho_j} - s\sqrt{\rho_k})}{(2 + s)(2 - s)}. \tag{16}$$

If inequality (16) is true for $\rho_j \geqslant \rho_k$, then it is also true for $\rho_j \leqslant \rho_k$ because the right-hand side becomes smaller. Then, for $\rho_j \geqslant \rho_k$, inequality (16) is satisfied if inequality (15) is also satisfied, i.e., if the right-hand side of inequality (15) is larger than the right-hand side of inequality (16), i.e. if:

$$\frac{(2 + s')^2(2\sqrt{\rho_j} - s\sqrt{\rho_k})^2}{(2 + s)^2(2 - s)^2} \geqslant \frac{(2 + s')\sqrt{\rho_j}(2\sqrt{\rho_j} - s\sqrt{\rho_k})}{(2 + s)(2 - s)}.$$

After some algebra, this inequality becomes $(2 + s')(2 - s\sqrt{\rho_k/\rho_j}) \geqslant (2 + s)(2 - s)$, which is more difficult to satisfy if $s' = s$, and in this case becomes $(2-s\sqrt{\rho_k/\rho_j})/(2-s) \geqslant 1$. Then, if $\rho_k/\rho_j \leqslant 1$ this inequality is always satisfied because the numerator is larger than the denominator. Consequently, if inequality (16) is satisfied for $\rho_k/\rho_j \leqslant 1$, it is also satisfied for $\rho_k/\rho_j \geqslant 1$. Symmetry is just a particular case. Therefore, if there exist incentives for data exchange, then both firms' equilibrium quantities increase.

Now consider the prices. The firm j price after the data exchange is lower than before, which is given by Expression (12), i.e., $p'_j \leqslant p_j$, if the following inequality is satisfied:

$$\frac{2(a+c)-(a-c)s'-cs'^2}{(4-s'^2)} \leqslant \frac{2(a+c)-(a-c)s\sqrt{\rho_k/\rho_j}-cs^2}{(4-s^2)}.$$

After some algebra this inequality reduces to $\sqrt{\rho_k/\rho_j} \leqslant (2s' + s^2)/(s(2+s'))$. The right-hand side increases in s' and decreases in s, therefore, it takes the minimum value $\sqrt{\rho_k/\rho_j} \leqslant 1$ at $s = s'$. Therefore, if $\rho_k/\rho_j \leqslant 1$, firm j always decreases prices after the data exchange, which includes the symmetric case, but not necessarily the case $\rho_k/\rho_j > 1$. ■

References

Acquisti, A., Taylor, C., Wagman, L., 2016. The economics of privacy. Journal of Economic Literature 54 (2), 442–92.

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

Bergemann, D., Bonatti, A., Gan, T., 2020. The economics of social data. Cowles Foundation Discussion Paper No. 2203R.

Bonatti, A., Cisternas, G., 2020. Consumer scores and price discrimination. The Review of Economic Studies 87 (2), 750–791.

Bos, I., Vermeulen, D., 2020. On the microfoundation of linear oligopoly demand. The BE Journal of Theoretical Economics.

Bounie, D., Dubus, A., Waelbroeck, P., 2021. Selling strategic information in digital competitive markets. The RAND Journal of Economics 52 (2), 283–313.

Chamley, C. P., 2004. Rational herds: Economic models of social learning. Cambridge University Press.

Clarke, R. N., 1983. Collusion and the incentives for information sharing. The Bell Journal of Economics, 383–394.

Clavorà Braulin, F., 2021. The effects of personal information on competition: Consumer privacy and partial price discrimination. ZEW-Centre for European Economic Research Discussion Paper (21-007).

Clavorà Braulin, F., Valletti, T., 2016. Selling customer information to competing firms. Economics Letters 149, 10–14.

De Corniere, A., Taylor, G., 2019. A model of biased intermediation. The RAND Journal of Economics 50 (4), 854–882.

Dixit, A., 1979. A model of duopoly suggesting a theory of entry barriers. The Bell Journal of Economics 10 (1), 20–32.

Fried, D., 1984. Incentives for information production and disclosure in a duopolistic environment. The Quarterly Journal of Economics 99 (2), 367–381.

Fudenberg, D., Villas-Boas, M., 2012. Price discrimination in the digital economy. In: The Oxford Handbook of the Digital Economy.

Gal-Or, E., 1985. Information sharing in oligopoly. Econometrica 53 (2), 329–343.

Gu, Y., Madio, L., Reggiani, C., 2019. Data brokers co-opetition. CESifo Working Paper No. 7523.

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

Li, L., 1985. Cournot oligopoly with information sharing. The RAND Journal of Economics 16 (4), 521–536.

Liu, Q., Serfes, K., 2006. Customer information sharing among rival firms. European Economic Review 50 (6), 1571–1600.

MacKay, D. J., 2003. Information theory, inference and learning algorithms. Cambridge University Press.

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

Raith, M., 1996. A general model of information sharing in oligopoly. Journal of Economic Theory 71 (1), 260–288.

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

Shapiro, C., 1986. Exchange of cost information in oligopoly. The Review of Economic Studies 53 (3), 433–446.

Shapley, L., Shubik, M., 1969. Price strategy oligopoly with product variation. Kyklos 22 (1), 30–44.

Shen, Q., Villas-Boas, M., 2018. Behavior-based advertising. Management Science 64 (5), 2047–2064.

Shiller, B. R., 2014. First-degree price discrimination using big data. Brandeis Univ., Department of Economics.

Shy, O., Stenbacka, R., 2016. Customer privacy and competition. Journal of Economics & Management Strategy 25 (3), 539–562.

Singh, N., Vives, X., et al., 1984. Price and quantity competition in a differentiated duopoly. The RAND Journal of Economics 15 (4), 546–554.

Taylor, C., Wagman, L., 2014. Consumer privacy in oligopolistic markets: Winners, losers, and welfare. International Journal of Industrial Organization 34, 80–84.

Vives, X., 1984. Duopoly information equilibrium: Cournot and bertrand. Journal of Economic Theory 34 (1), 71–94.

Vives, X., 1990. Trade association disclosure rules, incentives to share information, and welfare. The RAND Journal of Economics 21 (3).

Vives, X., 1999. Oligopoly pricing: old ideas and new tools. MIT press.

Westenbroek, T., Dong, R., Ratliff, L. J., Sastry, S. S., 2020. Competitive statistical estimation with strategic data sources. IEEE Transactions on Automatic Control 65 (4), 1537–1551.