The Economics of Data

David Easley Shiyang Huang Liyan Yang Zhuo Zhong*

March 18, 2018

Abstract

We analyze the economic consequences of selling consumer data to oligopoly producers. Without data sales, producers keep secret their private consumer data, leading to efficiency loss and in some cases, to a prisoners dilemma for producers. In the presence of an independent data vendor who maximizes its own profits with smart contracts, data sales causes producers to effectively share their consumer data in equilibrium, thereby improving total surplus. This setting is consistent with a situation in which data is owned by consumers and analyzing such a setting provides a way to quantify the economic value of consumer data. When data is owned by producers, a data vendor á la a trade association is likely to maximize the total profits of producers, and its presence can address the prisoners' dilemma for producers. Our analysis provides implications for the debates about data ownership and privacy.

JEL Classification: D61; G14; M41

Keywords: Fintech, Industry organization, Welfare, Prisoners dilemma, Data ownership

*David Easley (dae3@cornell.edu) is at Cornell University; Shiyang Huang (huangsy@hku.hk) is at University of Hong Kong; Liyan Yang (liyan.yang@rotman.utoronto.ca) is at University of Toronto; Zhuo Zhong (zhuo.zhong@unimelb.edu.au) is at University of Melbourne.

1 Introduction

In the last two decades, the technology improvement in data storage and analysis has made data an important input for modern business. With data advantage, tech-companies like Amazon and Google surpass the traditional business giants, such as General Electronic or Mobil. Many consider data as the "new oil" that future business will thrive on. However, there are many concerns accompanied with the new data economy, one of which is who should own the data. The current debate on data ownership centers around issues like data privacy or fairness (Acquisti, 2015). An important but less explored question is how the data ownership affects the economic efficiency. The answer to this question improves the understanding of the springing data economy.

Based on the classical setting in the literature of privately informed duopoly competition (Gal-Or, 1985; Vives, 1984), we propose a model to study how data sales and data ownership affect the economic outcome. Specifically, we consider a two-period model with two generations of consumers. There are two firms, A and B, and one data vendor. Each firm produces in two markets, her "local" markets and the "global" market. In local markets, the firm is a monopoly. Whereas in the global market, firms compete on quantity, which is a Cournot duopoly competition. Consumers have correlated uncertain preferences. Thus, firms could use the transaction data from the early-generation consumers to forecast the demand of the later-generation consumers. Firms use the data from both the local and global market to make better production policies. Because the local market's data is not accessed to the rival firm, the local market data becomes the proprietary information for firms. Different from the local market, data in the global market are public information. The data vendor collects transaction data from firms' local markets. As a result, firms could purchase the rival firm's local market data in making production decisions.

Our model is able to capture the emerging feature of the modern business practice. (1) The duopoly setting captures the recent concentration trend of U.S. product markets. Grullon, Larkin, and Michaely (2017) show that the Herfindahl-Hirschman index (a measure of concentration) has systematically increased in over 75% of U.S. industries for the last two decades. The market share of the largest firm in an industry has grown significantly. (2) Firms utilize information to adjust for production plans. For example, the retail giant Walmart uses past data to predict the "rush hour" of shopping and assign associates at the counters accordingly.[1] (3). The emergence of data sales from data centers, such as IBM and Oracle. In 2015, IBM spent over 2 billion to acquire Weather Company. A year later, utilizing data from Weather company, IBM released a hyperlocal weather forecast — at a 0.2-mile to 1.2-mile resolution — to provide enterprise clients with short-term customized forecasts.[2]

We have several main results. First, we show the presence of the data vendor can change firms' behaviors, and consequently moves the equilibrium to achieve a higher total surplus. To illustrate this point, we first consider a benchmark economy without the data vendor. Consistent with the literature (Gal-Or, 1985, Vives, 1984), we find no firms disclose their proprietary information. Both firms want to retain their competitive advantage on information. Strikingly, no information sharing prevails even if full disclosure could generate higher profits for both firms. In essence, firms are stuck in the "prisoner's dilemma" equilibrium. However, we show that the emergence of the data vendor solves the prisoner's dilemma and increases the total surplus. With the data vendor, each firm can purchase data on the rival's proprietary information. In equilibrium, firms purchase all of her rival's proprietary information. This leads to the full information sharing equilibrium. Consumer surplus and total surplus increase in this equilibrium. The incremental surplus comes from that the purchased data improves firms' planning on their production. Under certain conditions, the equilibrium with data vendor can be a Pareto improvement to all participants (e.g., firms) in the economy.

The second result is about how data ownership affects data sales and economic outcomes.

[1] See "5 ways Walmart uses big data to help customers", https://blog.walmart.com/innovation/20170807/5-ways-walmart-uses-big-data-to-help-customers

[2] For more detailed reports, see https://www.marketwatch.com/story/ibm-finally-reveals-why-it-bought-the-weather-company-2016-06-15

We discuss two settings of data ownership. In the first setting, data is sold via an independent profit-maximizing data vendor in which data is originally owned by consumers. In the second setting, data is owned by firms and firms form a data vendor to maximize their total profits in selling data. We find that data ownership matters for data sales and the total surplus. When the data vendor is an independent profit-maximizing institute, it always sells all of the rival's information to each firm. As a result, the optimal total surplus will always be achieved. However, this is not always the case when the data vendor is owned by firms. When the size of the local market is small, the data vendor owned by firms will not sell any information to firms and the total surplus is the same as in the no information sharing equilibrium. When the size of the local market is large, the data vendor always sells all of the rival's information to each firm, which achieves the same total surplus as the first setting.

Our model bridges the literature on privately informed duopoly competition (see Vives, 1984; Gal-Or, 1985; Gal-Or, 1986; Darrough, 1993; Raith, 1996; and Bagnoli and Watts, 2015) with the literature on information sales (see Admati and Pfleiderer, 1986, Admati and Pfleiderer, 1987, Admati and Pfleiderer, 1988). Section 2 illustrates our model setting. In Section 3, we analyze an economy without the data vendor. This serves as the benchmark economy. We then introduce the data vendor in Section 4. In Section 5, we internalize the data vendor through analyzing the data ownership. We conclude with Section 6.

2 The model

We build our analysis in a duopoly setting in which two firms face random demand and compete on quantity. Each firm possesses private information regarding the random demand. This setting has been extensively studied in the industrial-organization literature (e.g., Vives, 1984; Gal-Or, 1985; Gal-Or, 1986; Raith, 1996; and Bagnoli and Watts, 2015). We extend it by introducing a data vendor from whom the two firms can purchase information (or data). We show that the presence of a data vendor can change firms' behaviors, which moves the equilibrium to a better allocation that achieves a higher total surplus. To convey our idea most parsomoniously, we do not explore how such a data vendor emerges for now and we delegate such an exploration to Section 5.

The economy lasts for two periods, $t = 0, 1$. Both firms, labeled as A and B, last for two periods, and in each period, they produce goods to maximize expected profits. There are two generations of consumers, each living for one period and consuming goods produced by firms. Within each generation, there are three types of consumers: A-type, B-type, and AB-type. An A-type consumer buys goods only from firm A; a B-type consumer buys goods only from firm B; and an AB-type consumer buys either from firm A or from firm B. Consumers' preference is random and is driven by both a common component and an idiosyncratic component. The common component persists through two generations of consumers and hence, the date-0 consumer data helps firms to forecast the date-1 demand.

In our setting, the date-0 product market equilibrium serves to generate the consumer data that can be potentially sold and firms' date-1 information structure. In period 0, each firm decides its production decisions and sells products to their respective consumers, giving rise to the market-clearing prices. The price for AB-type consumers are observed by both firms. By contrast, the equilibrium prices for A-type consumers and for B-type consumers constitute the private information of firm A and firm B, respectively. Since date-0 consumer data contains useful information for future date-1 demand, both firms have incentives to observe the private data owned by their rivals. In our setting, the data vendor satisfies this demand for data; the data vendor has access to the date-0 consumer data and sells it to firms. In the remaining of this section, we describe in greater detail the product markets and the information market, and then define the equilibrium concept.

2.1 Consumption, supply, and product markets

2.1.1 Product demand from consumers

The consumption decisions of consumers generate to the demand for firms' products. As mentioned above, three types of consumers—A-type, B-type, and AB-type—exist in each period. We interpret A-type consumers and B-type consumers respectively as each firm's "local" market consumers, and AB-type consumers as firms' "global" market consumers. Firms behave as monopolists in their local markets and compete in the global market. There are M local markets for each firm, where M is a positive integer. In each local market, there is one representative consumer. In the global market, there exist $N \geq 1$ representative AB-type consumers (and thus the global market is relatively larger than a typical local market).

This local/global setting is consistent with the granular structure of many industries. For instance, in developed economies, small towns tend to have their own favorite grocery stores, while in big cities, many major grocery stores coexist and compete. Another interpretation of our setting is to perceive consumers arriving sequentially to buy products from firms (e.g., in the context of online shopping). Firms face only one consumer at a time; those loyal consumers who only want to buy goods from a particular firm constitute that firm's local markets, while those consumers who are indifferent to brands constitute the global market. We are agonistic about interpretations and label local markets as X-markets and the global market as the Y-market.

We denote by U_A, U_B, and U_{AB} the utilities for type-A, type-B, and type-AB consumers, respectively. Following the literature (e.g., Singh and Vives, 1984), consumers derive utility from consuming the products produced by firms according to the following quasi-linear forms:

$$U_A(x^t_{A,i}) = \tilde{s}^t_{A,i} x^t_{A,i} - \frac{(x^t_{A,i})^2}{2} - p^t_{A,i} x^t_{A,i}, \;\; i = 1, 2, ..., M; \tag{1}$$

$$U_B(x^t_{B,j}) = \tilde{s}^t_{B,j} x^t_{B,j} - \frac{(x^t_{B,j})^2}{2} - p^t_{B,j} x^t_{B,j}, \;\; j = 1, 2, ..., M; \tag{2}$$

$$U_{AB}(y^t_k) = \tilde{s}^t_{AB,k} y^t_k - \frac{(y^t_k)^2}{2} - p^t_y y^t_k, \;\; k = 1, 2, ..., N. \tag{3}$$

Here, variables $x_{A,i}^t$ and $x_{B,j}^t$ are the the quantities consumed by the generation-t consumers in firm A's ith X-market and in firm B's jth X-market, respectively. Variables $p_{A,i}^t$ and $p_{B,j}^t$ represent the product prices in these local markets. Similarly, variable y_k^t represents the demand of a typical generation-t consumer k in the global Y-market, and p_y^t is the product price at the Y-market in period t.

Variables $\tilde{s}_{A,i}^t$, $\tilde{s}_{B,j}^t$, and $\tilde{s}_{AB,k}^t$ capture preference shocks. Preference shocks contain two random components, a time invariant common component $\tilde{\theta}$ and an idiosyncratic component $\tilde{\varepsilon}$:

$$\tilde{s}_{A,i}^t = \tilde{\theta} + \tilde{\varepsilon}_{A,i}^t, i = 1, 2, ..., M,$$

$$\tilde{s}_{B,j}^t = \tilde{\theta} + \tilde{\varepsilon}_{B,j}^t, j = 1, 2, ..., M,$$

$$\tilde{s}_{AB,k}^t = \tilde{\theta} + \tilde{\varepsilon}_{AB,k}^t, k = 1, 2, ..., N,$$

where $\tilde{\theta} \sim \mathcal{N}\left(0, \tau_\theta^{-1}\right)$, $\tilde{\varepsilon}_{A,i}^t \sim \mathcal{N}\left(0, \tau_\varepsilon^{-1}\right)$, $\tilde{\varepsilon}_{B,j}^t \sim \mathcal{N}\left(0, \tau_\varepsilon^{-1}\right)$, and $\tilde{\varepsilon}_{AB,k}^t \sim \mathcal{N}\left(0, \tau_\varepsilon^{-1}\right)$ with $\tau_\theta > 0$ and $\tau_\varepsilon > 0$. We assume that $\left\{\tilde{\theta}, \left\{\tilde{\varepsilon}_{A,i}^t\right\}_i, \left\{\tilde{\varepsilon}_{B,j}^t\right\}_j, \left\{\tilde{\varepsilon}_{AB,k}^t\right\}_k\right\}$ are mutually independent. Consumers know their own preference shocks when making purchase decisions.

We have normalized the mean of preference shocks as zero. This normalization does not affect our results. We have also assumed that the common component is the same across all consumers. Our mechanism still works as long as there is some correlation among consumers' preference shocks. In addition, in preference specification (3), we assume that the products of both firms are perfect substitutes for AB-type consumers. This assumption is made for the sake of simplicity. The results still go through if the products of both firms are not perfect substitutes in the Y-market.

Each consumer maximizes her preference taking the product prices as given. Solving consumers' utility-maximization problems leads to the following inverse demand functions in the X-markets and Y-market, respectively:

$$p_{A,i}^t = \tilde{s}_{A,i}^t - x_{A,i}^t, i = 1, 2, ..., M; \quad (4)$$

$$p_{B,j}^t = \tilde{s}_{B,j}^t - x_{B,j}^t, j = 1, 2, ..., M; \quad (5)$$

$$p_y^t = \frac{1}{N}\left[\sum\nolimits_{k=1}^{N} \tilde{s}_{AB,k}^t - \sum\nolimits_{k=1}^{N} y_k^t\right]. \quad (6)$$

2.1.2 Product supply from firms

Firms live for two periods. In each period, firms maximize the expected profits conditional on their information. These profit-maximization decisions lead to the product supply in the product markets.

Date-0 product markets When making date-0 production decisions, firms have not received any information yet, except their priors about the model structure. Thus, firms choose production quantities to maximize unconditional expected profits taking as given the demand functions from consumers and the production quantities of their rivals. Firm A's optimal production quantities $\left(X_{A,1}^0, ..., X_{A,M}^0, Y_A^0\right)$ are determined by

$$\max_{\{X_{A,i}^0\}_{i=1}^M, Y_A^0} \mathbb{E}[\underbrace{\sum\nolimits_{i=1}^M p_{A,i}^0 X_{A,i}^0}_{X\text{-market}} + \underbrace{p_y^0 Y_A^0}_{Y\text{-market}}], \quad (7)$$

where $p_{A,i}^0$ and p_y^0 are given respectively by demand functions (4) and (6) with $t = 0$. Firm B's decisions can be characterized similarly by changing notations.

In an X-market, the corresponding firm behaves as a monopolist and thus the two firms make decisions independently. In the Y-market, firms' optimal productions form a Nash equilibrium. When making production decisions, each firm needs to take into account the other firm's production and the market-clearing condition (i.e., $Y_A^0 + Y_B^0 = \sum_{k=1}^N y_k^0$). The equilibrium computation is standard and thus omitted. We summarize the result in the following lemma.

Lemma 1 (Date-0 product market equilibrium)

In the data-0 product markets, the equilibrium prices $(\{p^{0*}_{A,i}\}_{i=1}^{M}, \{p^{0*}_{B,j}\}_{j=1}^{M}, p^{0*}_{y})$*, production quantities* $(\{X^{0*}_{A,i}\}_{i=1}^{M}, Y^{0*}_{A}, \{X^{0*}_{B,j}\}_{j=1}^{M}, Y^{0*}_{B})$*, and expected profits* ($\mathbb{E}\Pi^{0*}_{A}$ *and* $\mathbb{E}\Pi^{0*}_{B}$) *are:*

$$\begin{aligned}
p^{0*}_{A,i} &= \tilde{s}^{0}_{A,i}, X^{0*}_{A,i} = 0, \text{ for } i = 1, ..., M; \\
p^{0*}_{B,j} &= \tilde{s}^{0}_{B,j}, X^{0*}_{B,j} = 0, \text{ for } j = 1, ..., M; \\
p^{0*}_{y} &= \frac{1}{N} \sum\nolimits_{k=1}^{N} \tilde{s}^{0}_{AB,k}, Y^{0*}_{A} = 0, Y^{0*}_{B} = 0; \\
\mathbb{E}\Pi^{0*}_{A} &= 0 \text{ and } \mathbb{E}\Pi^{0*}_{B} = 0.
\end{aligned}$$

The equilibrium prices in the date-0 X-markets reveal consumers' preference shocks and hence this price data is useful for firms to make forecast about next period demand. The production quantities and profits are equal to 0 in equilibrium. This is due to the fact that we have normalized both the mean of preference shocks and the production cost at zero. If we relax this normalization, then both quantities and profits become non zero in equilibrium, and under this alternative specification, prices and sales (i.e., prices multiplied by quantities) convey the same information. In our current simplified setting, when using the wording "consumer data," we refer to the date-0 price data, $\{p^{0*}_{A,i}\}_{i=1}^{M}$ and $\{p^{0*}_{B,j}\}_{j=1}^{M}$. To ease expressions, we label these price vectors as follows: $\mathbf{P}^{0}_{A} \equiv \{p^{0*}_{A,i}\}_{i=1}^{M}$ and $\mathbf{P}^{0}_{B} \equiv \{p^{0*}_{B,j}\}_{j=1}^{M}$.

Date-1 product markets After the date-0 product markets clear, the price data is formed. Both firms observe the equilibrium Y-market price, p^{0*}_{y}. Firm A privately observes all of its X-market prices $\mathbf{P}^{0}_{A}$. Similarly, firm B privately observes its own X-market prices $\mathbf{P}^{0}_{B}$. This forms the basis of the firms' starting information structure in period 1. As we mention before, all the price data is also available to a data vendor, who in turn sells the data to firms. We will discuss the data market in the next subsection in detail. The general idea is that firm A buys price data about firm B's date-0 X-market, and vice versa.

Let $\mathcal{F}_{A} \equiv \{p^{0*}_{y}, \mathbf{I}_{A}, \mathbf{P}^{0}_{A}\}$ denote firm A's information set, where $\mathbf{I}_{A}$ indicates the vector of price data purchased by firm A. Firm A's date-1 production quantities $\left(\left\{X_{A,i}^1\right\}_i, Y_A^1\right)$ are determined by

$$\max_{\left\{X_{A,i}^1\right\}_{i=1}^M, Y_A^1} \mathbb{E}[\underbrace{\sum_{i=1}^M p_{A,i}^1 X_{A,i}^1}_{X\text{-market}} + \underbrace{p_y^1 Y_A^1}_{Y\text{-market}} |\mathcal{F}_A], \tag{8}$$

where the prices $p_{A,i}^1$ and p_y^1 are given by inverse demand functions (4) and (6) with $t = 1$, respectively. We can write down a similar profit-maximization problem for Firm B.

Similar to date 0, firms behave as monopolists in their respective X-markets and make production decisions separately. Now their optimal productions are no longer constant, but instead depend on their information sets. For instance, the optimal production policies for firm A in the ith X-markets is $X_{A,i}^{1*} = X_{A,i}^1(\mathcal{F}_A)$. In the Y-market, we need to consider the strategic interactions between the two firms, and their optimal production decisions form a Bayesian Nash equilibrium. We delegate the derivation of the date-1 product market equilibrium to Section 4, and now turn to the description of the data market which determines firms' information sets $\mathcal{F}_A$ and $\mathcal{F}_B$.

2.2 Data vendor and data market

2.2.1 The vendor's problem

In the data market, a data vendor sells the collected date-0 consumer data to firms who in turn use the purchased data to improve their date-1 production decisions. In the baseline model described by this section, we follow the literature on information sales in financial markets (e.g., Admati and Pfleiderer (1986)) and assume that the data vendor maximizes its own profits and behaves as a monopolist in the data market. In Section 5, we argue that this assumption is consistent with an equilibrium outcome in a setting in which the date-0 consumers own the data. Nonetheless, we note that our central message that information sales can improve social welfare does not depend on this assumption (see Section 5).

Data transactions are completed at the beginning of date 0. We will discuss the details of the transaction games shortly in Section 2.2.2. The outcome of these transactions is that firm A pays cost C_A to buy m_A amount of data and firm B pays cost C_B to buy m_B amount of data. As we discussed before, the consumer data is in the form of X-market prices and thus, the amount of consumer data refers to the number of X-market prices. We follow the literature (e.g., Gal-Or, 1985; Li, McKelvey, and Page, 1987; Vives, 1988; and Hwang, 1993) and assume that after firms make their data purchase decisions, the purchase amount (m_A, m_B) becomes common knowledge and is observable to both firms before they make their date-1 production decisions (of course, the specific values of the m prices are only observable to the firm who has purchased the data). In the terminology of Hauk and Hurkens (2001), firms do not engage in "secret information acquisition."[3]

Each firm only wants to buy the X-market prices of its rival. These prices are originally the private information of each firm who collects this information from its own local X-market transactions. Thus, with data purchase, firms effectively observe part or all of their rivals' private data. We label this resulting data exchange outcome as "data allocation."

Definition 1 (Data allocation)

A data allocation, denoted by (m_A, m_B) with $m_A \in \{0, 1, ..., M\}$ and $m_B \in \{0, 1, ..., M\}$, refers to a situation in which, when making their date-1 production decisions, firm A observes m_A date-0 X-market prices $p^{0*}_{B,j}$ of firm B, and firm B observes m_B date-0 X-market prices $p^{0*}_{A,i}$ of firm A.

We follow Admati and Pfleiderer (1986) and assume that the data vendor can implement any admissible data allocation through information sales. This is natural given that the data vendor is a monopolist in the data market. In Section 2.2.2, we will describe how the data vendor achieves this implementation by offering right contracts. Given data allocation (m_A, m_B), we use $\mathbb{E}\Pi^1_A(m_A, m_B)$ to denote firm A's expected profit resulting from the date-1 product market equilibrium. Specifically, we insert the optimal production policies into the objective function of (8) and take unconditional expectations to compute $\mathbb{E}\Pi^1_A(m_A, m_B)$. If

[3] If firms engage in "secret" information purchase (i.e., (m_A, m_B) is not observable to firms when making production decisions), then a firm will take its rival's production policies as given when considering the information value through a deviation analysis. We have shown that our results are robust under this alternative assumption.

firm A does not buy any data from the data vendor, then its expected profit is $\mathbb{E}\Pi_A^1(0, m_B)$. Thus, firm A's willingness to pay for an amount m_A of data given that its rival has purchased an amount m_B of data is

$$C_A(m_A, m_B) = \mathbb{E}\Pi_A^1(m_A, m_B) - \mathbb{E}\Pi_A^1(0, m_B). \tag{9}$$

Since the monopolist extracts all surplus, $C_A(m_A, m_B)$ constitutes its profit from selling data to firm A. We can define firm B's willingness to pay similarly and label it as $C_B(m_A, m_B)$.

A profit-maximizing data vendor's problem is to choose a data allocation to maximize its own profits as follows:

$$\max_{(m_A, m_B) \in \{0,1,\ldots,M\}^2} \left[C_A(m_A, m_B) + C_B(m_A, m_B)\right]. \tag{10}$$

Equations (9) and (10) share the same spirit as Admati and Pfleiderer (1986) who study how a monopolistic data vendor sells information in financial markets with different precision levels. In their model, the data price that the seller can charge is computed as the difference between the certainty equivalent of a trader who is equipped with the information and the certainty equivalent of a trader who is uninformed (their equation (3.1)). This corresponds to our equation (9). Similar to our equation (10), the data vendor in Admati and Pfleiderer (1986) extracts all surplus by choosing a distribution of information precision levels (see their equation (3.2)).

2.2.2 Microstructure of data sales

We now describes two mechanisms through which the data vendor implements a particular data allocation (m_A, m_B). In the first mechanism, the vendor simultaneously offers two take-it-or-leave-it contracts to both firms, while in the second mechanism, the vendor offers contracts sequentially. Both games feature a unique equilibrium in terms of data allocation, and the equilibrium data prices lead to equation (9). In the latter sections, we are agnostic to these two implementation mechanisms, and just focus on the vendor's maximization problem (10).

Simultaneous offering with contingent contracts

The data prices in the contracts offered by the data vendor are contingent on data allocations. Putting it in context, suppose that data is owned by consumers and the date-0 local consumers form the data vendor (see Section 5.1 for more discussions). For illustrative purpose, we assume that the data vendor wants to implement data allocation $(m_A, m_B) = (100, 100)$. Then, the data vendor may present the following two offers to firms, for example:

Contract A (on A-type consumer data): "*If you purchase 1000 data points about A-type consumers and no one else buys any data, then you pay $30; and if you purchase 1000 data points about A-type consumers and someone else also buys some data, then you pay $40.*"

Contract B (on B-type consumer data): "*If you purchase 1000 data points about B-type consumers and no one else buys any data, then you pay $30; and if you purchase 1000 data points about B-type consumers and someone else also buys some data, then you pay $40.*"

Given that only firm A is interested in contract B and only firm B is interested in contract A, the above two contracts are effectively the following: "*If the data allocation is* $(m_A, 0) = (1000, 0)$, *then firm A pays* $t_A = 30$*; and if the data allocation is* $(m_A, m_B) = (1000, 1000)$, *then firm A pays* $t_A = 40$;" and "*If the data allocation is* $(0, m_B) = (0, 1000)$, *then firm B pays* $t_B = 30$*; and if the data allocation is* $(m_A, m_B) = (1000, 1000)$, *then firm B pays* $t_B = 40$." The contents of both contracts are observable to both firms.

These contracts correspond to the concept of "smart contracts" in the context of FinTech. Smart contracts are computer programs that execute "if this happens then do that," run and verified by many computers to ensure trustworthiness in a blockchain environment.[4] If these contingent contracts are available, then the data vendor can use them to modify the payoff

[4]See Cong and He (2018) for more discussions on smart contracts. An informal discussion on this concept can be found at: https://bitsonblocks.net/2016/02/01/a-gentle-introduction-to-smart-contracts/.

matrix of firms at the information purchase stage, such that the unique Nash equilibrium leads to data allocation (m_A, m_B) (see Section 4 for more details). There are multiple contingent contracts that implement (m_A, m_B), but for all of these contracts, the vendor's ultimate profits are given by $C_A(m_A, m_B) + C_B(m_A, m_B)$.

Sequential offering with simple contracts

In the absence of "smart contracts," we can consider a four-stage game in which the vendor offers *simple* contracts sequentially. The game's extensive form is drawn in Figure 1. In the first stage, the data vendor contacts firm A and offers a take-it-or-leave-it contract which states that "firm A can pay a cost t_A to buy an amount m_A of data." In the game, the cost t_A is the vendor's choice variable with an action space $\mathbb{R}_+$, and the data amount m_A is a fixed parameter that is exogenous to the game. Receiving the offer, firm A decides to accept or reject the offer in the second stage. If firm A accepts the offer, then it will pay t_A and purchase m_A amount of data, and if it rejects, it will not buy data.[5] In the third stage, observing firm A's choice, the data vendor then offers another contract to firm B which says that "firm B can pay a cost t_B to buy an amount m_B of data." In the last stage, firm B decides to take or reject the offer.

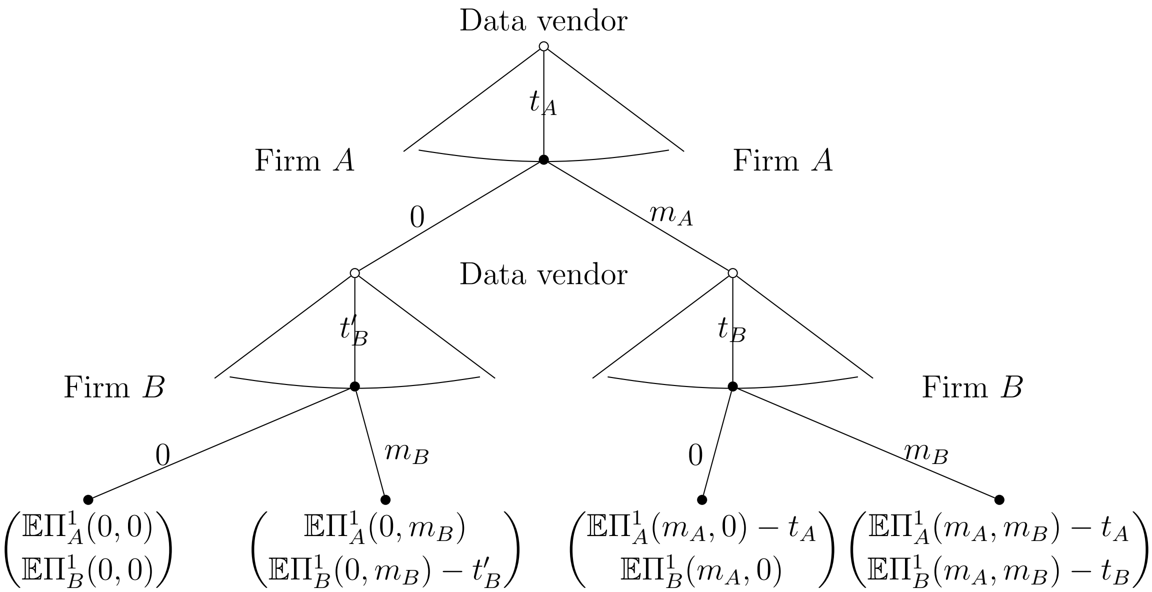
Lemma 2 (Sequential contract offering)

In the four-stage game described above, there is a unique sequential Nash equilibrium in which the data vendor sets data price $t_A^* = C_A(m_A, m_B)$ in the first stage, firm A accepts the offer in the second stage, then the vendor sets data price $t_B^* = C_B(m_A, m_B)$ in the third stage, and firm B also accepts the offer in the last stage.

Intuitively, the data vendor would like to sell its data to both firms and thus, it will choose the right data prices such that both firms would like to purchase the data. More formally, we solve the model by backward induction. In the last stage, at each node, firm B will buy data if and only if the data price t_B is sufficiently low. In the third stage, anticipating the

[5] When a firm is indifferent between buying and not buying data, we assume that the firm will always choose to buy data. This makes sense because if not, the data vendor can always slightly lower the data price to get a positive profit.

Figure 1: Sequential contract offering



optimal response of firm B in the last stage, the data vendor will charge the price just to the level at which firm B does not want to switch from buying, so that firm B always buys data in any possible equilibrium path. Back to stage 2, anticipating that firm B will always buy data, firm A will buy data if and only if the data price t_A is no larger than $C_A(m_A, m_B)$. To achieve the maximum profit, the data vendor sets data price at $t_A^* = C_A(m_A, m_B)$. Thus, the four-stage game implements the data allocation (m_A, m_B).

In reality, the data vendor could implement sequential sales in several ways. First, the data vendor as a monopolist has the full discretion over the timing of sales, so that it can literally do sequential data sales by contacting firms one by one. This practice is popular in many real over-the-counter (OTC) markets. Second, the data vendor can allow one firm to place a pre-order for data and then set the late-stage data price conditioning on the pre-order outcome. Pre-ordering is customary in book and video game industries. The development of Initial Coin Offerings (ICOs), in particular the utility-token sales, can facilitate pre-ordering by directly providing token buyers with future access to the token seller's products.[6]

[6] For more discussion regarding ICOs, see Li and Mann (2018).

2.3 Timeline and equilibrium concept

The timeline of the economy is described by Panel A of Figure 2. At the beginning of date 0, the data vendor designs contracts to sell consumer data to firms. Firms then make production decisions and the product markets clear. The date-0 equilibrium product prices generate the consumer data, which is observable to the data vendor at the end of date 0. At the beginning of date 1, the data vendor delivers the promised consumer data to firms. Then, firms make optimal production decisions based on their information sets. Finally, the product markets clear, consumers purchase goods and consume, and firms realize their profits.

Panel B of Figure 2 describes the flow of products and information. Specifically, the data vendor's information stems from the date-0 product prices in firms' local X-markets. Then, the data vendor provides firm A, at a cost C_A, with a certain number of firm B's date-0 X-market prices. Firm A in turn employs this information, together with its knowledge about its own X-market prices and the Y-market price in the previous period, to make production decisions. There is a similar information flow between the data vendor and firm B. Firms A and B behave as monopolists in their own X-markets, but they compete in the global Y-market.

Recall that Lemma 1 has characterized the date-0 product market equilibrium and that Lemma 2 has provided a sequential equilibrium to microfound the data prices. Thus, our equilibrium definition focuses only on the data vendor's profit optimization problem on date 0 and the product market equilibrium on date 1. We adopt an equilibrium concept in the sense of a sequential Nash equilibrium.

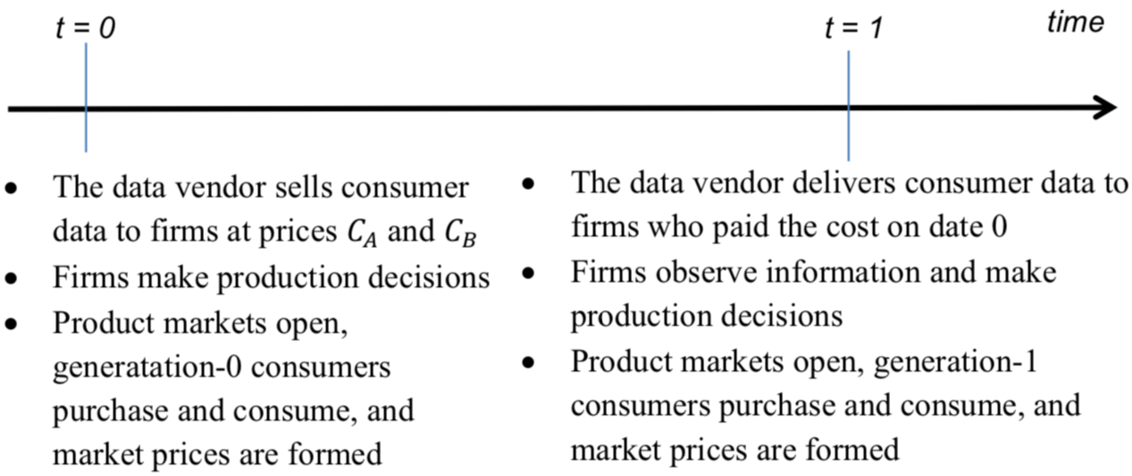
Definition 2 (Equilibrium)

An equilibrium consists of a date-0 data allocation (m_A^*, m_B^*) *and date-1 production policies,* $(\{X_{A,i}^1(\mathcal{F}_A)\}_{i=1}^M, Y_A^1(\mathcal{F}_A))$ *and* $(\{X_{B,j}^1(\mathcal{F}_B)\}_{j=1}^M, Y_B^1(\mathcal{F}_B))$*, such that:*

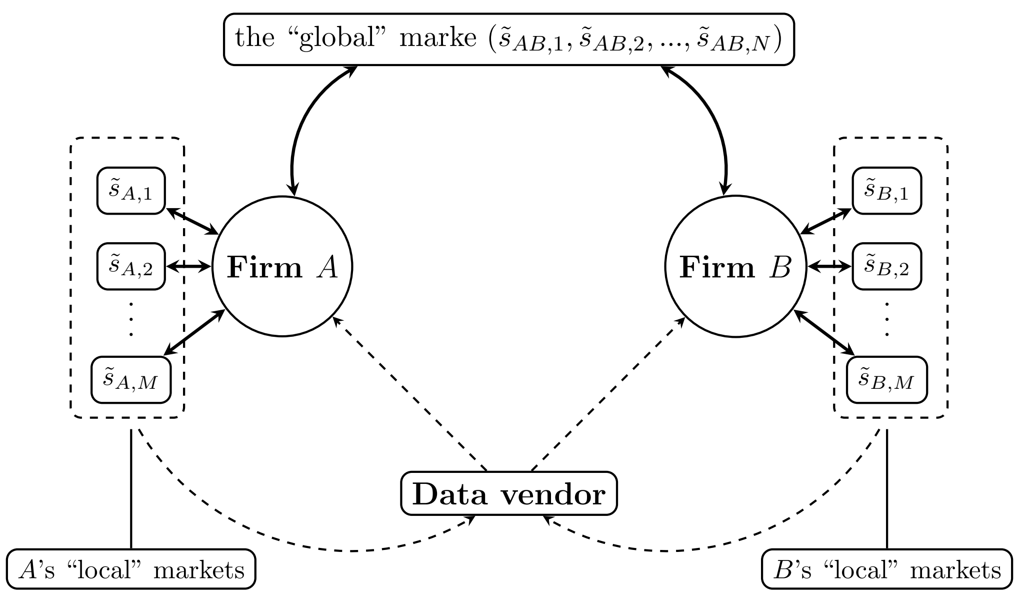
(a) Given the equilibrium amount (m_A^*, m_B^*) *of purchased data, the date-1 policies* $X_{A,i}^1(\mathcal{F}_A)$ *and* $X_{B,j}^1(\mathcal{F}_B)$ *maximize the conditional profits in firm A's ith X-market and in firm B's jth X-market, respectively; and* $(Y_A^1(\mathcal{F}_A), Y_B^1(\mathcal{F}_B))$ *form a Bayesian Nash equilibrium*

Figure 2: Model overview

Panel A: Timeline



Panel B: The market structure



Panel A of Figure 2 shows the timeline of the model. Panel B illustrates the market structure. In Panel B, the solid line indicates real goods transactions and the dashed line indicates information collection and sales.

in the Y-market.

(b) The equilibrium amount (m_A^*, m_B^*) of sold data is determined by (10), and the data prices (C_A^*, C_B^*) are set accordingly as $C_A^* = C_A(m_A^*, m_B^*)$ and $C_B^* = C_B(m_A^*, m_B^*)$.

We solve the equilibrium by backward induction. That is, we compute the date-1 product market equilibrium for any given (m_A, m_B). This allows us to figure out the expression of firms' profits, $\mathbb{E}\Pi_A^1(m_A, m_B)$ and $\mathbb{E}\Pi_B^1(m_A, m_B)$. We then solve the data vendor's profit-maximization problem (10), which leads to the equilibrium data allocation (m_A^*, m_B^*). To set the stage for our analysis, in the next section we first examine a benchmark economy without a data vendor.

3 What happens without a data vendor?

In this section, we first analyze a benchmark economy without a data vendor. We then provide some background discussions on how the literature has strived to improve on the equilibrium outcome in the benchmark economy, namely, by considering information sharing among firms. However, free information sharing is not viable or costly in the case of Cournot competition and demand uncertainty. By contrast, our paper shows that information sales—both in the benchmark model of Section 2 and in the variant model of Section 5—can achieve the desired welfare improvement.

3.1 Product market equilibrium in the benchmark economy

Without a data vendor, the date-1 Y-market in our economy degenerates to the classical duopoly setting with privately informed Cournot firms (e.g., Gal-Or, 1985; Darrough, 1993). The information structure of firms is endogenously determined by the date-0 product market equilibrium. Specifically, the date-0 Y-market price p_y^{0*} serves as the public information shared by both firms. Firm A's date-0 X-market prices $\mathbf{P}_A^0 \equiv \{p_{A,i}^{0*}\}_{i=1}^M$ are firm A's private information, while firm B's date-0 X-market prices $\mathbf{P}_B^0 \equiv \{p_{B,j}^{0*}\}_{j=1}^{M}$ are firm B's private information. By Lemma 1, the date-0 market prices reveal preference shocks of date-0 consumers. Formally, firm A's information set and firm B's information set are respectively:

$$\mathcal{F}_A = \{p_y^{0*}, \mathbf{P}_A^0\} = \left\{ \frac{\sum_{k=1}^{N} \tilde{s}_{AB,k}^0}{N}, \tilde{s}_{A,1}^0, ..., \tilde{s}_{A,M}^0 \right\},$$
$$\mathcal{F}_B = \{p_y^{0*}, \mathbf{P}_B^0\} = \left\{ \frac{\sum_{k=1}^{N} \tilde{s}_{AB,k}^0}{N}, \tilde{s}_{B,1}^0, ..., \tilde{s}_{B,M}^0 \right\}.$$

We follow the literature (e.g., Gal-Or, 1985 and Darrough, 1993) and consider the date-1 production policies that are linear in firms' information variables. Given that the date-0 X-market prices have the same precision level in predicting the persistent component $\tilde{\theta}$ in the future demand, it is intuitive to specify that the coefficients on these prices are the same. We therefore conjecture the following date-1 production policies for firms A and B:

$$X_{A,i}^1 = \Phi_{A_0}^X + \Phi_{A_1}^X (P_A^0 - \mu), i = 1, ..., M, \tag{11}$$

$$X_{B,j}^1 = \Phi_{B_0}^X + \Phi_{B_1}^X (P_B^0 - \mu), j = 1, ..., M, \tag{12}$$

$$Y_A^1 = \Phi_{A_0}^Y + \Phi_{A_1}^Y (P_A^0 - \mu), \tag{13}$$

$$Y_B^1 = \Phi_{B_0}^Y + \Phi_{B_1}^Y (P_B^0 - \mu), \tag{14}$$

where

$$P_A^0 \equiv \frac{1}{M} \sum\nolimits_{i=1}^{M} p_{A,i}^{0*} = \frac{1}{M} \sum\nolimits_{i=1}^{M} \tilde{s}_{A,i}^0, \tag{15}$$
$$P_B^0 \equiv \frac{1}{M} \sum\nolimits_{j=1}^{M} p_{B,j}^{0*} = \frac{1}{M} \sum\nolimits_{j=1}^{M} \tilde{s}_{B,j}^0, \tag{16}$$

are two price indices of date-0 X-market data, and

$$\mu \equiv \mathbb{E}(\tilde{\theta} | p_y^{0*}) = \frac{N\tau_\varepsilon}{N\tau_\varepsilon + \tau_\theta} p_y^{0*} \tag{17}$$

is the posterior about $\tilde{\theta}$ given the public information, the Y-market price p_y^{0*}.

Equation (11) maximizes firm A's conditional expected profit in each of its local X-markets in period 1. Note that since all the X-markets are symmetric, the optimal production policies are the same across all M local markets. Similarly, equation (12) maximizes firm B's expected profit in its date-1 X-markets. Equations (13) and (14) form a linear Bayesian Nash equilibrium in the global Y-market in period 1. The following lemma characterizes the linear date-1 product market equilibrium without a data vendor.

Lemma 3 (Date-1 product market equilibrium without data sales)

In the economy without a data vendor, there exists a unique linear equilibrium in which

$$\begin{aligned}
X_{A,i}^1 &= \frac{1}{2}\left[\mu + \frac{M\tau_\varepsilon}{M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta}(P_A^0 - \mu)\right], i = 1, ..., M, \\
X_{B,j}^1 &= \frac{1}{2}\left[\mu + \frac{M\tau_\varepsilon}{M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta}(P_B^0 - \mu)\right], j = 1, ..., M, \\
Y_A^1 &= \frac{N}{3}\mu + \frac{MN\tau_\varepsilon}{3M\tau_\epsilon + 2(N\tau_\varepsilon + \tau_\theta)}(P_A^0 - \mu), \\
Y_B^1 &= \frac{N}{3}\mu + \frac{MN\tau_\varepsilon}{3M\tau_\varepsilon + 2(N\tau_\varepsilon + \tau_\theta)}(P_B^0 - \mu).
\end{aligned}$$

The equilibrium expected profits of firm A and firm B in period 1 are

$$\begin{aligned}
\mathbb{E}\Pi_A^{1*} = \mathbb{E}\Pi_B^{1*} = &\frac{M}{4}\left[\frac{N\tau_\varepsilon}{(N\tau_\varepsilon + \tau_\theta)\tau_\theta} + \frac{M\tau_\varepsilon}{(M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)(N\tau_\varepsilon + \tau_\theta)}\right] \\
&+ \frac{N}{9}\frac{N\tau_\varepsilon}{(N\tau_\varepsilon + \tau_\theta)\tau_\theta} + \frac{MN\tau_\varepsilon(M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)}{\left[3M\tau_\varepsilon + 2\left(N\tau_\varepsilon + \tau_\theta\right)\right]^2\left(N\tau_\varepsilon + \tau_\theta\right)}.
\end{aligned}$$

3.2 Data allocation, efficiency, and disclosure

3.2.1 Data allocation and welfare improvement

The equilibrium outcome characterized by Lemma 3 can be potentially improved in terms of social welfare by changing data allocations. Specifically, we consider an artificial situation in which both firms' private information $\mathbf{P}_A^0$ and $\mathbf{P}_B^0$, in addition to the public information p_y^{0*}, are commonly observed by the two firms. This corresponds to data allocation $(m_A, m_B) =$

(M, M) defined in Definition 1, and we label it with "**MM**". Under data allocation **MM**, both firms have the same information set, which is $\mathcal{F}_A^{\mathbf{MM}} = \mathcal{F}_B^{\mathbf{MM}} = \{p_y^{0*}, \mathbf{P}_A^0, \mathbf{P}_B^0\}$. Equipped with this new information set, firms still maximize their conditional expected profits in both local and global markets. The original data allocation in Lemma 3 is $(m_A, m_B) = (0, 0)$, and we label it with "$\emptyset\emptyset$" to indicate that both firms keep their private information secret.

We define the welfare variables—consumer surplus (CS) and total surplus (TS)—as follows:

$$CS \equiv \underbrace{\sum\nolimits_{i=1}^{M} \frac{1}{2}\mathbb{E}(X_{A,i}^1)^2}_{A\text{-type}} + \underbrace{\sum\nolimits_{i=1}^{M} \frac{1}{2}\mathbb{E}(X_{B,i}^1)^2}_{B\text{-type}} + \underbrace{\frac{1}{2}\mathbb{E}(Y_A^1 + Y_B^1)^2}_{AB\text{-type}}, \tag{18}$$

$$TS \equiv \underbrace{CS}_{\text{consumer surplus}} + \underbrace{\mathbb{E}\Pi_A^1 + \mathbb{E}\Pi_B^1}_{\text{producer surplus}}, \tag{19}$$

where $X_{A,i}^1, X_{B,i}^1, Y_A^1, Y_B^1, \mathbb{E}\Pi_A^1$, and $\mathbb{E}\Pi_B^1$ are the production policies and profits reached in the date-1 product market equilibrium when firms are equipped with their information sets.

Proposition 1 (Welfare gains)

Relative to data allocation $\emptyset\emptyset$, under data allocation **MM***, consumer surplus and total surplus are always higher, and firm profits are higher if and only if there are sufficiently many local markets. That is, $CS^{\mathbf{MM}} > CS^{\emptyset\emptyset}, TS^{\mathbf{MM}} > TS^{\emptyset\emptyset}$ for any M; and $\mathbb{E}\Pi_A^{1,\mathbf{MM}} = \mathbb{E}\Pi_B^{1,\mathbf{MM}} > \mathbb{E}\Pi_A^{1,\emptyset\emptyset} = \mathbb{E}\Pi_B^{1,\emptyset\emptyset}$ if and only if $M > \hat{M}$, where $\hat{M}$ is a constant given by equation (OA.56) in the appendix.*

Intuitively, when firms are equipped with better information, they can better collectively accommodate consumers' needs, which improve consumer surplus and total surplus. However, because of the strategic competing behavior, firms are worse off in the Y-market when their private information becomes public. This profit loss can be compensated by their more informed production decisions in their respective local X-markets, and if the number of these local markets is sufficiently large, the overall profit effect of sharing information is positive.

Figure 3: Efficiency and disclosure

Panel A: Welfare variables

Data allocation	Small M, $M = N/10$			Large M, $M = 10N$		
	TS	CS	$\mathbb{E}\Pi_A^1 + \mathbb{E}\Pi_B^1$	TS	CS	$\mathbb{E}\Pi_A^1 + \mathbb{E}\Pi_B^1$
$\emptyset\emptyset$	227.894	224.972	2.922	132474.67	132196.63	278.04
MM	240.379	237.463	2.916	171190.61	170836.85	353.76

Panel B: The payoff matrix for firms

$M = N/10$

		Firm B ND	Firm B D
Firm A	ND	$(1.461, 1.461)$	$(1.570, 1.349)$
	D	$(1.349, 1.570)$	$(1.458, 1.458)$

$M = 10N$

		Firm B ND	Firm B D
Firm A	ND	$(139.02, 139.02)$	$(179.02, 136.77)$
	D	$(136.77, 179.02)$	$(176.88, 176.88)$

Panel A shows the total surplus, consumer surplus, and total profits for the corresponding data allocation. Panel B is the payoff matrix for firms for the corresponding action non-disclosure, "ND", or disclosure, "D." In this numerical example, we assume $\tau_\theta = 1, \tau_\varepsilon = 0.001$ and $N = 100$. We consider two values of M: $M = N/10 = 10$ and $M = 10N = 1000$.

3.2.2 Voluntary and mandatory disclosure

Although data allocation **MM** improves on data allocation $\emptyset\emptyset$, it is not clear how such a data allocation is achieved in the first place. The information-sharing literature has considered whether firms would like to voluntarily share their private information, for instance, by forming a trade association that discloses the signals reported by its member firms (Gal-Or, 1985; Darrough, 1993; and see Vives (2016) for a survey). However, it is shown that withholding information is always a dominant strategy for firms in oligopoly settings with Cournot competition and demand uncertainty; that is, data allocation **MM** is not supported in equilibrium with voluntary disclosure.

To illustrate, let us consider the following numerical example. We set $\tau_\theta = 1, \tau_\varepsilon = 0.001$, and $N = 100$, and M can take two values: $M = \frac{N}{10} = 10$ or $M = 10N = 1000$. Consistent with Proposition 1, independent of the value of M, both consumer surplus and total surplus are higher under data allocation **MM** in Panel A of Figure 3. Also, when M is high, firms profits are higher under **MM**, and when M is low, firms profits are lower under **MM**.

In the context of voluntary information sharing, each firm faces a choice of disclosure (D) or nondisclosure (ND) of its own private information. This leads to the payoff matrices in Panel B of Figure 3. Each cell in this matrix is the equilibrium profits resulting from the date-1 product market equilibrium. For instance, if both firms choose not to disclose information, then the profits of each firm are given by the expression of $\mathbb{E}\Pi_A^{1*}$ and $\mathbb{E}\Pi_B^{1*}$ in Lemma 3. We can see that withholding information is a dominant strategy for each firm, so that (ND, ND) constitutes the unique Nash equilibrium at the information-sharing stage for both values of M. In particular, when M is high, the resulting payoff matrix is the "prisoners' dilemma,"[7] which predicts that firms would have been better off if both of them could disclose, which, however, is not a viable agreement in a noncooperative setting.

Given that voluntary disclosure is not viable, the literature also suggests mandatory disclosure through regulatory agencies such as the SEC or the FASB that, in theory, can force firms to disclose the information that firms wish hidden (e.g., see Darrough, 1993). However, mandatory disclosure can be costly. The cost stems not only from the administrative cost of implementing the disclosure rules but also from some other economic costs. Firms could take strategic actions to respond to regulatory requirements, for instance, by adding noises or a large amounts of nonmaterial and raw information of little value in the public disclosure.[8] The root reason for this kind of cost is that mandatory disclosure regulations run against

[7]Darrough (1993) also identifies a prisoners' dilemma in an information-sharing setting, although for a different reason. Specifically, in Darrough's setting, a prisoners' dilemma arises when firms' products are sufficiently different. By contrast, the firms' products are perfect substitute in our setting, and the prevalence of a prisoners' dilemma depends on the number of local markets.

[8]Evidence supporting this argument is provided by extensive studies on Regulation Fair Disclosure (Reg FD) which, promulgated by the SEC in 2000, mandates that all publicly traded companies must disclose material information to the general public at the same time. For instance, Bailey, Li, Mao, and Zhong (2003) find that the Reg FD could make the public communication become "sound bites" with "boilerplate" disclosures. A survey conducted by Security Industry Association shows that 72% of analysts interviewed during the survey mention that information communicated by issuers to the public is of lower quality after the Reg FD regulation (http://www.sia.com/testimony/html/kaswell5 -17.html). Cohen, Lou, and Malloy (2017) document that firms could "cast" their conference calls and thus control the information flow released to the public even after Reg FD. Bushee, Matsumoto, and Miller (2004) find that Reg FD had a significant negative impact on managers' decisions to continue hosting conference calls and on their decisions regarding the optimal time to hold.

firms' private incentives to maximize their own profits. In the following two sections, we will show that data sales instead can incentivize firms to reach the more efficient data allocation **MM**.

4 Welfare-improving data sales

We now solve the model with data sales described in Section 2, in which the data vendor maximizes its own profits. We first solve the date-1 product market equilibrium for any given amount of data purchase, (m_A, m_B), and then solve the optimal data sales (m_A^*, m_B^*). Finally, we discuss the welfare consequences of data sales.

4.1 Product market equilibrium

Suppose that on date 0, firms A and B have respectively purchased m_A and m_B local market prices from the data vendor. Note that the consumer data purchased by firm A is about firm B's date-0 X-markets, and vice versa. We assume that the consumer identities of the sold data are anonymous. The data vendor can achieve this goal by randomly sampling from the pool of all date-0 consumers. Nonetheless, we assume that the data vendor ensures that the data is indeed useful for firms (i.e., the data bought by firm A is drawn from firm B's X-market prices and vice versa). Let us label the randomly drawn consumers by $\{j_1, ..., j_{m_A}\}$ and $\{i_1, ..., i_{m_B}\}$ for the two sold data sets. The data sets purchased by firms A and B are, respectively,

$$\mathbf{I}_A = \left\{ p^{0*}_{B,j_1}, ..., p^{0*}_{B,j_{m_A}} \right\} \text{ and } \mathbf{I}_B = \left\{ p^{0*}_{A,i_1}, ..., p^{0*}_{A,i_{m_B}} \right\}.$$

As in Section 3, we still consider linear equilibria in which optimal production policies are linear in firms' information variables. Also, given the symmetry of the purchased market data, it is natural to specify that the coefficients on the purchased prices are the same. Thus, we conjecture the following date-1 production policies:

$$X_{A,i}^1 = \Phi_{A_0}^X + \Phi_{A_1}^X (P_A^0 - \mu) + \Phi_{A_2}^X (I_A - \mu), i = 1, ..., M, \tag{20}$$

$$X_{B,j}^1 = \Phi_{B_0}^X + \Phi_{B_1}^X (P_B^0 - \mu) + \Phi_{B_2}^X (I_B - \mu), j = 1, ..., M, \tag{21}$$

$$Y_A^1 = \Phi_{A_0}^Y + \Phi_{A_1}^Y (P_A^0 - \mu) + \Phi_{A_2}^Y (I_A - \mu), \tag{22}$$

$$Y_B^1 = \Phi_{B_0}^Y + \Phi_{B_1}^Y (P_B^0 - \mu) + \Phi_{B_2}^Y (I_B - \mu), \tag{23}$$

where

$$I_A \equiv \frac{1}{m_A} \sum\nolimits_{a=1}^{m_A} p_{B,j_a}^{0*} = \frac{1}{m_A} \sum\nolimits_{a=1}^{m_A} s_{B,j_a}^0, \tag{24}$$

$$I_B \equiv \frac{1}{m_B} \sum\nolimits_{b=1}^{m_B} p_{A,i_b}^{0*} = \frac{1}{m_B} \sum\nolimits_{b=1}^{m_B} s_{A,i_b}^0, \tag{25}$$

where the second equality in (24) and (25) follows from Lemma 1, and P_A^0, P_B^0, and μ are given by equations (15), (16), and (17), respectively.

Equations (20) and (21) maximize expected profits in the local X-markets respectively for firm A and firm B. Equations (22) and (23) form a linear Bayesian Nash equilibrium in the global Y-market. The following proposition characterizes the product market equilibrium

Proposition 2 (Product market equilibrium)

For any given data purchase (m_A, m_B), there exists a linear product market equilibrium characterized by equations (20)–(23), where the Φ-coefficients are given in the appendix. The equilibrium expected profits $\mathbb{E}\Pi_A^{1*}$ and $\mathbb{E}\Pi_B^{1*}$ are given in the appendix.

4.2 Equilibrium data sales

At the beginning of date 0, the data vendor designs contracts to maximize firms' willingness to pay for data, $C_A(m_A, m_B)$ and $C_B(m_A, m_B)$, given by equation (9). It turns out that the data vendor's profit is maximized when both firms purchase the maximum amount of data. These results are formalized in the following proposition.

Proposition 3 (Optimal data sales)

In equilibrium, the data vendor sells all of its data to firms, that is, $m_A^* = m_B^* = M$*. The resulting data prices are*

$$C_A^* = C_B^* = \left(\frac{M}{4} + \frac{N}{9}\right) \frac{M\tau_\varepsilon}{(2M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)(M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)}.$$

Setting $m_A = m_B = M$ in Proposition 2, we obtain the overall sequential equilibrium for the economy with a profit-maximizing data vendor.

Proposition 4 (Overall equilibrium)

On date 0, the data vendor sells all of its data to firms. On date 1, the optimal production policies in product markets are:

$$Y_A^{1*} = Y_B^{1*} = \frac{N}{3}\left(\mu + \frac{M\tau_\varepsilon}{2M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta}[P_A^0 - \mu + P_B^0 - \mu]\right),$$
$$X_{A,i}^{1*} = X_{B,j}^{1*} = \frac{1}{2}\left(\mu + \frac{M\tau_\varepsilon}{2M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta}[P_A^0 - \mu + P_B^0 - \mu]\right),$$

for $i, j = 1, ..., M$*. The equilibrium date-1 expected profits (gross of data price* C^**) are*

$$\mathbb{E}\Pi_A^{1*} = \mathbb{E}\Pi_B^{1*} = \frac{M}{4}\left[\frac{N\tau_\varepsilon}{(N\tau_\varepsilon + \tau_\theta)\,\tau_\theta} + \frac{2M\tau_\varepsilon}{(2M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)(N\tau_\varepsilon + \tau_\theta)}\right]$$
$$+ \frac{N}{9}\left[\frac{N\tau_\varepsilon}{(N\tau_\varepsilon + \tau_\theta)\,\tau_\theta} + \frac{2M\tau_\varepsilon}{(2M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)(N\tau_\varepsilon + \tau_\theta)}\right].$$

4.3 Intuitions, implementation, and welfare

We now use a numerical example in Figure 4 to illustrate better what is going on in the economy with data sales. The parameter values in Figure 4 are the same as those in Figure 3 with $M = 1000$. In Panel A of Figure 4, we plot the payoff matrix for firms. By Proposition 3, in equilibrium, the data vendor will implement data allocation **MM** and charge a price C^* for the data. Thus, firms' actions are either to reject the data vendor's contracts and not purchase data, or to accept the contracts and acquire an amount M of consumer data. Here, we allow the data vendor to be able to design "smart contracts" which allow data prices to depend on data allocations. Specifically, the contracts state the following: "*If data allocation is* $(0, M)$*, then firm A pays price* t_{0M}*; if data allocation is* $(M, 0)$*, then firm B pays price* t_{0M}*; and if data allocation is* (M, M)*, then both firms pay a price* t_{MM}." By Proposition 3, we know that the equilibrium value of t_{MM} must equal C^*, which is 41.11 in this example. We now explain why this is the case and what values t_{0M} can take.

Figure 4: Data sales and efficiency

Panel A: The payoff matrix for firms

		Firm B: 0	Firm B: M
Firm A	0	$(139.02, 139.02)$	$(136.77, 179.02 - t_{0M})$
	M	$(179.02 - t_{0M}, 136.77)$	$(176.88 - t_{MM}, 176.88 - t_{MM})$

Panel B: Welfare variables

Data allocation	TS	CS	$\mathbb{E}\Pi_A^1 + \mathbb{E}\Pi_B^1$	Data vendor's profits
∅∅	132474.67	132196.63	278.04	0
MM	171190.61	170836.85	273.54	80.22

Panel A is the payoff matrix for firms for the corresponding actions — purchasing "0" local market price or "M" local market prices. The payoff is the expected profit net cost of buying information from the data vendor. t_{0M} is the cost of data when one firm buys 0 and the other buys M, and t_{MM} is the cost when both firms purchase M local market prices. Panel B shows the total surplus, consumer surplus, total profits for firms, and total profits for the data vendor. The total surplus includes the data vendor's profits. In this numerical example, we assume $\tau_\theta = 1, \tau_\varepsilon = 0.001$ and $N = 100$. Since we focus on the "prison dilemma" problem, we consider only when M is large, i.e., $M = 10N$.

The first observation is the following. Suppose that the data prices t_{0M} and t_{MM} are set at 0. Then, comparing the payoff matrix in Panel A of Figure 4 with that in Panel B of Figure 3, we find that the former is a transpose of the latter. Note that in Figure 3, firms' actions are disclosing or not disclosing information and thus, there, firms are considering whether to *supply* information for free. In contrast, in Figure 4, firms are considering whether to buy information at a cost, which is about the *demand* side of data. This switch between supply and demand perspectives transposes the payoff matrix, which in turn changes the equilibrium data allocations.

Recall that in equilibrium, the data vendor wants to implement data allocation (M, M), which leads to the highest profits. One way of implementation is to choose appropriate values of t_{0M} and t_{MM}, such that purchasing data is a dominant strategy for both firms. This requires the following:

$$179.02 - t_{0M} > 139.02 \Rightarrow t_{0M} < 40,$$
$$176.88 - t_{MM} > 136.77 \Rightarrow t_{MM} < 40.11.$$

Thus, by setting $t_{MM} = 40.11$ and $t_{0M} < 40$, the data vendor can sell data to both firms, collecting a total profit of $2 \times t_{MM} = 80.22$. Intuitively, the upper bounds of t_{0M} and t_{MM} are firms' willingness to pay at data allocations $(0, M)$ and (M, M), respectively. In our setting, there is strategic complementarity in firms' data purchase behavior: Firm A's willingness to pay is higher when firm B is buying data than when firm B is not (i.e., $C_A(M, M) = C_B(M, M) = 40.11 > 40 = C_A(M, 0) = C_B(M, 0)$). In consequence, the data vendor can achieve the highest profit when both firms buy data, because in this case, not only the data vendor is selling to two instead of one firm, but also each firm is willing to pay more, relative to the case in which only one firm buys data. This complementarity result holds true in general as formalized in the following proposition.

Proposition 5 (Complementarity)

In equilibrium, firms' information purchase decisions are a strategic complement, that is, $\frac{\partial C_A(M,m_B)}{\partial m_B} > 0$ *and* $\frac{\partial C_B(m_A,M)}{\partial m_A} > 0$.

The presence of a data vendor effectively moves the equilibrium data allocation from $\emptyset\emptyset$ to **MM**. That is, in the benchmark economy without data sales, both firms keep their private information secret and thus no firm can see its rival's private information (which corresponds to data allocation $\emptyset\emptyset$). Here, with data purchase, both firms can observe the private information of their respective rivals, although at a cost. This leads to the data allocation **MM**. By Proposition 1, both consumer surplus CS and total surplus TS are improved with the introduction of a data vendor, where CS and TS are still defined by equations (18) and (19), respectively. Panel B of Figure 4 confirms these results.

Proposition 6 (Welfare-improving data sales)

The introduction of a profit-maximizng data vendor improves both consumer surplus and total surplus in equilibrium.

4.4 When are the firms better off?

In Panel B of Figure 4, firms' equilibrium profits are lower with data sales than without. Thus, although introducing a data vendor improves total surplus, it is not a Pareto improvement. Recall that by Proposition 1, when the number M of local markets is sufficiently high, both firms can be better off under data allocation **MM** than under data allocation $\emptyset\emptyset$, if they had not paid any costs to acquire the data. The underlying reason that firms get worse off in Panel B of Figure 4 (where M is relatively high) is due to the assumption that the data vendor has all the market power in the data market. If we relax this assumption, then firms can be better off as well with the introduction of a data vendor for a sufficiently high M, so that the introduction of a data vendor indeed leads to a Pareto improvement.

Specifically, let us consider a setting in which firms could bargain with the data vendor in the data market. This may be reasonable given that both the data vendor and the two firms are big players in the data market. Now suppose that a firm can negotiate over the data price when receiving sales contracts from the data vendor, and the data price C is set through Nash bargaining between the data vendor and the firm. We use $\beta \in (0, 1)$ to denote the data vendor's bargaining power. Our baseline model corresponds to the degenerate case with $\beta = 1$.

The bargaining outcome depends on each agent's utility in the events of agreement versus no agreement. For a firm, say, firm A, the utility when agreeing on a data price C_A is $\mathbb{E}\Pi_A^1(m_A, m_B) - C_A$. If no agreement is reached, then firm A's outside option value is

$\mathbb{E}\Pi_A^1(0, m_B)$. The data vendor's gain from agreement is the data price C_A. The bargaining outcome maximizes the Cobb-Douglas product of the utility gains from agreement:

$$\max_{C_A} C_A^{\beta} \left(\mathbb{E}\Pi_A^1(m_A, m_B) - C_A - \mathbb{E}\Pi_A^1(0, m_B)\right)^{1-\beta}.$$

The solution leads to the data price as follows:

$$C_A\left(m_A, m_B\right) = \beta \times \left[\mathbb{E}\Pi_A^1(m_A, m_B) - \mathbb{E}\Pi_A^1(0, m_B)\right]. \tag{26}$$

Comparing the above expression with equation (9), we find that the only difference in this generalized setting is the scaling fraction β. This fraction does not affect the data vendor's profit-maximization problem and thus, when the data vendor designs contracts to implement data allocations, it still chooses $m_A^* = m_B^* = M$. The resulting overall equilibrium is given by Proposition 4 with a smaller data price:

$$C_A^* = C_B^* = \beta \left(\frac{M}{4} + \frac{N}{9}\right) \frac{M\tau_\varepsilon}{(2M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)(M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)}.$$

As a result, the net profits of firms become higher. In particular, when the data vendor has small bargaining power and when there are sufficiently many local markets, firms are better off with data sales, so that the introduction of a data vendor leads to a Pareto improvement.

Proposition 7 (Nash bargaining)

The introduction of a data vendor benefits firms and a Pareto improvement if and only if the following two conditions are satisfied:

$$M > \hat{M} \text{ and } \beta < 1 - \left(\frac{M}{4} + \frac{N}{9}\right)^{-1} \frac{N\left[6M\tau_\varepsilon + 5(N\tau_\varepsilon + \tau_\theta)\right]\left(2M\tau_\epsilon + N\tau_\varepsilon + \tau_\theta\right)}{9\left[3M\tau_\epsilon + 2(N\tau_\varepsilon + \tau_\theta)\right]^2},$$

where $\hat{M}$ is a constant defined by equation (OA.56) in the appendix.

5 Data ownership and vendor formation

In previous sections, we assume that the data vendor maximizes its own profits and we do not explore how such a profit-maximizing vendor arises. The data vendor's emergence and its resulting objective function may depend on the original data ownership. In this section, we first discuss how an independent profit-maximizing data vendor can endogenously arise in cases in which the data is originally owned by platforms (such as Amazon.com, Inc.) or by consumers. We then examine a variation setting in which the data is owned by firms and show how firms can form a data vendor to maximize their total profits. Our analysis provides useful insights for the current debates on data ownership and privacy.

5.1 Platforms, consumers, and independent data vendors

Nowadays, numerous consumers data were held by many transaction and settlement platforms, such as Amazon, Paypal, and Taobao. If the ownership of these data belongs to these platforms, then these platforms can sell the accumulated data to firms who in turn use the data to make more informed production decisions. In this case, these platforms correspond directly to the data vendor in Section 2, and their objective is to maximize their own profits.

When consumers make purchase decisions in these platforms, they may have implicitly concurred to give up their data ownership to these platforms by signing some agreements without carefully reading the contents. Now consumers start to understand that their data have value and that they are due some compensation. Some startups, under the concept of "data locker," have already taken this kind of initiatives to give consumers more control over their own data and the opportunity to earn compensation.[9] The recent development of blockchain technology makes such a compensation easier to implement, because this new technology is well suited for effectively defining and protecting data ownership. One issue in this context is that consumers do not know how much their data are worth in terms of dollars

[9] "Data mining offers rich seam," February 18, 2013, *Financial Times*.

and how to trade off this monetary benefit against the potential cost of leaking privacy.[10] Our analysis in the previous sections provides an upper bound for the potential market value of data, namely, the profits earned by the data vendor.

Formally, suppose that the transaction data is originally owned by the data-0 consumers in our setting. These consumers can seize the compensation by forming a profit-maximizing data vendor, for example, via an initial coin offering (ICO).[11] By Proposition 3, the total profits accruing to the data vendor are $C_A^* + C_B^* = \left(\frac{M}{4} + \frac{N}{9}\right) \frac{2M\tau_\varepsilon}{(2M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)(M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta)}$. If data-0 consumers' valuations toward privacy are lower than $C_A^* + C_B^*$, then it is beneficial for them to sell their transaction data.[12] When consumers' privacy concerns are heterogeneous, their decisions can be different; intuitively, in equilibrium, those consumers who care about privacy the least would like to contribute their data and become a shareholder of the data vendor. We leave a formal analysis of this kind for future research.

5.2 Firms as data owners form the data vendor

It is also natural to assume that the data-0 consumer data is owned by firms, since they are important participants in producing such data. As Section 3 shows, when firms are original data owners, they have no incentives to share their private consumer data, although information sharing can be better for them if there is a sufficiently large number of local markets. However, the information sharing considered by the literature is sharing "for free." Then, how about sharing "for a price"? For instance, suppose that firms can form a data vendor who purchases data from and sells data to firms. Can such a data vendor move the equilibrium data allocation from $\emptyset\emptyset$ to **MM**, as achieved in Section 4?

In relation to the information-sharing literature (e.g., Gal-Or, 1985; Vives, 2006), the data vendor corresponds to a trade association examined by the literature. In the literature,

[10] "Fuel of the future—Data is giving rise to a new economy," May 6, 2017, *Economist*. Also see Acquisti (2014) for related discussions.

[11] See Li and Mann (2018) for an analysis on how a data vendor can be formed via an ICO.

[12] Existing experimental studies suggest that consumers' valuations about privacy are relatively small, ranging from 0.50 to 45 US dollars (see Section 5 of Acquisti (2014)).

a trade association collects information from firms at no cost and distributes information to firms for free. Here, the data vendor, which is the counterpart of a trade association, pays a price to a firm that contributes data to the vendor, and charges a price from a firm that acquires data from the vendor. Given that both firms are the shareholders of the data vendor, now it is natural to assume that the data vendor maximizes the total profits of both firms (as opposed to the vendor's own profits in Section 2), and retains no profits for itself.

In this case, the data vendor has incentives to move data allocation from $\emptyset\emptyset$ to **MM** if and only if the number M of local markets is sufficiently high, since by Proposition 1, firms are better off if and only if M is high. Since the data vendor retains no profits, the data transactions are equivalent to the following transfers between firms: Firm A makes transfer t_A to firm B for firm B's private consumer data and firm B makes t_B transfer to firm A for firm A's private consumer data. Does there exist a set of transfers (t_A, t_B) that supports the data allocation **MM**, when the number M of local markets is sufficiently large? The answer to this question is positive. Now let explain how.

Given the data vendor now behaves like a two-sided market, we need to consider both the data supply and demand from firms, which correspond respectively to Figures 3 and 4 in previous sections. In Figure 5, we adopt the same parameter values as those in Figure 4. Panel A of Figure 5 describes the payoff matrix when firms supply data to the data vendor. This corresponds to Panel B of Figure 3, which assumes $t_A = t_B = 0$ (i.e., public disclosure means no compensation for supplying data). From the payoff matrix, we see that when $t_A \geq 2.25$ and $t_B \geq 2.25$, supplying information is the dominant strategy for both firms. Intuitively, when data prices are sufficiently high, both firms are willing to sell their data. Panel B of Figure 5 draws the payoff matrix when firms demand data from the vendor. This corresponds to Panel A of Figure 3 (with t_{0M} and t_{MM} replaced with t_A and t_B). Apparently, when $t_A \leq 40$ and $t_B \leq 40$, the data prices are sufficiently low such that firms always want to buy data from the data vendor. Taken together, we conclude that any transfer $(t_A, t_B) \in [2.25, 40]^2$ can support data allocation **MM**.

Figure 5: Data sales as transfers

Panel A: The payoff matrix when firms supply information

		Firm B: 0	Firm B: M
Firm A	0	$(139.02, 139.02)$	$(179.02, 136.77 + t_A)$
	M	$(136.77 + t_B, 179.02)$	$(176.88 + t_B, 176.88 + t_A)$

Panel B: The payoff matrix when firms demand information

		Firm B: 0	Firm B: M
Firm A	0	$(139.02, 139.02)$	$(136.77, 179.02 - t_B)$
	M	$(179.02 - t_A, 136.77)$	$(176.88 - t_A, 176.88 - t_B)$

Panel C: Welfare variables

Data allocation	TS	CS	$\mathbb{E}\Pi_A^1 + \mathbb{E}\Pi_B^1$ net transfers
$\emptyset\emptyset$	132474.67	132196.63	278.04
MM	171190.61	170836.85	353.76

Panel A is the payoff matrix for firms when they supply information. Panel B is the payoff matrix for firms when they demand information. Panel C reports the total surplus, consumer surplus and firms' net profits, i.e., the total profits net transfers. In this numerical example, we assume $\tau_\theta = 1, \tau_\varepsilon = 0.001$ and $N = 100$, $M = 10N$.

Proposition 8

Any transfer (t_A, t_B) in the following rectangular set can support data allocation **MM***:*

$$(t_A, t_B) \in \left[\frac{MN\tau_\varepsilon\left[6M\tau_\varepsilon + 5(N\tau_\varepsilon + \tau_\theta)\right]}{9\left(M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta\right)\left[3M\tau_\varepsilon + 2(N\tau_\varepsilon + \tau_\theta)\right]^2}, \frac{M\tau_\varepsilon\Theta}{36\left(M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta\right)\left(2M\tau_\varepsilon + N\tau_\varepsilon + \tau_\theta\right)\left[3M\tau_\varepsilon + 2(N\tau_\varepsilon + \tau_\theta)\right]^2}\right]^2,$$

where

$$\Theta = 81M^3\tau_\varepsilon^2 + 33M^2N\tau_\varepsilon^2 + 108M^2\tau_\varepsilon(N\tau_\varepsilon + \tau_\theta) + 44MN\tau_\varepsilon(N\tau_\varepsilon + \tau_\theta) + 36M(N\tau_\varepsilon + \tau_\theta)^2 + 16N(N\tau_\varepsilon + \tau_\theta)^2.$$

The above set is non-empty for a sufficiently large M.

Proposition 8 also illustrates why voluntary disclosure is not viable in Section 3. Specifically, as we mentioned above, voluntary disclosure essentially sets $t_A = t_B = 0$, which does not lie in the rectangular set. Intuitively, when $t_A = t_B = 0$, the data price is so low that no firms want to supply information, leading to an equilibrium data allocation $\emptyset\emptyset$.

Our discussions in Sections 5.1 and 5.2 suggest that data ownership may matter for social welfare through changing the objectives of the data vendor. Specifically, if data belongs to consumers or platforms, the data vendor is likely to maximize its own profits, and data sales always changes the equilibrium data allocation from $\emptyset\emptyset$ to **MM** independent of the number M of local markets. This change in data allocation increases total surplus. However, if firms own the data and form a data vendor that maximizes the total profits of both firms, then data sales changes data allocation and improves total surplus only for sufficiently large M. This observation suggests that it may be better to give ownership to consumers than to firms, provided that consumers can effectively monetize the value of their transaction data.

6 Conclusion

In this paper, we study the value of data based on the classical duopoly competition setting. In our model, duopoly firms use past consumption data to predict future demand and adjust their production plan accordingly. Data is valuable as it optimizes firms production behavior. But firms are not able to explore the full advantage of data, since they are not willing to share their private information. The emergence of the data vendor resolves this issue. With the data vendor, the equilibrium becomes the full information sharing equilibrium. The social welfare is increased. In some cases, the increase is the Pareto improvement for all participants in the economy.

Our model highlights the importance of the data vendor to the efficiency of the real economy. We further explore the question of who should own the data to form such an organization. It is straightforward that when consumers own the data, they will organize such a data vendor to share the value of the data. This leads to full information sharing economy. On the other hand, when firms own the data, it becomes less straightforward. Firms have to balance the benefit from more information with the cost of losing their competitive advantage. We show that when firms build the data vendor as a two-sided market, then there exists a set of transfers between firms that can support the full information sharing economy. That said, not any transfers can support the full information sharing economy. For example, when transfers are set to zeros. The equilibrium becomes no information sharing.

Data are to this century what oil was to the last one. There are many concerns over the springing data economy, such as the issue of data ownership. While many discussions are from a legal or technology perspective, we approach this issue from an economic one. We find data ownership can potentially affect the economic efficiency. While assigning ownership to consumers improves welfare, assigning ownership to firms may not always achieve the same target. As so, we believe regulators should devote more consideration to the case when the data ownership is assigned to firms.

Appendix

Proof for Proposition 1

Proof. From the first order condition of consumers' utility maximization, we get the total expected consumer surplus for A-type and B-type consumers as

$$CS_A = \sum_{i=1}^{M} CS_{A,i} = \sum_{i=1}^{M} \frac{1}{2}\mathbb{E}(x^1_{A,i})^2 = M\frac{1}{2}\mathbb{E}\left[\mathbb{E}\left[\left(x^1_{A,i}\right)^2 \mid \mathcal{F}_A\right]\right] = \frac{M}{2}\mathbb{E}\Pi^1_{A,X}. \tag{A.1}$$

The last equality is implied by the market clearing condition. Similarly, the market clearing condition implies the AB-type consumer's surplus is,

$$\begin{aligned} CS_{AB} &= \frac{1}{2}\mathbb{E}\left[\left(Y^1_A + Y^1_B\right)^2\right] && \text{(A.2)} \\ &= \frac{1}{2}\mathbb{E}\left[(Y^1_A)^2 + 2Y^1_A Y^1_B + (Y^1_B)^2\right] \\ &= \frac{N}{2}\left(\mathbb{E}\Pi^1_{A,Y} + \mathbb{E}\Pi^1_{B,Y}\right) + \mathbb{E}(Y^1_A Y^1_B). \end{aligned}$$

1. Applying the optimal production in $\Delta\mathcal{F}_A, \Delta\mathcal{F}_B = \emptyset$ and $\Delta\mathcal{F}_A = \mathbf{M}, \Delta\mathcal{F}_B = \mathbf{M}$, we have

$$CS_A(\emptyset,\emptyset) = CS_B(\emptyset,\emptyset) = \frac{M}{2}\mathbb{E}\Pi^1_{A,X}(\emptyset,\emptyset) = \frac{M^2}{8}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{M\tau_\epsilon}{(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}\right). \tag{A.3}$$

For AB-type consumers,

$$\mathbb{E}\Pi^1_{A,Y}(\emptyset,\emptyset) = \mathbb{E}\Pi^1_{B,Y}(\emptyset,\emptyset), \tag{A.4}$$

$$\mathbb{E}(Y^1_A Y^1_B)(\emptyset,\emptyset) = \Phi^2_{A_0} + \Phi^2_{A_1}\frac{1}{N\tau_\epsilon+\tau_\theta} = \frac{N^2}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \left(\frac{MN\tau_\epsilon}{3M\tau_\epsilon + 2(N\tau_\epsilon+\tau_\theta)}\right)^2\frac{1}{N\tau_\epsilon+\tau_\theta}. \tag{A.5}$$

Therefore,

$$CS_{AB}(\emptyset,\emptyset) = \frac{2N^2}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{2M^2N^2\tau_\epsilon^2 + MN^2\tau_\epsilon(N\tau_\epsilon+\tau_\theta)}{(3M\tau_\epsilon+2(N\tau_\epsilon+\tau_\theta))^2(N\tau_\epsilon+\tau_\theta)}. \tag{A.6}$$

The total consumer surplus is

$$CS(\emptyset,\emptyset) = 2CS_A(\emptyset,\emptyset) + CS_{AB}(\emptyset,\emptyset). \tag{A.7}$$

And the total surplus is

$$TS(\emptyset,\emptyset) = 2\mathbb{E}\Pi_A^1(\emptyset,\emptyset) + CS(\emptyset,\emptyset). \tag{A.8}$$

2. Repeating above with $\Delta\mathcal{F}_A = \mathbf{M}, \Delta\mathcal{F}_B = \mathbf{M}$, we get

$$CS_A(\mathbf{M},\mathbf{M}) = CS_B(\mathbf{M},\mathbf{M}) = \frac{M^2}{8}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{M\tau_\epsilon}{(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}\right), \tag{A.9}$$

$$CS_{AB}(\mathbf{M},\mathbf{M}) = N\mathbb{E}\Pi_{A,Y}^1(\mathbf{M},\mathbf{M}) + \mathbb{E}(Y_A^1Y_B^1)(\mathbf{M},\mathbf{M}) \tag{A.10}$$

$$= \frac{2N^2}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{2M^2N^2\tau_\epsilon^2 + MN^2\tau_\epsilon(N\tau_\epsilon+\tau_\theta)}{(3M\tau_\epsilon+2(N\tau_\epsilon+\tau_\theta))^2(N\tau_\epsilon+\tau_\theta)}, \tag{A.11}$$

$$TS(\mathbf{M},\mathbf{M}) = 2\mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}) + CS(\mathbf{M},\mathbf{M}). \tag{A.12}$$

Direct computation shows that

$$CS(\emptyset,\emptyset) < CS(\mathbf{M},\mathbf{M}), TS(\emptyset,\emptyset) < TS(\mathbf{M},\mathbf{M}). \tag{A.13}$$

■

Proof for Proposition 2

Proof. Since firm A and firm B are symmetric in choosing optimal production, we use firm A to illustrate the optimal production decision. Similar argument can be applied to firm B to yield the expression in Proposition 2.

Given any belief $(\hat{m}_A, \hat{m}_B)$, firm A's optimal production decision is to maximize her profits in the X-market and the Y-market. Combining the market clearing condition ($x^1_{A,i} = X^1_{A,i}, \sum_{i=1}^N y^1_{AB,i} = Y^1_A + Y^1_B$) which implies $P^1_{A,i} = \tilde{s}^1_{A,i} - X^1_{A,i}, p^1_y = \frac{\sum_{i=1}^N \tilde{s}^1_{AB,i} - Y^1_A - Y^1_B}{N}$, firm A's production decision is

$$\max_{x^1_{A,i}} \mathbb{E}[P^1_{A,i} X^1_{A,i} \mid \mathcal{F}_A] = \max_{x^1_{A,i}} \mathbb{E}\left[\left(\tilde{s}^1_{A,i} - X^1_{A,i}\right) X^1_{A,i} \mid \mathcal{F}_A\right], \text{for } i = 1, 2, ..., M; \tag{A.14}$$

$$\max_{Y^1_A} \mathbb{E}\left[\left(\frac{\sum_{i=1}^N \tilde{s}^1_{AB,i} - Y^1_A - \hat{Y}^1_B}{N}\right) Y^1_A \mid \mathcal{F}_A\right]. \tag{A.15}$$

The optimal production is

$$X^1_{A,i} = \frac{1}{2}\mathbb{E}\left(\tilde{\theta} \mid \mathcal{F}_A\right), \tag{A.16}$$

$$Y^1_A = \frac{N}{2}\mathbb{E}(\tilde{\theta} \mid \mathcal{F}_A) - \frac{1}{2}\mathbb{E}(\hat{Y}^1_B \mid \mathcal{F}_A). \tag{A.17}$$

Since $\mathcal{F}_A = \left\{P^0_A, p^0_y, I_A\right\}$, from the Bayesian updating, we have

$$\mathbb{E}(\tilde{\theta} - \mu \mid \mathcal{F}_A) = \begin{bmatrix} \frac{M\tau_\epsilon}{M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta)} \\ \frac{\tau_\epsilon m_A}{M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta)} \end{bmatrix} [P^0_A - \mu \quad I_A - \mu], \tag{A.18}$$

$$\mathbb{E}(P^0_B - \mu \mid \mathcal{F}_A) = \begin{bmatrix} \frac{M\tau_\epsilon - \tau_\epsilon m_A}{M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta)} \\ \frac{\tau_\epsilon m_A(2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}{M\tau_\epsilon(M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta))} \end{bmatrix} [P^0_A - \mu \quad I_A - \mu], \tag{A.19}$$

$$\mathbb{E}(I_B - \mu \mid \mathcal{F}_A) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} [P^0_A - \mu \quad I_A - \mu], \tag{A.20}$$

where $\mu = \frac{N\tau_\epsilon}{N\tau_\epsilon + \tau_\theta} p^0_y$.

For the Y-market, combining with the conjecture linear strategy,

$$Y_A^1 = \hat{\Phi}^Y_{A_0} + \hat{\Phi}^Y_{A_1}(P_A^0 - \mu) + \hat{\Phi}^Y_{A_2}(I_A - \mu), \tag{A.21}$$

$$Y_B^1 = \hat{\Phi}^Y_{B_0} + \hat{\Phi}^Y_{B_1}(P_B^0 - \mu) + \hat{\Phi}^Y_{B_2}(I_B - \mu), \tag{A.22}$$

(where $\hat{\Phi}^Y_{A_{(0,1,2)}}$ and $\hat{\Phi}^Y_{B_{(0,1,2)}}$ depends on $(\hat{m}_A, \hat{m}_B)$), we have

$$\begin{aligned} Y_A^1 &= \begin{bmatrix} 1 & P_A^0 - \mu & I_A - \mu \end{bmatrix} \\ &\times \left\{ \frac{N}{2} \begin{bmatrix} \mu \\ \frac{M\tau_\epsilon}{M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta)} \\ \frac{\tau_\epsilon m_A}{M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta)} \end{bmatrix} - \frac{\hat{\Phi}^Y_{B_0}}{2} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{\hat{\Phi}^Y_{B_1}}{2} \begin{bmatrix} 0 \\ \frac{M\tau_\epsilon - \tau_\epsilon m_A}{M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta)} \\ \frac{\tau_\epsilon m_A (2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}{M\tau_\epsilon (M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta))} \end{bmatrix} - \frac{\hat{\Phi}^Y_{B_2}}{2} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\} \\ &= \begin{bmatrix} 1 & P_A^0 - \mu & I_A - \mu \end{bmatrix} \begin{bmatrix} \frac{N}{2}\mu - \frac{\hat{\Phi}^Y_{B_0}}{2} \\ \frac{N}{2} M\tau_\epsilon \Lambda_A - \hat{\Phi}^Y_{B_1} \frac{M\tau_\epsilon - \tau_\epsilon m_A}{2} \Lambda_A - \hat{\Phi}^Y_{B_2} \frac{1}{2} \\ \frac{N}{2} \tau_\epsilon m_A \Lambda_A - \hat{\Phi}^Y_{B_1} \frac{\tau_\epsilon m_A (2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}{2M\tau_\epsilon} \Lambda_A \end{bmatrix}, \end{aligned} \tag{A.23}$$

where $\Lambda_A^{-1} = M\tau_\epsilon + \tau_\epsilon m_A + (N\tau_\epsilon + \tau_\theta)$.

Comparing with the conjecture strategy, we get

$$\begin{aligned} \underbrace{\begin{bmatrix} \Phi^Y_{A_0} \\ \Phi^Y_{A_1} \\ \Phi^Y_{A_2} \end{bmatrix}}_{\Phi^Y_A} &= \begin{bmatrix} \frac{N}{2}\mu - \frac{\hat{\Phi}^Y_{B_0}}{2} \\ \frac{N}{2} M\tau_\epsilon \Lambda_A - \hat{\Phi}^Y_{B_1} \frac{M\tau_\epsilon - \tau_\epsilon m_A}{2} \Lambda_A - \hat{\Phi}^Y_{B_2} \frac{1}{2} \\ \frac{N}{2} \tau_\epsilon m_A \Lambda_A - \hat{\Phi}^Y_{B_1} \frac{\tau_\epsilon m_A (2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}{2M\tau_\epsilon} \Lambda_A \end{bmatrix} \\ &= \frac{N}{2} \Lambda_A \begin{bmatrix} \Lambda_A^{-1} \mu \\ M\tau_\epsilon \\ \tau_\epsilon m_A \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & (M\tau_\epsilon - \tau_\epsilon m_A)\Lambda_A & 1 \\ 0 & \frac{\tau_\epsilon m_A (2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}{M\tau_\epsilon} \Lambda_A & 0 \end{bmatrix} \begin{bmatrix} \hat{\Phi}^Y_{B_0} \\ \hat{\Phi}^Y_{B_1} \\ \hat{\Phi}^Y_{B_2} \end{bmatrix}. \end{aligned} \tag{A.24}$$

For the X-market,

$$X^1_{A,i} = \frac{1}{2}\left(\mu + \begin{bmatrix} \frac{M\tau_\epsilon}{M\tau_\epsilon+\tau_\epsilon m_A+(N\tau_\epsilon+\tau_\theta)} \\ \frac{\tau_\epsilon m_A}{M\tau_\epsilon+\tau_\epsilon m_A+(N\tau_\epsilon+\tau_\theta)} \end{bmatrix} [P^0_A - \mu \quad I_A - \mu]\right). \tag{A.25}$$

■

Proof for Proposition 3

Proof. The expected profit in the X-market is

$$\begin{aligned}
\mathbb{E}\Pi^1_{A,X} =& M\mathbb{E}\left[\mathbb{E}\left[\left(x^1_{A,i}\right)^2 \mid \mathcal{F}_A\right]\right] &\text{(A.26)}\\
=& \frac{M}{4}\left\{\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{(M\tau_\epsilon)^2}{(M\tau_\epsilon+\tau_\epsilon m_A+(N\tau_\epsilon+\tau_\theta))^2}\mathbb{E}\left[\left(P^0_A-\mu\right)^2\right]\right. &\text{(A.27)}\\
&\left.+\frac{\tau_\epsilon^2 m_A^2}{(M\tau_\epsilon+\tau_\epsilon m_A+(N\tau_\epsilon+\tau_\theta))^2}\mathbb{E}\left[(I_A-\mu)^2\right] + \frac{2M\tau_\epsilon^2 m_A}{(M\tau_\epsilon+\tau_\epsilon m_A+(N\tau_\epsilon+\tau_\theta))^2}\mathbb{E}\left[\left(P^0_A-\mu\right)(I_A-\mu)\right]\right.\\
=& \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{M\tau_\epsilon+\tau_\epsilon m_A}{(M\tau_\epsilon+\tau_\epsilon m_A+(N\tau_\epsilon+\tau_\theta))(N\tau_\epsilon+\tau_\theta)}\right).
\end{aligned}$$

Obviously, $\mathbb{E}\Pi^1_{A,X}$ increases in m_A.

The expected profit in the Y-market is

$$\begin{aligned}
\mathbb{E}\Pi^1_{A,Y} &= \frac{1}{N}\mathbb{E}(Y^1_A)^2 &\text{(A.28)}\\
&= \frac{1}{N}\left(\mathbb{E}(\Phi^2_{A_0}) + \Phi^2_{A_1}\mathbb{E}(P^0_A-\mu)^2 + \Phi^2_{A_2}\mathbb{E}(I_A-\mu)^2 + 2\Phi_{A_1}\Phi_{A_2}\mathbb{E}(P^0_A-\mu)(I_A-\mu)\right)\\
&= \frac{1}{N}\mathbb{E}(\Phi^2_{A_0}) + \underbrace{\frac{1}{N\tau^y_\theta}\left((\Phi_{A_1}+\Phi_{A_2})^2 + \Phi^2_{A_1}\frac{(N\tau_\epsilon+\tau_\theta)}{M\tau_\epsilon} + \Phi^2_{A_2}\frac{(N\tau_\epsilon+\tau_\theta)}{\tau_\epsilon m_A}\right)}_{\pi}.
\end{aligned}$$

After direct computation, we obtain

$$\frac{\partial\pi}{\partial m_A} = \frac{\tau_\epsilon(N\tau_\epsilon+\tau_\theta)(\hat{\Phi}_{B_1}(2M\tau_\epsilon+(N\tau_\epsilon+\tau_\theta)) - MN\tau_\epsilon)^2}{4(M\tau_\epsilon)^2(M\tau_\epsilon+\tau_\epsilon m_A+(N\tau_\epsilon+\tau_\theta))^2} > 0. \tag{A.29}$$

Hence, $\mathbb{E}\Pi^1_{A,Y}$ increases in m_A. As $\mathbb{E}\Pi^1_A = \mathbb{E}\Pi^1_{A,X} + \mathbb{E}\Pi^1_{A,Y}$, $\mathbb{E}\Pi^1_A$ increases in m_A.

From above, we know that firms' expected profits increase in both X and Y market, when they purchase more information. This is true regardless of their belief on the opponent's information purchase. From Section **??**, we know that the data vendor will choose m_A^*, m_B^* to maximize $C_A + C_B$. Hence, $m_A^* = M, m_B^* = M$.

Note $C_A^* = \mathbb{E}\Pi_A^1(m_A^* = M, m_B^* = M) - \mathbb{E}\Pi_A^1(0, m_B^* = M)$ satisfies firm A's participation constraints in purchasing information, and hence, is supported in equilibrium. Same argument applies to $C_B^* = \mathbb{E}\Pi_B^1(m_A^* = M, m_B^* = M) - \mathbb{E}\Pi_B^1(0, m_B^* = M)$.

To solve for C_A^*, C_B^*, we substitute $m_A = M, m_B = M$ into production policies in Proposition 2 and compute $\mathbb{E}\Pi_A^1(M, M)$,

$$\begin{aligned}\mathbb{E}\Pi_A^1(M, M) =& \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon + \tau_\theta)\tau_\theta} + \frac{2M\tau_\epsilon}{(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(N\tau_\epsilon + \tau_\theta)}\right) \\ &+ \frac{N}{9}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon + \tau_\theta)\tau_\theta} + \frac{2M\tau_\epsilon}{(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(N\tau_\epsilon + \tau_\theta)}\right).\end{aligned} \tag{A.30}$$

Then, we repeat these computation for $m_A = 0, m_B = M$ and obtain the product policy

$$\begin{bmatrix}\Phi_{A_0}^Y \\ \Phi_{A_1}^Y \\ \Phi_{A_2}^Y\end{bmatrix} = \begin{bmatrix}\frac{N\mu}{3} \\ \frac{MN\tau_\epsilon}{3(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)} \\ 0\end{bmatrix}. \tag{A.31}$$

Plugging these to the profit computation, we obtain $\mathbb{E}\Pi_A^1(0, M)$,

$$\begin{aligned}\mathbb{E}\Pi_A^1(0, M) =& \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon + \tau_\theta)\tau_\theta} + \frac{M\tau_\epsilon}{(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(N\tau_\epsilon + \tau_\theta)}\right) \\ &+ \frac{N}{9}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon + \tau_\theta)\tau_\theta} + \frac{M\tau_\epsilon}{(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(N\tau_\epsilon + \tau_\theta)}\right).\end{aligned} \tag{A.32}$$

At last, we obtain

$$C_A^* = C_B^* = \mathbb{E}\Pi_A^1(M, M) - \mathbb{E}\Pi_A^1(0, M) = \left(\frac{M}{4} + \frac{N}{9}\right)\frac{M\tau_\epsilon}{(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(N\tau_\epsilon + \tau_\theta)}. \tag{A.33}$$

■

Proof for Proposition 4

Proof. According to Definition 1, the equilibrium with the data vendor is obtained by combining Proposition 2 and 3. More specifically, let $m_A = M, m_B = M$ in Proposition 2. ■

Proof for Proposition 6

Proof. Since the equilibrium of the economy with the data vendor reaches full information sharing and the equilibrium of the economy without the data vendor is no information sharing, we know from Proposition 1, $CS(\mathbf{M}, \mathbf{M}) > CS(\emptyset, \emptyset)$ and $TS(\mathbf{M}, \mathbf{M}) > TS(\emptyset, \emptyset)$. ■

Proof for Proposition 7

Proof. The change in the expected profit for firm A when comparing (M, M) and $(0, 0)$ is

$$\begin{aligned}
&\mathbb{E}\Pi_A^1(M,M) - C_A(M,M) - \mathbb{E}\Pi_A^1(0,0) \qquad \text{(A.34)}\\
&= (1-\beta)\left(\frac{M}{4} + \frac{N}{9}\right)\frac{M\tau_\epsilon}{(2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))(M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))} + \mathbb{E}\Pi_{A,Y}^1(0,M) - \mathbb{E}\Pi_{A,Y}^1(0,0)\\
&= (1-\beta)\left(\frac{M}{4} + \frac{N}{9}\right)\frac{M\tau_\epsilon}{(2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))(M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}\\
&\quad - \frac{MN\tau_\epsilon(6M\tau_\epsilon + 5(N\tau_\epsilon + \tau_\theta))}{9(M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))(3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta))^2},
\end{aligned}$$

which is greater than 0, if $M > \hat{M}$, and if

$$1 - \beta > \left(\frac{M}{4} + \frac{N}{9}\right)^{-1}\frac{N(6M\tau_\epsilon + 5(N\tau_\epsilon + \tau_\theta))(2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}{9(3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta))^2}. \qquad \text{(A.35)}$$

■

Proof for Proposition 8

Proof. By symmetry, we focus only on firm A. To ensure supplying information is a dominant strategy, we need following condition

$$\text{Given } B \text{ does not supply, } A \text{ will supply} \Rightarrow \mathbb{E}\Pi_A^1(\emptyset,\emptyset) \le \mathbb{E}\Pi_A^1(\emptyset,\mathbf{M}) + t_B, \tag{A.36}$$

and

$$\text{Given } B \text{ supplies, } A \text{ will supply} \Rightarrow \mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) \le \mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}) + t_B. \tag{A.37}$$

Hence, if

$$\max\left\{\mathbb{E}\Pi_A^1(\emptyset,\emptyset) - \mathbb{E}\Pi_A^1(\emptyset,\mathbf{M}), \mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) - \mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M})\right\} \le t_B, \tag{A.38}$$

then both conditions are met. In the online appendix we compute the folliwng

$$\mathbb{E}\Pi_A^1(\emptyset,\emptyset) - \mathbb{E}\Pi_A^1(\emptyset,\mathbf{M}) = \frac{MN\tau_\epsilon\left(6M\tau_\epsilon + 5(N\tau_\epsilon+\tau_\theta)\right)}{9\left(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(3M\tau_\epsilon + 2(N\tau_\epsilon+\tau_\theta)\right)^2}, \tag{A.39}$$

$$\mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) - \mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}) = \frac{5MN\tau_\epsilon}{36\left(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)}. \tag{A.40}$$

After computation, we get

$$\begin{aligned}&\left[\mathbb{E}\Pi_A^1(\emptyset,\emptyset) - \mathbb{E}\Pi_A^1(\emptyset,\mathbf{M})\right] - \left[\mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) - \mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M})\right] \\ &= \frac{M^2N\tau_\epsilon^2\left(3M\tau_\epsilon + 4(N\tau_\epsilon+\tau_\theta)\right)}{36\left(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(3M\tau_\epsilon + 2(N\tau_\epsilon+\tau_\theta)\right)^2} > 0.\end{aligned} \tag{A.41}$$

Therefore, by symmetry, we have

$$\frac{MN\tau_\epsilon\left(6M\tau_\epsilon + 5(N\tau_\epsilon+\tau_\theta)\right)}{9\left(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(3M\tau_\epsilon + 2(N\tau_\epsilon+\tau_\theta)\right)^2} \le t_A, t_B. \tag{A.42}$$

To ensure demanding information is a dominant strategy, we need following conditions,

$$\text{Given } B \text{ does not demand, } A \text{ will demand} \Rightarrow \mathbb{E}\Pi_A^1(\emptyset,\emptyset) \leq \mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) - t_A, \tag{A.43}$$

and

$$\text{Given } B \text{ demands, } A \text{ will demand} \Rightarrow \mathbb{E}\Pi_A^1(\emptyset,\mathbf{M}) \leq \mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}) - t_A. \tag{A.44}$$

Hence, if

$$t_A \leq \min\left\{\mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) - \mathbb{E}\Pi_A^1(\emptyset,\emptyset), \mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}) - \mathbb{E}\Pi_A^1(\emptyset,\mathbf{M})\right\}, \tag{A.45}$$

then both conditions are met. From computation, we have

$$\mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) - \mathbb{E}\Pi_A^1(\emptyset,\emptyset) = \frac{M\tau_\epsilon\Theta}{36\left(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta\right)\left(2M\tau_\epsilon+N\tau_\epsilon+\tau_\theta\right)\left(3M\tau_\epsilon+2(N\tau_\epsilon+\tau_\theta)\right)^2}, \tag{A.46}$$

$$\mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}) - \mathbb{E}\Pi_A^1(\emptyset,\mathbf{M}) = \frac{M\tau_\epsilon\left(9M+4N\right)}{36\left(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta\right)\left(2M\tau_\epsilon+N\tau_\epsilon+\tau_\theta\right)}. \tag{A.47}$$

where

$$\Theta = 81M^3\tau_\epsilon^2 + 33M^2N\tau_\epsilon^2 + 108M^2\tau_\epsilon(N\tau_\epsilon+\tau_\theta) + 44MN\tau_\epsilon(N\tau_\epsilon+\tau_\theta) + 36M(N\tau_\epsilon+\tau_\theta)^2 + 16N(N\tau_\epsilon+\tau_\theta)^2 \tag{A.48}$$

Direct computation shows that

$$\begin{aligned}&\left[\mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) - \mathbb{E}\Pi_A^1(\emptyset,\emptyset)\right] - \left[\mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}) - \mathbb{E}\Pi_A^1(\emptyset,\mathbf{M})\right] \\ &= -\frac{M^2N\tau_\epsilon^2\left(3M\tau_\epsilon+4(N\tau_\epsilon+\tau_\theta)\right)}{36\left(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta\right)\left(2M\tau_\epsilon+N\tau_\epsilon+\tau_\theta\right)\left(3M\tau_\epsilon+2(N\tau_\epsilon+\tau_\theta)\right)^2} < 0.\end{aligned} \tag{A.49}$$

Therefore, by symmetry, we have

$$t_A, t_B \leq \frac{M\tau_\epsilon \Theta}{36\left(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta)\right)^2}. \tag{A.50}$$

Since

$$\begin{aligned} &\overbrace{\left[\mathbb{E}\Pi_A^1(\mathbf{M}, \emptyset) - \mathbb{E}\Pi_A^1(\emptyset, \emptyset)\right]}^{\text{upper bound}} - \overbrace{\left[\mathbb{E}\Pi_A^1(\emptyset, \emptyset) - \mathbb{E}\Pi_A^1(\emptyset, \mathbf{M})\right]}^{\text{lower bound}} \\ &= \frac{M\tau_\epsilon \Theta'}{36\left(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta\right)\left(3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta)\right)^2} \end{aligned} \tag{A.51}$$

where

$$\Theta' = 81M^3\tau_\epsilon^2 - 15M^2N\tau_\epsilon^2 + 108M^2\tau_\epsilon(N\tau_\epsilon + \tau_\theta) - 20MN\tau_\epsilon(N\tau_\epsilon + \tau_\theta) + 36M(N\tau_\epsilon + \tau_\theta)^2 - 4N(N\tau_\epsilon + \tau_\theta)^2 \tag{A.52}$$

Θ' is larger than 0 when $M \to \infty$. The set of transfers is not empty for large M. ■

References

Acquisti, Alessandro, 2015, From the economics of privacy to the economics of big data, *Working Paper* pp. 1–33.

Admati, Anat R., and Paul Pfleiderer, 1986, A monopolistic market for information, *Journal of Economic Theory* 39, 400–438.

———, 1987, Viable allocations of information in financial markets, *Journal of Economic Theory* 43, 76–115.

———, 1988, Selling and trading on information in financial markets, *American Economic Review* 78, 96–103.

Bagnoli, Mark, and Susan G. Watts, 2015, Competitive intelligence and disclosure, *RAND Journal of Economics* 46, 709–729.

Bailey, Warren, Haitao Li, Connie X. Mao, and Rui Zhong, 2003, Regulation fair disclosure and earnings information: Market, analyst, and corporate responses, *Journal of Finance* 58, 2487–2514.

Bushee, Brian J., Dawn A. Matsumoto, and Gregory S. Miller, 2004, Managerial and investor responses to disclosure regulation: The case of Reg FD and conference calls, *Accounting Review* 79, 617–643.

Cohen, Lauren, Dong Lou, and Christopher Malloy, 2017, Playing favorites: How firms prevent the revelation of bad news, *Working Paper*.

Cong, Lin William, and Zhiguo He, 2018, Blockchain disruption and smart contracts, *Working Paper*.

Darrough, Masako N., 1993, Disclosure policy and competition: Cournot vs. Bertrand, *Accounting Review* 68, 534–561.

Gal-Or, Esther, 1985, Information sharing in oligopoly, *Econometrica* 53, 329–343.

——— , 1986, Information transmission – Cournot and Bertrand equilibria, *Review of Economic Studies* 53, 85–92.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely, 2017, Are U.S. industries becoming more concentrated?, *Working Paper*.

Hauk, Esther, and Sjaak Hurkens, 2001, Secret information acquisition in Cournot markets, *Economic Theory* 18, 661–681.

Hwang, Hae-shin, 1993, Optimal information acquisition for heterogenous duopoly firms, *Journal of Economic Theory* 59, 385–402.

Li, Jiasun, and William Mann, 2018, Initial coin offering and platform building, *Working Paper*.

Li, Lode, Richard D McKelvey, and Talbot Page, 1987, Optimal research for cournot oligopolists, *Journal of Economic Theory* 42, 140–166.

Raith, Michael, 1996, A general model of information sharing in oligopoly, *Journal of Economic Theory* 71, 260–288.

Singh, Nirvikar, and Xavier Vives, 1984, Price and quantity competition in a differentiated duopoly, *RAND Journal of Economics* 15, 546–554.

Vives, Xavier, 1984, Duopoly information equilibrium: Cournot and Bertrand, *Journal of Economic Theory* 34, 71–94.

——— , 1988, Aggregation of information in large Cournot markets, *Econometrica* 56, 851.

——— , 2016, Information sharing among firms, in *The New Palgrave Dictionary of Economics* . pp. 1–4 (Palgrave Macmillan UK: London).

Online appendix

Proof for Lemma 3 and Proposition 1

Proof. To proof Lemma 3 and Proposition 1, we need to compute firms expected profits for the new information allocation space of $\{\emptyset, \mathbf{M}\}\} \times \{\emptyset, \mathbf{M}\}$

Given any information allocation, a firm's optimal production decision is to maximize her profits in the X-market and the Y-market. To keep our discussion succinct, we only illustrate the production decision of firm A. Similar argument can be applied to firm B. Combining the market clearing condition ($x^1_{A,i} = X^1_{A,i}, \sum_{i=1}^N y^1_{AB,i} = Y^1_A + Y^1_B$) which implies $P^1_{A,i} = \tilde{s}^1_{A,i} - X^1_{A,i}, p^1_y = \frac{\sum_{i=1}^N \tilde{s}^1_{AB,i} - Y^1_A - Y^1_B}{N}$, firm A's production decision is

$$\max_{x^1_{A,i}} \mathbb{E}[P^1_{A,i} X^1_{A,i} \mid \mathcal{F}_A] = \max_{x^1_{A,i}} \mathbb{E}\left[\left(\tilde{s}^1_{A,i} - X^1_{A,i}\right) X^1_{A,i} \mid \mathcal{F}_A\right], \text{for } i = 1, 2, ..., M; \tag{OA.1}$$

$$\max_{Y^1_A} \mathbb{E}\left[\left(\frac{\sum_{i=1}^N \tilde{s}^1_{AB,i} - Y^1_A - Y^1_B}{N}\right) Y^1_A \mid \mathcal{F}_A\right]. \tag{OA.2}$$

The optimal production is

$$X^1_{A,i} = \frac{1}{2}\mathbb{E}\left(\tilde{\theta} \mid \mathcal{F}_A\right), \tag{OA.3}$$

$$Y^1_A = \frac{N}{2}\mathbb{E}(\tilde{\theta} \mid \mathcal{F}_A) - \frac{1}{2}\mathbb{E}(Y^1_B \mid \mathcal{F}_A). \tag{OA.4}$$

Plugging the optimal production to the expected profit function, we get the total expected profit in the X-market is

$$\mathbb{E}\Pi^1_{A,X} = M\mathbb{E}\left[\mathbb{E}\left[\left(X^1_{A,i}\right)^2 \mid \mathcal{F}_A\right]\right]. \tag{OA.5}$$

The total expected profit in the Y-market is

$$\mathbb{E}\Pi^1_{A,Y} = \frac{1}{N}\mathbb{E}\left[\mathbb{E}\left[\left(Y^1_A\right)^2 \mid \mathcal{F}_A\right]\right]. \tag{OA.6}$$

In above equations, $\mathcal{F}_A$ depends firm A's endowned information $(\mathbf{P}_A^0, p_y^0)$ and the increment information $\Delta\mathcal{F}_A$ from the new information allocation.

1. $\Delta\mathcal{F}_A = \emptyset, \Delta\mathcal{F}_B = \emptyset$.

   In this case, neither firm A nor B obtains new information. Since A and B are symmetric in this case, we only illustrate the equilibrium production and profits for firm A. Since $P_{A,i}^0 \in \mathbf{P}_A^0$ equals $\tilde{s}_{A,i}^0$ which follows an iid Normal for $i = 1, 2, ..., M$. We can summarize the information set $\mathbf{P}_A^0$ with with $P_A^0 = \frac{\sum_{P_{A,i}^0 \in \mathbf{P}_A^0} P_{A,i}^0}{M}$ in Bayesian updating. P_A^0 is essentially the average price across all A-type consumers. With the similar argument, we use $P_B^0 = \frac{\sum_{P_{B,i}^0 \in \mathbf{P}_B^0} P_{B,i}^0}{M}$ to summarize the information set $\mathbf{P}_B^0$. P_B^0 is essentially the average price across all B-type consumers. Hence, we have the following from Bayesian updating

$$\mathbb{E}(\tilde{\theta} \mid \mathcal{F}_A) = \mu + \frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu), \tag{OA.7}$$

$$\mathbb{E}(P_B^0 \mid \mathcal{F}_A) = \mu + \frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu), \tag{OA.8}$$

   where $\mu = \frac{N\tau_\epsilon}{N\tau_\epsilon + \tau_\theta} p_y^0$. Substituting these into the optimal production choice, and combining with the following conjecture of linear strategy

$$Y_A^1 = \Phi_{A_0} + \Phi_{A_1}(P_A^0 - \mu), \tag{OA.9}$$

   we have optimal production in the Y-mraket as follows

$$Y_A^1 = \frac{N\mu}{3} + \frac{MN\tau_\epsilon}{3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta)}(P_A^0 - \mu). \tag{OA.10}$$

   Symmetrically, we can solve for Y_B^1, which is

$$Y_B^1 = \frac{N\mu}{3} + \frac{MN\tau_\epsilon}{3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta)}(P_B^0 - \mu). \tag{OA.11}$$

In the X-market, we have

$$X_{A,i}^1 = \frac{1}{2}\left(\mu + \frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu)\right), \tag{OA.12}$$

$$X_{B,i}^1 = \frac{1}{2}\left(\mu + \frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_B^0 - \mu)\right). \tag{OA.13}$$

The total expected profit in the Y-market and X-market is

$$\begin{aligned}\mathbb{E}\Pi_{A,Y}^1(\emptyset,\emptyset) &= \frac{1}{N}\mathbb{E}\left[\frac{N\mu}{3} + \frac{MN\tau_\epsilon}{3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta)}(P_A^0 - \mu)\right]^2 \qquad \text{(OA.14)}\\ &= \frac{N}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon + \tau_\theta)\tau_\theta} + \frac{MN\tau_\epsilon(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}{(3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta))^2(N\tau_\epsilon + \tau_\theta)},\end{aligned}$$

$$\mathbb{E}\Pi_{A,X}^1(\emptyset,\emptyset) = \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon + \tau_\theta)\tau_\theta} + \frac{M\tau_\epsilon}{(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(N\tau_\epsilon + \tau_\theta)}\right). \tag{OA.15}$$

And

$$\mathbb{E}\Pi_A^1(\emptyset,\emptyset) = \mathbb{E}\Pi_{A,X}^1(\emptyset,\emptyset) + \mathbb{E}\Pi_{A,Y}^1(\emptyset,\emptyset) = \mathbb{E}\Pi_B^1(\emptyset,\emptyset). \tag{OA.16}$$

This proves Lemma 3.

2. $\Delta\mathcal{F}_A = \emptyset, \Delta\mathcal{F}_B = \mathbf{M}$.

For the second case, we consider firm A gets no new information but firm B gets all firm A's private information. For firm A, her information set is $\mathcal{F}_A = \left\{P_A^0, p_y^0\right\}$. For firm B, her information set is $\mathcal{F}_B = \left\{P_A^0, P_B^0, p_y^0\right\}$. From Bayesian updating, we have

$$\mathbb{E}(\tilde{\theta} \mid \mathcal{F}_A) = \mu + \frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu), \tag{OA.17}$$

$$\mathbb{E}(P_B^0 \mid \mathcal{F}_A) = \mu + \frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu), \tag{OA.18}$$

$$\mathbb{E}(\tilde{\theta} \mid \mathcal{F}_B) = \mu + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_B^0 - \mu) + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu) \tag{OA.19}$$

$$\mathbb{E}(P_A^0 \mid \mathcal{F}_B) = P_A^0. \tag{OA.20}$$

We conjecture the linear strategy

$$Y_A^1 = \Phi_{A_0} + \Phi_{A_1}(P_A^0 - \mu), \tag{OA.21}$$

$$Y_B^1 = \Phi_{B_0} + \Phi_{B_1}(P_B^0 - \mu) + \Phi_{B_2}(P_A^0 - \mu). \tag{OA.22}$$

Combining the conjecture linear strategy with the optimal production decision, we have

$$Y_A^1 = \frac{N}{2}\left(\mu + \frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu)\right) - \frac{1}{2}\left(\Phi_{B_0} + \Phi_{B_1}\frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu) + \Phi_{B_2}(P_A^0 - \mu)\right) \tag{OA.23}$$

$$Y_B^1 = \frac{N}{2}\left(\mu + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_B^0 - \mu) + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu)\right) - \frac{1}{2}\left(\Phi_{A_0} + \Phi_{A_1}(P_A^0 - \mu)\right). \tag{OA.24}$$

After computation, we have

$$Y_A^1 = \frac{N\mu}{3} + \frac{MN\tau_\epsilon}{3(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}(P_A^0 - \mu), \tag{OA.25}$$

$$Y_B^1 = \frac{N\mu}{3} + \frac{MN\tau_\epsilon}{2(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}(P_A^0 - \mu) + \frac{MN\tau_\epsilon(M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta))}{6(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}(P_B^0 - \mu). \tag{OA.26}$$

In the X-market, we have

$$X_{A,i}^1 = \frac{1}{2}\left(\mu + \frac{M\tau_\epsilon}{M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu)\right), \tag{OA.27}$$

$$X_{B,i}^1 = \frac{1}{2}\left(\mu + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu) + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_B^0 - \mu)\right). \tag{OA.28}$$

The expected profit in the Y-market and the X-market is

$$\mathbb{E}\Pi^1_{A,Y}(\emptyset,\mathbf{M}) = \frac{N}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{MN\tau_\epsilon}{9(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}, \tag{OA.29}$$

$$\mathbb{E}\Pi^1_{B,Y}(\emptyset,\mathbf{M}) = \frac{N}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{MN\tau_\epsilon(8M\tau_\epsilon+13(N\tau_\epsilon+\tau_\theta))}{36(2M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}, \tag{OA.30}$$

$$\mathbb{E}\Pi^1_{A,X}(\emptyset,\mathbf{M}) = \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{M\tau_\epsilon}{(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}\right), \tag{OA.31}$$

$$\mathbb{E}\Pi^1_{B,X}(\emptyset,\mathbf{M}) = \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{2M\tau_\epsilon}{(2M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}\right). \tag{OA.32}$$

And

$$\mathbb{E}\Pi^1_{A}(\emptyset,\mathbf{M}) = \mathbb{E}\Pi^1_{A,X}(\emptyset,\mathbf{M}) + \mathbb{E}\Pi^1_{A,Y}(\emptyset,\mathbf{M}), \tag{OA.33}$$

$$\mathbb{E}\Pi^1_{B}(\emptyset,\mathbf{M}) = \mathbb{E}\Pi^1_{B,X}(\emptyset,\mathbf{M}) + \mathbb{E}\Pi^1_{B,Y}(\emptyset,\mathbf{M}). \tag{OA.34}$$

3. $\Delta\mathcal{F}_A = \mathbf{M}, \Delta\mathcal{F}_B = \emptyset$.

Case III is symmetric to Case II. Hence, we have The expected profit in the Y-market and the X-market is

$$\mathbb{E}\Pi^1_{A,Y}(\mathbf{M},\emptyset) = \frac{N}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{MN\tau_\epsilon(8M\tau_\epsilon+13(N\tau_\epsilon+\tau_\theta))}{36(2M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}, \tag{OA.35}$$

$$\mathbb{E}\Pi^1_{B,Y}(\mathbf{M},\emptyset) = \frac{N}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{MN\tau_\epsilon}{9(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}, \tag{OA.36}$$

$$\mathbb{E}\Pi^1_{A,X}(\emptyset) = \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{2M\tau_\epsilon}{(2M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}\right), \tag{OA.37}$$

$$\mathbb{E}\Pi^1_{B,X}(\mathbf{M},\emptyset) = \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon+\tau_\theta)\tau_\theta} + \frac{M\tau_\epsilon}{(M\tau_\epsilon+N\tau_\epsilon+\tau_\theta)(N\tau_\epsilon+\tau_\theta)}\right). \tag{OA.38}$$

And

$$\mathbb{E}\Pi_A^1(\mathbf{M}, \emptyset) = \mathbb{E}\Pi_{A,X}^1(\mathbf{M}, \emptyset) + \mathbb{E}\Pi_{A,Y}^1(\mathbf{M}, \emptyset), \tag{OA.39}$$

$$\mathbb{E}\Pi_B^1(\mathbf{M}, \emptyset) = \mathbb{E}\Pi_{B,X}^1(\mathbf{M}, \emptyset) + \mathbb{E}\Pi_{B,Y}^1(\mathbf{M}, \emptyset). \tag{OA.40}$$

4. $\Delta\mathcal{F}_A = \mathbf{M}, \Delta\mathcal{F}_B = \mathbf{M}$.

In the final case, both firm A and B gets all new information. Since firm A and firm B are symmetric in this case, we only illustrate the optimal choice of firm A. For firm A, her information set is now $\mathcal{F}_A = \left\{P_A^0, p_y^0, P_B^0\right\}$. Based on Bayesian updating, we have

$$\mathbb{E}(\tilde{\theta} \mid \mathcal{F}_A) = \mu + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_A^0 - \mu) + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P_B^0 - \mu), \tag{OA.41}$$

$$\mathbb{E}(P_B^0 \mid \mathcal{F}_A) = P_B^0. \tag{OA.42}$$

Plugging this into the optimal production choice, and combining with the conjecture of linear strategy

$$Y_A^1 = \Phi_{A_0} + \Phi_{A_1}(P_A^0 - \mu) + \Phi_{A_2}(P_B^0 - \mu) \tag{OA.43}$$

$$Y_B^1 = \Phi_{B_0} + \Phi_{B_1}(P_B^0 - \mu) + \Phi_{B_2}(P_A^0 - \mu). \tag{OA.44}$$

We have optimal production in the Y-mraket,

$$Y_A^1 = \frac{N\mu}{3} + \frac{MN\tau_\epsilon}{3(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}(P_A^0 - \mu) + \frac{MN\tau_\epsilon}{3(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}(P_B^0 - \mu), \tag{OA.45}$$

$$Y_B^1 = \frac{N\mu}{3} + \frac{MN\tau_\epsilon}{3(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}(P_A^0 - \mu) + \frac{MN\tau_\epsilon}{3(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}(P_B^0 - \mu). \tag{OA.46}$$

In the X-market, we have

$$X^1_{A,i} = X^1_{B,i} = \frac{1}{2}\left(\mu + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P^0_A - \mu) + \frac{M\tau_\epsilon}{2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta}(P^0_B - \mu)\right). \tag{OA.47}$$

The total expected profit in the Y-market and the X-market is

$$\mathbb{E}\Pi^1_{A,Y}(\mathbf{M},\mathbf{M}) = \frac{N}{9}\frac{N\tau_\epsilon}{(N\tau_\epsilon + \tau_\theta)\tau_\theta} + \frac{2MN\tau_\epsilon}{9(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(N\tau_\epsilon + \tau_\theta)}, \tag{OA.48}$$

$$\mathbb{E}\Pi^1_{A,X}(\mathbf{M},\mathbf{M}) = \frac{M}{4}\left(\frac{N\tau_\epsilon}{(N\tau_\epsilon + \tau_\theta)\tau_\theta} + \frac{2M\tau_\epsilon}{(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(N\tau_\epsilon + \tau_\theta)}\right). \tag{OA.49}$$

And

$$\mathbb{E}\Pi^1_A(\mathbf{M},\mathbf{M}) = \mathbb{E}\Pi^1_{A,X}(\mathbf{M},\mathbf{M}) + \mathbb{E}\Pi^1_{A,Y}(\mathbf{M},\mathbf{M}) = \mathbb{E}\Pi^1_B(\mathbf{M},\mathbf{M}). \tag{OA.50}$$

Since

$$\mathbb{E}\Pi^1_A(\mathbf{M},\mathbf{M}) - \mathbb{E}\Pi^1_A(\emptyset,\emptyset) \tag{OA.51}$$

$$= \frac{M^2\tau_\epsilon}{4(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)} - \frac{MN\tau_\epsilon(3M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)}{9(2M\tau_\epsilon + N\tau_\epsilon + \tau_\theta)(3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta))^2}$$

$$= \frac{M\tau_\epsilon}{36(2M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))(M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}\frac{1}{(3M\tau_\epsilon + 2(N\tau_\epsilon + \tau_\theta))^2}$$

$$\times (M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))(3M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta)) \tag{OA.52}$$

$$\times \underbrace{\left[9\frac{M}{(M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta))}\left(3M\tau_\epsilon + (N\tau_\epsilon + \tau_\theta) + 2(N\tau_\epsilon + \tau_\theta) + \frac{(N\tau_\epsilon + \tau_\theta)^2}{3M\tau_\epsilon + \tau_y}\right) - 4N\right]}_{g(M)}, \tag{OA.53}$$

it is clear that the sign of $\mathbb{E}\Pi^1_A(\mathbf{M},\mathbf{M}) - \mathbb{E}\Pi^1_A(\emptyset,\emptyset)$ depends only on $g(M)$.

We find that

$$g(0) = -4N < 0, g(+\infty) > 0, \tag{OA.54}$$

and

$$\frac{\partial g}{\partial M} = 9\tfrac{(N\tau_\epsilon+\tau_\theta)}{(M\tau_\epsilon+\tau_y)^2}\left(3M\tau_\epsilon + (N\tau_\epsilon+\tau_\theta) + 2(N\tau_\epsilon+\tau_\theta) + \tfrac{(N\tau_\epsilon+\tau_\theta)^2}{3M\tau_\epsilon+(N\tau_\epsilon+\tau_\theta)}\right) \tag{OA.55}$$
$$+9\tfrac{M}{(M\tau_\epsilon+(N\tau_\epsilon+\tau_\theta))}\left(3\tau_\epsilon - \tfrac{3\tau_\epsilon(N\tau_\epsilon+\tau_\theta)^2}{(3M\tau_\epsilon+(N\tau_\epsilon+\tau_\theta))^2}\right) > 0.$$

Thus there exists one unique solution of $g(M) = 0$. We denote the solution as $\hat{M}$, i.e.,

$$\left[9\frac{\hat{M}}{(\hat{M}\tau_\epsilon + (N\tau_\epsilon+\tau_\theta))}\left(3\hat{M}\tau_\epsilon + (N\tau_\epsilon+\tau_\theta) + 2(N\tau_\epsilon+\tau_\theta) + \frac{(N\tau_\epsilon+\tau_\theta)^2}{3\hat{M}\tau_\epsilon + N\tau_\epsilon+\tau_\theta}\right) - 4N\right] = 0. \tag{OA.56}$$

We have when $M > \hat{M}$, $g(M) > 0$; and when $M < \hat{M}$, $g(M) < 0$. This suggest that when $M > \hat{M}$, $\mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}) - \mathbb{E}\Pi_A^1(\emptyset,\emptyset) > 0$.

This proves $\hat{M}$ in Proposition 1.

Direct computation shows that

$$\mathbb{E}\Pi_A^1(\mathbf{M},\emptyset) > \mathbb{E}\Pi_A^1(\emptyset,\emptyset), \mathbb{E}\Pi_A^1(\mathbf{M},\mathbf{M}). \tag{OA.57}$$

Thus, the unique pure strategy Nash equilibrium is neither firm chooses to share information, i.e., $\Delta\mathcal{F}_A = \Delta\mathcal{F}_B = \emptyset$. This proves the Nash equilibrium in Table 3. ■