Competitive Markets for Personal Data

Simone Galperti — UC, San Diego
Tianhao Liu — Columbia University
Jacopo Perego — Columbia University

August 20, 2023

Abstract

Consumers supply a crucial input for the modern economy: their personal data. Yet, they often have limited control over who uses it and are imperfectly compensated in return. This status quo can lead to inefficiencies. Could a competitive market for personal data do better? We study a stylized competitive economy where consumers own their data and can sell it to a platform. The platform then uses this data to interact the corresponding consumers with a third-party merchant, from whom they can buy a product. We find that, despite its competitive nature, this economy is inefficient. The market failure stems from consumers exerting an externality on each other when selling their data, which is enabled by how the platform optimally uses this data. We propose two solutions to this inefficiency. The first one introduces a "data union," which manages consumers' data on their behalf and compensates them accordingly. The second one envisions trading in markets that determine not only who gets the data but also how this data is used.

JEL Classification Numbers: C72, D82, D83

Keywords: Personal Data, Markets, Competitive Equilibrium, Externalities

We are thankful to Nima Haghpanah, Dana Foarta, Matthew Gentzkow, Paolo Siconolfi, Philipp Strack, Alexis Akira Toda for helpful comments. This research is supported by grants from the NSF (Galperti: SES-2149289; Perego: SES-2149315). Galperti also gratefully acknowledges financial support from the UPenn's CTIC and the Warren Center for Network & Data Sciences.

1 Introduction

Consumer data has become a crucial input into the modern economy. For instance, this data allows marketers to access consumers and understand their preferences, thereby fueling the large online-advertising industry along with numerous other multi-billion-dollar industries. Since each consumer is the primary supplier of data about herself, one could expect that she would control when and how her data is supplied, and receive appropriate compensation when this happens. In practice, however, consumers typically have minimal control over their data and are at best only indirectly compensated for it—for example, through mechanisms like barter.[1]

Such an arrangement, which appears far from ideal market conditions, could harbor inefficiencies and lead to market failure (Seim et al., 2022). For example, if consumers have limited control over their data, they may not be able to avoid the externalities associated with how it is used. Additionally, if consumers are inadequately compensated, they may supply inefficiently low or high quantities of data. This paper delves into these concerns by comparing various designs of a data economy and by evaluating their welfare implications. Our main contribution is to show that—due to the specific ways data is used by platform intermediaries—even a perfectly competitive economy where consumers own their data and can sell it at the market price can be inefficient. We propose two alternative market designs that could correct for this inefficiency.

All the economies we study share a stylized common structure. Consumers own their data and can sell it to a platform at a given price. In addition, the platform provides a service to the consumers who sell their data. It intermediates them with a third-party merchant, from whom they can buy a product. As an intermediary, the platform uses the consumers' data to inform the merchant about their willingness to pay for the product, enabling the merchant to extract surplus from them. Therefore, how the platform uses this data affects the merchant's profits, the consumers' surplus, and, ultimately, the price of data. Holding this basic structure fixed, we consider several economies that differ in how the price of data is determined.

To set a simple benchmark, we begin by studying an economy where the platform is a price maker. It sets prices by making each consumer a take-it-or-leave-it offer to acquire her data. We show that, in any equilibrium of this economy, data are allocated and used in a (constrained) efficient way. That is, the sum of the platform's and consumers' welfare is maximized, subject to the requirement that the platform uses the acquired data in a sequentially rational way. The

[1] As of today, the vast majority of consumers' data is either held as a proprietary asset by large digital platforms, or is traded in opaque brokerage markets with minimal consumer involvement (Federal Trade Commission, 2014).

converse also holds: Any (constrained) efficient data allocation can be supported by an equilibrium of this economy. While efficient, these equilibria feature a welfare distribution that—as expected—is skewed in favor of the platform. Since the platform has all the bargaining power, it appropriates the value created by consumers' data, leaving consumers' welfare minimized.

We then consider a competitive economy where the platform is a price taker. Prices are pinned down by market clearing, as if multiple identical platforms competed to acquire the consumers' data. We show that the (constrained) efficiency of this competitive economy crucially relies on the platform's objective. Specifically, when the platform's and the merchant's objectives are sufficiently aligned, the equilibrium allocation is (constrained) efficient and consumers' welfare is maximized. By contrast—and perhaps counterintuitively—when the platform's objective is sufficiently aligned with that of the consumers, the equilibrium can be inefficient and consumers' welfare can be as low as in the benchmark economy of the previous paragraph.

The inefficiency of this competitive economy stems from the fact that consumers exert an externality on each other when selling their data. The externality arises endogenously from the way the platform uses the data (Galperti, Levkun, and Perego, 2023). For this reason, the platform's objective is critical to the efficiency of the economy. More specifically, when the platform cares too much about consumers' surplus, it will tend to withhold some information from the merchant to prevent surplus extraction. To do so, the platform needs to pool consumers of different types. In such a way, the merchant cannot deduce with certainty the type of each consumer in the pool. The composition of the pool determines the beliefs of the merchant. If a consumer refuses to join the pool—e.g., by deciding not to sell her data—her choice affects the beliefs of the merchant and, thus, the welfare of all consumers in the pool.

We then discuss two alternative market designs that correct the aforementioned inefficiency of the competitive economy: one relies on regulation, the other on more-complete markets.

In the first design, we introduce a "data union" into the economy, which we model as follows.[2] Consumers voluntarily relinquish their data to the union, which manages it on their behalf. The data union then sells part of it to the platform and returns all the proceeds from the sale to the consumers. The compensation a consumer receives needs to be sufficiently high to prevent this consumer from leaving the union. We show that any equilibrium of the data-union economy is (constrained) efficient and consumers' welfare is maximized, regardless of the platform's objective.

The second alternative design is inspired by classic models of competitive economies with

[2] To the best of our knowledge, the idea of a data union was first discussed in Posner and Weyl (2018).

externalities (e.g., Arrow (1969) and Laffont (1976)). We consider a "Lindahl" economy in which the consumers and the platform trade not only the data ownership but also how the data is used.[3] We show that equilibria of this economy are (unconstrained) efficient and consumers' welfare is maximized, regardless of the objective of the platform. The converse also holds: Any (unconstrained) efficient data allocation can be supported as an equilibrium of the Lindahl economy. From a methodological perspective, these equilibria can be conveniently characterized as solutions of a grand information-design problem.

Related Literature. This paper contributes to the literature on data markets (for a review, see Acquisti et al. (2016), Bergemann and Bonatti (2019), and Bergemann and Ottaviani (2021)). Our model of a competitive data market is rooted in a general-equilibrium tradition but leverages the recent progress of the information-design literature (for a review, see Bergemann and Morris (2019) and Kamenica (2019)). We model the interaction between the platform and the merchant in a stylized way, following the steps of Bergemann et al. (2015) and Galperti et al. (2023). The key modeling choice is that the platform is an intermediary and, thus, may have incentives to withhold information from the merchant.[4]

Our main focus is on the upstream market where the platform acquires consumers' data. The basic structure of our model borrows elements from Choi et al. (2019), Ichihashi (2021), Acemoglu et al. (2022), and Bergemann et al. (2022): The platform buys data from consumers and uses it to provide information to a merchant, who in turn sells a product to the consumers. In contrast to these papers, our focus is on competitive markets and our goal is to emphasize a novel market failure. This failure does not arise from exogenous correlation in consumers' data but, rather, from how the platform endogenously uses this data. The source of inefficiency is a "pooling" externality originally introduced by Galperti, Levkun, and Perego (2023). Our contribution is to build a tractable competitive economy around this idea, identify conditions under which this externality can indeed lead to market failure, and propose solutions to it.

[3] In a similar spirit, the European Union's GDPR requires that "the specific purposes for which personal data are processed should be [...] determined at the time of the collection" (Regulation 2016/679 (39)).

[4] Bergemann et al. (2018), Yang (2022), Bonatti et al. (2022), and Bonatti and Bergemann (2023), for instance, study richer interactions where the platform optimally designs information and sells it to a merchant, possibly under incomplete information.

2 Model

This section introduces the basic ingredients of our model, which builds on Bergemann et al. (2015) and Galperti et al. (2023). We consider a platform (*it*), a merchant (*he*), and a unit mass of consumers (*she*).

The merchant produces widgets at zero marginal cost and wants to sell them to the consumers. Each consumer has unit demand for the widget and we denote her willingness to pay by $\omega \in \Omega$. Let Ω be a finite subset of the positive real line and let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω. Each consumer owns a *data record* that consists of a verifiable list of her personal characteristics (gender, age, etc.) and identifiers (IP address, telephone number, etc.). The data record carries information about the consumer's ω and also provides access to her. To keep the notation simple, let us assume that the data record fully reveals the consumer's ω.[5]

The model has two periods. In the first period, the platform and the consumers trade the data records. On the supply side, given a vector of prices $p = (p(\omega))_{\omega\in\Omega} \in \mathbb{R}^{\Omega}$, each type-$\omega$ consumer decides whether to sell her record to the platform. Denote by $\zeta(\omega) \in [0, 1]$ the probability that she sells her record. Without loss of generality, we assume that consumers of the same type behave symmetrically. Thus, $\zeta(\omega)\bar{q}(\omega)$ is the total supply of ω-records. We assume that a consumer who does not sell her record to the platform forgoes the opportunity of purchasing the merchant's widget and simply enjoys a final payoff of $r(\omega) \geq 0$.[6] On the demand side, the platform decides how many records of each type to demand. Let $q = (q(\omega))_{\omega\in\Omega} \in \mathbb{R}^{\Omega}_{+}$ denote the composition of the *database* demanded by the platform.

In the second period, the platform uses the acquired database q to mediate the interaction between the corresponding consumers and the merchant. More specifically, the platform acts as an information intermediary: It provides the merchant with information about the consumers in the database. In other words, the platform solves a standard information-design problem where the relative frequency of consumers' types is given by q. In this problem, the platform commits to an information structure that maps the records of the consumers in its database into random signals. Given the signal received, the merchant sets a fee $a \in A$ for each consumer, who then purchases the widget only if the merchant's fee a is lower than her willingness to

[5]It is straightforward to allow for records that are only partially informative about ω, a model of which is discussed in Galperti et al. (2023) (Section 4). Instead, the assumption that a data record bundles "access" to a consumer with verifiable information about her type is more substantive. See Ali et al. (2022) for a model where these two components are separated.

[6]This implies that the platform cannot access a consumer without first purchasing her record, and that a consumer cannot access the merchant without first selling her record to the platform.

pay ω. Therefore, the consumer's trading surplus and the merchant's profit are $u(a,\omega) = \max\{\omega - a, 0\}$ and $\pi(a,\omega) = a\mathbb{1}(\omega \geq a)$, respectively. For each interaction, the platform's payoff is $v(a,\omega) = \gamma_u u(a,\omega) + \gamma_\pi \pi(a,\omega)$, where $\gamma_u, \gamma_\pi \geq 0$. That is, the platform's payoff is a linear combination of the consumer's trading surplus and the merchant's profits. To avoid trivial cases, we assume $\gamma_u + \gamma_\pi > 0$.

By standard arguments (e.g., see Bergemann and Morris (2016)), the platform's problem in the second period can be formulated as choosing a recommendation mechanism $x : \Omega \to \Delta(A)$ that solves:

$$(\mathcal{P}_q): \qquad V(q) = \max_{x:\Omega\to\Delta(A)} \sum_{a,\omega} v(a,\omega)x(a|\omega)q(\omega)$$
$$\text{such that} \quad \sum_{\omega} \big(\pi(a,\omega) - \pi(a',\omega)\big)x(a|\omega)q(\omega) \geq 0 \qquad \forall\, a, a' \in A.$$

To summarize, we have introduced a profile of four endogenous variables (p, ζ, q, x): prices p, supply decisions ζ for each consumer, a demanded database q of data records, a mechanism x for problem $\mathcal{P}_q$. The next section will discuss two equilibrium concepts—one for a "monopsonist economy" and one for a "competitive economy"—that differ in how prices p are determined. For the time being, it is useful to introduce a notion of consistency of a profile (p, ζ, q, x) that will be useful for both these equilibrium concepts.

Definition 1. *A profile (p, ζ, q, x) is **consistent** if*

(a). Given x and p, ζ solves the consumers' problem. That is, for all ω,

$$\zeta(\omega) \in \arg\max_{z\in[0,1]} z\Big(p(\omega) + \sum_a x(a|\omega)u(a,\omega)\Big) + (1-z)r(\omega).$$

(b). Given q, x solves the platform's problem $\mathcal{P}_q$ in the second period.

(c). Markets clear. That is, for all ω, $q(\omega) = \zeta(\omega)\bar{q}(\omega)$.

We now briefly discuss these three requirements. Condition (a) requires that each consumer of type ω choose $\zeta(\omega)$ optimally. In the first period, this consumer anticipates that the platform will acquire a database q and use it to implement mechanism x. Therefore, she is willing to sell her record at price $p(\omega)$ only if $p(\omega) + \sum_a u(a,\omega)x(a|\omega) \geq r(\omega)$, where $\sum_a u(a,\omega)x(a|\omega)$ captures her expected trading surplus. Condition (b) requires that the mechanism x that the consumers expect the platform to implement is sequentially rational for the platform. That is, after having acquired database q, the platform is indeed willing to implement mechanism x. Finally, condition (c) requires that the demand of each record type equals its supply.

Next, we specify how prices p are determined. Section 3 studies an economy in which the platform is a price maker and thus chooses p. Section 4, instead, studies a competitive economy in which the platform is a price taker and p is determined by market clearing.

3 The Monopsonist Economy as an Efficiency Benchmark

In this subsection, we introduce a benchmark that will be useful throughout the paper. As in Bergemann et al. (2022), we assume the platform is a monopsonist, i.e., the only buyer of data records. It exerts its bargaining power by setting the prices $p \in \mathbb{R}^{\Omega}$. Specifically, the platform makes a take-it-or-leave-it offer $p(\omega)$ to each type-ω consumer who then decides whether to sell her record. An equilibrium of the monopsonist economy is defined as follows:

Definition 2. *A profile (p^*, ζ^*, q^*, x^*) is an equilibrium of the monopsonist economy if it solves*

$$\begin{aligned} \max_{(p,\zeta,q,x)} \quad & V(q) - \sum p(\omega) q(\omega) \\ \textit{such that} \quad & (p, \zeta, q, x) \textit{ is consistent.} \end{aligned}$$

In words, a profile is an equilibrium of the monopsonist economy if the platform chooses prices and allocation to maximize its profit, subject to the consistency requirement. To understand the properties of equilibria of the monopsonist economy, let us first introduce a notion of constrained efficiency. Given an allocation (q, x), we denote the aggregate payoff that the platform and the consumers obtain by

$$W(q,x) = \sum_{a,\omega} \Big(v(a,\omega) + u(a,\omega) \Big) x(a|\omega) q(\omega) + \sum_{\omega} \Big(\bar{q}(\omega) - q(\omega) \Big) r(\omega). \tag{1}$$

Definition 3. *An allocation (q°, x°) is **constrained efficient** if it solves*

$$\begin{aligned} (\mathcal{SB}): \qquad W^\circ \;=\; & \max_{q,x} \quad W(q,x) \\ & \textit{such that} \quad q \le \bar{q}, \\ & \textit{and} \quad x \textit{ solves } \mathcal{P}_q. \end{aligned}$$

In this benchmark, a planner allocates data records between consumers and the platform without any compensation to the consumers. The planner seeks to maximize $W(q, x)$. The planner's problem is constrained in two ways. First, by feasibility, namely, the planner cannot allocate more records than those that exist. Second, the planner needs to choose a mechanism x that is sequentially rational for the platform given q. That is, the planner cannot force a mechanism onto the platform that it will want to change in the second period.[7]

[7] Remark A.1 in Appendix A shows that a constrained efficient allocation exists.

Fix an equilibrium (p^*, ζ^*, q^*, x^*) of the monopsonist economy. In this equilibrium, consumers welfare is

$$\mathcal{U}(p^*, \zeta^*, q^*, x^*) = \sum_{\omega} q^*(\omega) p^*(\omega) + \sum_{\omega, a} u(a, \omega) x^*(a|\omega) q^*(\omega) + \sum_{\omega} (\bar{q}(\omega) - q^*(\omega)) r(\omega),$$

while the platform's payoff is $\mathcal{V}(p^*, \zeta^*, q^*, x^*) := V(q^*) - \sum_\omega p(\omega) q(\omega)$. Letting $R := \sum_\omega \bar{q}(\omega) r(\omega)$ and using Definition 1(a) and the fact that the platform cannot earn a negative payoff, we obtain that

$$R \leq \mathcal{U}(p^*, \zeta^*, q^*, x^*) \leq W^\circ \qquad \text{and} \qquad 0 \leq \mathcal{V}(p^*, \zeta^*, q^*, x^*) \leq W^\circ - R.$$

We are now ready to characterize the equilibria of the monopsonist economy

Proposition 1. *Let (p^*, ζ^*, q^*, x^*) be an equilibrium of the monopsonist economy. The implied allocation (q^*, x^*) is constrained efficient. Moreover, the platform's equilibrium payoff is maximized, i.e., $\mathcal{V}(p^*, \zeta^*, q^*, x^*) = W^\circ - R$, while the consumer welfare is minimized, i.e., $\mathcal{U}(p^*, \zeta^*, q^*, x^*) = R$. Conversely, any constrained efficient allocation (q°, x°) can be supported as an equilibrium of the monopsonist economy by setting prices equal to $p(\omega) = r(\omega) - \mathbb{E}_{x^\circ}[u(a, \omega)]$ and consumers' supply decisions equal to $\zeta(\omega) = q^\circ(\omega)/\bar{q}(\omega)$ for all ω.*

This result shows that there is an equivalence between the monopsonist problem and the planner's problem from Definition 3. Each equilibrium in the monopsonist economy corresponds to a solution of the planner's problem, and conversely, every planner's problem solution can be supported as an equilibrium in the monopsonist economy. Since a constrained-efficient allocation exists (Remark A.1), this result also establishes the existence of an equilibrium of the monopsonist economy. The monopsonist economy is therefore constrained efficient. It maximizes the aggregate welfare obtained by the platform and the consumers. However, since the platform has all the bargaining power, consumers' welfare is minimized, i.e., it equals their outside option.

4 The Competitive Data Economy

In this section, we study whether by shifting bargaining power away from the platform, an economy can remain efficient while increasing consumers' welfare. Specifically, we do so by studying a competitive economy in which the platform is assumed to be a price taker.[8] The solution concept for this economy is a standard notion of competitive equilibrium, which we define as follows.

Definition 4. *A profile* (p^*, ζ^*, q^*, x^*) *is an equilibrium of the competitive economy if*

(a). *Given* p^*, q^* *solves the platform's problem in the first period, i.e.,*

$$q^* \in \arg\max_{q \in \mathbb{R}^{\Omega}_{+}} V(q) - \sum p^*(\omega) q(\omega). \tag{2}$$

(b). (p^*, ζ^*, q^*, x^*) *is consistent.*

We begin by establishing the existence of an equilibrium of the competitive economy.

Proposition 2. *An equilibrium of the competitive economy exists.*

This result is not immediate because, unlike in the typical Walrasian equilibrium, the consumer's problem depends not only on the prices p but also on the platform's mechanism x, which can change discontinuously in q. Nonetheless, the optimal mechanism x as a correspondence of q is upper-hemicontinuous (Lemma A.1). This property suffices to show existence.

Let (p^*, ζ^*, q^*, x^*) be an equilibrium of the competitive economy. Since $\mathcal{P}$ exhibits constant return to scale, i.e., $V(\beta q^*) = \beta V(q^*)$ for $\beta \geq 0$, the platform cannot earn a positive payoff in equilibrium because, if it did, it would be profitable to demand a larger database. That is, $\mathcal{V}(p^*, \zeta^*, q^*, x^*) = 0$ or, equivalently, $V(q^*) = \sum_{\omega} q^*(\omega) p^*(\omega)$.[9] Therefore, consumers welfare satisfies

$$R \leq \mathcal{U}(p^*, \zeta^*, q^*, x^*) = W(q^*, x^*) \leq W^{\circ}.$$

In other words, if the equilibrium allocation (q^*, x^*) were constrained efficient, consumer welfare would be maximized. This underscores the importance of focusing on the benchmark defined in (1): In a competitive setting, an equilibrium whose allocation is constrained efficient maximize consumer welfare.

Therefore, the central question is whether equilibria of the competitive economy are inducing allocations that are constrained efficient. In the next result we identify a sufficient condition

[8] In practice, we retain the assumption of a single platform and assume it is a price taker. Conceptually, however, this model is equivalent to one where a finite set of identical platforms compete with each other to acquire the data records from the consumers. In Galperti and Perego (2022), we show how a competitive economy with multiple heterogeneous platforms can be modelled in a general way.

[9] See Remark A.2 in Appendix A.

for constrained efficiency but also argue that the market can fail when such a condition is not satisfied.

Proposition 3. *Let (p^*, ζ^*, q^*, x^*) be an equilibrium of the competitive economy. If $\gamma_\pi > \gamma_u$, the equilibrium allocation (q^*, x^*) is constrained efficient. Therefore, consumer welfare is maximized, $\mathcal{U}(p^*, \zeta^*, q^*, x^*) = W^\circ$. If instead $\gamma_\pi \leq \gamma_u$, the equilibrium allocation can be inefficient.*

When $\gamma_\pi > \gamma_u$, the platform's and the merchant's objective are sufficiently aligned so that the platform finds it optimal to be fully transparent with the merchant regarding the types of consumers who belong to its database. In fact, for any q, there is a unique x solving $\mathcal{P}_q$ such that $x(a|\omega) = 1$ if and only if $\omega = a$. This implies that, for each consumer, the merchant sets a fee a that fully extract her value, i.e., $\mathbb{E}_x(u(a, \omega)) = 0$, for all ω. Therefore, consumers' welfare must entirely stem from the prices p^* that they receive from the platform for their records: $\sum_\omega q^*(\omega) p^*(\omega) = W^\circ$.

It is, perhaps, counterintuitive to see that the competitive economy maximizes consumers' welfare precisely when the platform has a stronger incentive to maximize the merchant's profit. The reason is that, the way data records are used by the platform in this case leaves no room for consumers creating externalities on each other when selling their records. In the absence of externalities, it is not surprising that the competitive economy is efficient. So the question is, why a consumer's decision to sell her record creates no externality on other consumers when $\gamma_\pi > \gamma_u$? The intuition for this traces back to Galperti et al. (2023). The optimal mechanism when $\gamma_\pi > \gamma_u$ does not involve pooling consumers with different types together. Whether or not a consumer participates in this mechanism, her decision does not affect the expected payoff of the consumers who do participate. In the absence of external effects, the platform can offer to each consumer the part of its payoff that this consumer contributed to creating, namely $p^*(\omega) = \mathbb{E}_{x^*}(v(a, \omega)) = \gamma_\pi \omega$, for all ω. Thus, a type-ω consumer will sell her record to the platform only if $r(\omega) \leq p^*(\omega) = \gamma_\pi \omega$, which is the same criterion the social planner would use to allocate records between consumers and the platform.

Conversely, when $\gamma_u \geq \gamma_\pi$, the equilibrium mechanism x^* may involve pooling. This is because the platform's objective is more aligned with consumers than the merchant and therefore tries to maximize the welfare of the former. As in Bergemann et al. (2015), the platform achieves so by creating segments where different types of consumers coexist. In this case, the participation decision of some consumers may be essential for delivering the promised payoff to other consumers. The next section illustrate this point.

To conclude, while the positive part of Proposition 3 can be interpreted as "good news," it comes with the caveat that, in many modern digital markets, platforms often have complex objectives that lead them to optimally withhold some information from the merchants (see, e.g., Xu and Yang, 2022). As discussed in Galperti et al. (2023), this usually leads to "pooling externalities," which would then generate the problems described above. Moreover, the qualitative takeaway of Proposition 3 does not depend on the notion of welfare we used in Definition 3, which excludes the merchant's profit. In Proposition B.1 (in Appendix B.1), we prove a similar result that use social welfare (i.e., the sum of consumers, platform, and merchant's payoffs) as the efficiency benchmark.

4.1 The Inefficiency of the Competitive Economy: An Example

In this section, we illustrate with a simple example why, when $\gamma_u \geq \gamma_\pi$, the competitive economy can be inefficient. The example sheds light on the nature of the inefficiency and how it relates to the externality first identified by Galperti et al. (2023). Suppose $\Omega = \{1, 2\}$ and that $\bar{q}(2) > \bar{q}(1)$. Every type has the same outside option $r(\omega) = \bar{r} \in (0, 1)$ for all ω. Finally, $\gamma_\pi = 0$, and thus we are in the case $\gamma_u > \gamma_\pi$.

To avoid uninteresting cases, we assume that $\bar{r} < \frac{1+\gamma_u}{2}$. When $\bar{r} \geq \frac{1+\gamma_u}{2}$, no trade is constrained efficient and $W^\circ = R$. That is, the planner would choose an allocation (q°, x°) where $q^\circ(\omega) = 0$ for all ω's. Indeed, the "cost" of allocating a pair of high- and low-type records to the platform is $2\bar{r}$, while the "benefit" is $1 + \gamma_u$. To see the latter, note that the platform would find it optimal to pool these two consumers to induce the merchant to charge a low fee, $a = 1$. In this case, the high-type consumer would enjoy a trading surplus of 1 and the platform would internalize γ_u of it. When no trade is constrained efficient, the competitive equilibrium is efficient.

When some trading is efficient, instead, we will show that market fails, either by inducing too little or too much trading. Let us begin by computing the constrained efficient allocation (q°, x°), which in this case is unique.

The Constrained-Efficient Allocation. Let $q^\circ(1) = q^\circ(2) = \bar{q}(1)$, which means the platform has all the low-type records and an equal amount of high-type records. In this case, the unique optimal mechanism x° is such that $x^\circ(1|\omega) = 1$ for all ω. When $\bar{r} < \frac{1+\gamma_u}{2}$, (q°, x°) is the unique constrained efficient allocation. To see why, first notice that, given any $q > 0$, an optimal mechanism is to set $x_q(1|\omega) = \min\{q(1), q(2)\}/q(\omega)$ for all ω. Given this, the aggregate welfare induced by the allocation (q, x_q) is $W(q, x_q) = (1 + \gamma_u) \min\{q(1), q(2)\} +$

$(1 - q(1) - q(2))\bar{r}$. This implies that an efficient allocation (q°, x°) must satisfy $q^\circ(1) = q^\circ(2)$. Thus, since by assumption $\bar{q}(1) < \bar{q}(2)$, setting $q^\circ(1) = q^\circ(2) = \bar{q}(1)$ uniquely maximizes $W(q, x_q)$. Therefore, $W^\circ = \bar{r} + \bar{q}(1)(1 + \gamma_u - 2\bar{r})$. △

We now compute the equilibria of the competitive economy. There are three parametric cases to consider (see Figure 1), all leading to a market failure.

Case 1, $\gamma_u < \bar{r}$: Inefficiently Low Trade. Suppose $\gamma_u < \bar{r}$, namely, the consumers' outside option is high relatively to γ_u. Consider a candidate equilibrium (p^*, ζ^*, q^*, x^*) defined as follows: $q^*(\omega) = \zeta^*(\omega) = 0$ for all ω; prices are $p^*(1) = \gamma_u$ and $p^*(2) = 0$; the mechanism is such that $x^*(\omega|\omega) = 1$ for all ω.[10] To verify that this is indeed an equilibrium of the competitive economy let us first show it is consistent (see Definition 1). First, given x^* and p^*, consumers do not have incentives to sell their records. A low-type consumer would lose $\bar{r}$ to only gain $p^*(1) = \gamma_u < \bar{r}$. A high-type consumer, instead, would lose $\bar{r}$ to gain nothing.[11] Second, since $q^* = 0$, x^* trivially solves $\mathcal{P}$. Third, q^* and ζ^* satisfy market clearing as $q^*(\omega) = \zeta^*(\omega)\bar{q}(\omega)$ for all ω. Next, we show that q^* solve the platform's problem in 4.(a). Given any deviation q, the platform obtains a gross payoff of $V(q) = \gamma_u \min\{q(1), q(2)\}$ at a cost of $\sum_\omega q(\omega)p^*(\omega) = \gamma_u q(1)$. Therefore, the net payoff from the deviation is weakly negative. Thus, (p^*, ζ^*, q^*, x^*) is an equilibrium of the competitive economy and consumer welfare is $W(q^*, x^*) = \bar{r} < W^\circ$. △

Why is trading not possible in this equilibrium? The competitive price for low-type records is sufficiently low that low-type consumers do not have incentive to sell. At the same time, it is sufficiently high that the platform has no incentive to buy. Without low-type consumers participation, the market unravels as high-type consumers can only lose from participating in any of the platform's mechanisms. The crux of the problem is that high-type consumers need low-type consumers to participate in order to earn a positive trading surplus. Indeed, a low fee, $a = 1$, is obedient when sufficiently many low-type consumers are pooled with the high-type consumers. The low-type consumers, however, do not internalize the positive externality that they create when selling their data. Moreover, equilibrium prices do not fully account for this externality either.

Before proceeding to the next case, we discuss three observations about this equilibrium.

[10] In fact, (q^*, x^*) is the unique equilibrium allocation, though there can be multiple equilibrium prices supporting it. Appendix B.2 provides a complete characterization.

[11] To see this, suppose a non-zero measure of high-type consumers sell, leading to a different $q \neq q^*$. The sequentially rational mechanism x_q would involve setting $x(2|2) = 1$, thus these consumers would obtain a payoff of zero.

First, note that Definition 4 does not a priori restrict prices to be positive. In this example, negative prices are simply not compatible with equilibrium. Second, note that, if they could, high-type consumers would be ready and able to subsidize low-type consumers to sell their data. However, this competitive economy is "too incomplete" for these side trades to take place. We will return to this point in Section 5.2, as a possible way to correct for this inefficiency. Finally, we observe that the equilibrium allocation discussed above is robust in the following sense: Even if a small number of low-type consumers were exogenously forced to sell, this would not trigger a cascade that lead the equilibrium towards constrained efficiency.

Case 2, $\gamma_u > 2\bar{r}$: Inefficiently High Trade. Suppose next that $\gamma_u > 2\bar{r}$. In this case, the unique equilibrium (p^*, ζ^*, q^*, x^*) is as follows. Prices are $p^*(1) = \gamma_u$ and $p^*(2) = 0$. Consumers choose $\zeta^*(1) = 1$ and $\zeta^*(2) = \min\{1, \frac{\bar{q}(1)}{\bar{r}\bar{q}(2)}\}$. The platform's database is $q^*(1) = \bar{q}(1)$ and $q^*(2) = \min\{\bar{q}(2), \frac{\bar{q}(1)}{\bar{r}}\}$. Notice that since $\bar{r} < 1$, $q^*(2) > \bar{q}(1) = q^*(1)$. Finally, let x^* be such that $x^*(1|1) = 1$ and $x^*(1|2) = \frac{q^*(1)}{q^*(2)}$. To verify that this is indeed an equilibrium of the competitive economy let us first show it is consistent (see Definition 1). First, given x^* and p^*, consumers do not have incentives to deviate. A low-type consumer strictly prefers to sell since $p^*(\omega) = \gamma_u > \bar{r}$. A high-type consumer is indifferent if $\bar{q}(1) \leq \bar{r}\bar{q}(2)$ or strictly prefers to sell otherwise. Second, q^* and ζ^* satisfies market clearing. Third, it is easy to verify that x^* solves $\mathcal{P}$ given q^* (also see Bergemann et al. (2015)). Finally, the platform does not want to deviate for the same reason as in Case 1 above. Thus, (p^*, ζ^*, q^*, x^*) is an equilibrium of the competitive economy and consumer welfare is

$$R < W(q^*, x^*) = \bar{q}(1)(1 + \gamma_u) + \bar{r} \max\left\{\bar{q}(2) - \frac{\bar{q}(1)}{\bar{r}}, 0\right\} < W^\circ.$$

$\triangle$

In this equilibrium, too many high-type consumers sell their data relative to what is optimal, i.e., $q^*(2) > q^\circ(2)$. They are attracted by the possibility of buying the widget at a low fee. The source of the inefficiency is the same as in Case 1. The mechanism x^* involves pooling high- and low-type consumers. When a high-type consumer sells her record to the platform and, thus, participates in the mechanism, she exerts a negative externality on other consumers. On the margin, her individual decision decreases the payoff of the high-type consumers who sold their data to the platform. Specifically, it decreases the probability they buy the widget at a low fee. This externality is not internalized by the consumer nor it is accounted for by the prices.

Case 3, $\bar{r} \leq \gamma_u \leq 2\bar{r}$: Inefficiently High or Low Trade. Finally, we analyze the residual parametric case, $\bar{r} \leq \gamma_u \leq 2\bar{r}$. In this region, two types of equilibria coexist. Their allocations are the same as those discussed in Case 1 and 2. Too see this, notice first that the same equilibrium presented in Case 2 exists in this region, and has the same implications. Second, there are equilibria with no trade, as in Case 1, which are however supported by different prices. For example, it is easy to check that there is an equilibrium with $q^* = \zeta^* = 0$, $x^*(\omega|\omega) = 1$ for all ω, and $p^*(1) = p^*(2) = \frac{\gamma_u}{2}$. △

Summing up, we identified two types of allocations that can be supported in equilibrium: either no trade (Case 1 and 3) or too much trade (Case 2 and 3). Generically, these are all allocations that can be supported in equilibrium.[12] Therefore, our example showed that all equilibria are inefficient when $\bar{r} < \frac{1+\gamma_u}{2}$.

From these examples, it is clear that the inefficiency we highlight stems from the platform's *endogenous* decision of pooling different types of data records together. This happens when the platform finds it optimal to withhold some information from the merchant.[13] In our model, withholding information is optimal only if $\gamma_u \geq \gamma_\pi$, namely when the platform and the consumer's payoffs are sufficiently aligned. Comparing the equilibrium outcome in the competitive economy with that of the monopsonist economy, we find that a competitive market does not even guarantee that consumer welfare will be strictly improved. Indeed, in Case 1, it does not. In Case 2, however, we show that consumer welfare is higher than R while the aggregate welfare is smaller than W°.

Pooling externalities, resulting from information withholding, are likely not the only way in which a competitive data economy can be inefficient. We conjecture that in our setting, "learning externalities" would generate inefficiencies that would further contribute to lowering consumer welfare below constrained efficiency. These types of externalities are discussed in Choi et al. (2019), Bergemann et al. (2022), Acemoglu et al. (2022), and Ichihashi (2021).

5 Remedies

This section discusses remedies to the inefficiency discussed in the previous section. We discuss two solutions. The first consists of establishing a "data union" that manages consumers'

[12]The knife-edge case of $\gamma_u = \bar{r}$ admits more equilibria, but all of them are inefficient. See Appendix B.2 for a complete characterization.

[13]As noted in Galperti et al. (2023), withholding information is a common practice for many digital platforms. For example, Google's "quality score" pools users' searches to increase competition among advertisers (see, e.g., Sayedi et al. (2014)); Uber conceals riders' destinations from drivers to increase riders' welfare; and Airbnb withholds hosts' profile pictures to decrease discrimination.

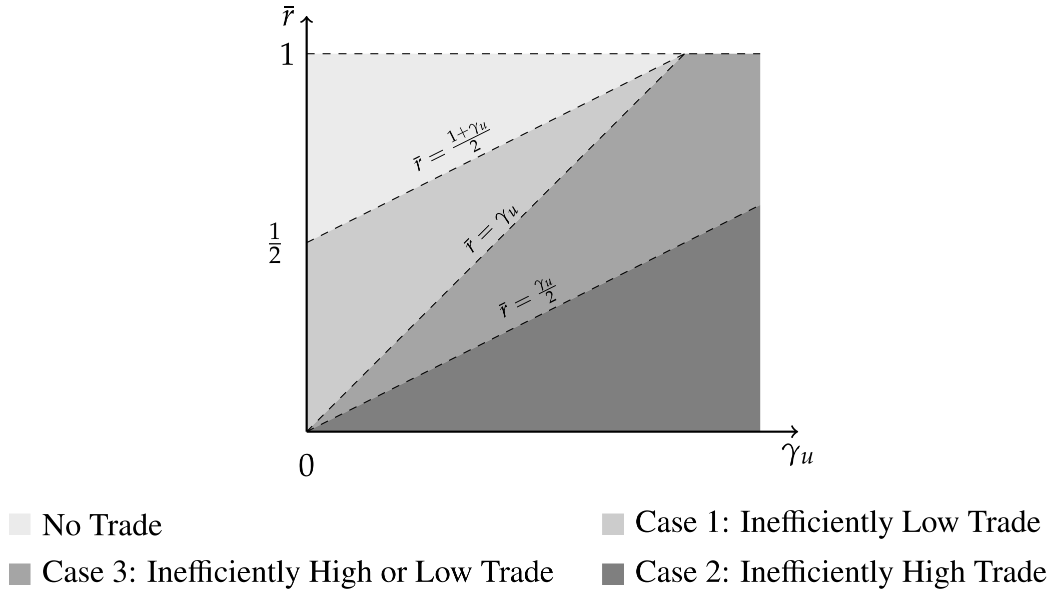

Figure 1: The Inefficiency of the Competitive Economy

data records on their behalf. The second solution consists of opening more markets where, in the spirit of Arrow (1969), the consumers and the platform can directly trade the externalities they create on each other.

5.1 Data Union

In this section, we consider the idea of a "data union," an agent that manages consumers' data records on their behalf, and compensate them by distributing the entirety of the proceedings it obtains from this activity. Versions of this idea have been proposed by Posner and Weyl (2018) and Seim et al. (2022). We illustrate how to design a data union that can implement allocations that are constrained efficient.

Our data union operates as follows. The data union is entrusted with managing the records of all consumers in the population, i.e., it has access to the database $\bar{q}$. It then chooses a database $q \leq \bar{q}$—a subset of the entire consumers population—and sells it to the platform at a price that extracts all the platform's expected payoff, namely, $V(q)$. Finally, the data union distributes the proceeds $V(q)$ to its members. That is, the union chooses $p \in \mathbb{R}^{\Omega}$ such that $\sum_{\omega} p(\omega)\bar{q}(\omega) = V(q)$. Note that payment to a consumer can be negative and can only depend on her type. The payment for each consumer must be sufficiently high to ensure that she does not leave the union.

More formally, the union solves the following problem:

$$\begin{aligned}
\max_{(p,q,x)} \quad & \sum_{\omega} p(\omega)\bar{q}(\omega) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega}(\bar{q}(\omega) - q(\omega))r(\omega) \\
\text{such that} \quad & q \leq \bar{q}, \\
\text{and} \quad & x \text{ solves } \mathcal{P}_q, \\
\text{and} \quad & \sum_{\omega} \bar{p}(\omega)q(\omega) = V(q), \\
\text{and} \quad & p(\omega) + \frac{q(\omega)}{\bar{q}(\omega)}\mathbb{E}_x(u(a,\omega)) + \Big(1 - \frac{q(\omega)}{\bar{q}(\omega)}\Big)r(\omega) \geq r(\omega).
\end{aligned}$$

The last constraint ensures that consumers have no incentive to leave the union. With probability $\frac{q(\omega)}{\bar{q}(\omega)}$, the data record of a type-ω consumer is sold to the platform and this consumer receives a payoff of $\mathbb{E}_x(u(a,\omega))$. With remaining probability, her data record is not sold, and this she maintains her privacy value $r(\omega)$.

Notice that, while consumers can decide to leave the union, they have no say in whether their data records will be sold to the platform or not. This allows the data union to induce allocations that, unlike the equilibria of the competitive economy discussed earlier, are always constrained efficient.

Proposition 4. *Let (p^*, q^*, x^*) be a solution to the data union's problem. The allocation (q^*, x^*) is constrained efficient and maximizes consumers' welfare. Conversely, if (q°, x°) is constrained efficient, there exists p° such that $(p^\circ, q^\circ, x^\circ)$ is a solution to the data union's problem.*

5.2 Lindahl Economy: Trading the Externalities

In this section, we show how the inefficiency highlighted in Section 4 can be corrected while preserving the competitive structure of the economy. We do so by following standard ways of modeling competitive economies with externalities (e.g., Arrow (1969) and Laffont (1976)).[14] Specifically, we allow consumers to trade the way their records are used by the platform. In other words, the platform needs to determine how it intends to use the consumer's record—i.e., which fee a it promises to deliver—at the time of the trade.[15] We will refer to this setting as the Lindahl Economy.

[14]See also, Bonnisseau et al. (2023).

[15]This is reminiscent of the European Union's data protection regulation (GDPR), which requires that "the specific purposes for which personal data are processed should be explicit and legitimate and determined at the time of the collection of the personal data" (see EU's Regulation 2016/679 (39)).

In the Lindahl economy, there is one market for each pair (a,ω). In market (a,ω), type-ω records can be traded for use a at a price $p(a,\omega)$. This means that a consumer of type ω chooses $\zeta(a,\omega) \in [0,1]$ for all a, such that $\sum_a \zeta(a,\omega) \leq 1$, where $\zeta(a,\omega)$ denotes the probability of selling the record to the platform for use a. Likewise, the platform chooses (q,x), where $x(a|\omega)q(\omega)$ represents the quantity of records of type ω that the platform demands for use a. It is implicit in the trade agreement that if the platform acquires a record for use a, the merchant must charge a fee a to the corresponding consumer. That is, the platform's problem is:

$$\begin{aligned} \max_{q,x} \quad & \sum_{a,\omega} \Big(v(a,\omega) - p(a,\omega)\Big) x(a|\omega)q(\omega) \\ \text{such that} \quad & \sum_{\omega} \big(\pi(a,\omega) - \pi(a',\omega)\big) x(a|\omega)q(\omega) \geq 0 \qquad \forall\, a,a' \in A \end{aligned} \tag{3}$$

It is instructive to compare the platform's problem above with the platform's problem in (2) from the competitive economy studied in Section 4. They only differ insofar as the Lindahl economy has richer markets, with prices that depend on a and not just on ω.[16] In particular, the timing in the two economies is the same.

The equilibrium definition in the Lindahl economy is an adaptation of Definition 4.

Definition 5. *A profile (p^*, ζ^*, q^*, x^*) is an equilibrium of the Lindahl economy if*

(a). Given p^*, (q^*, x^*) solves the platform's problem in (3).

(b). Given p^*, ζ^* solves the consumers problem. That is, for all ω,

$$\zeta^*(\cdot,\omega) \in \arg \max_{z \in \mathbb{R}^A_+ \ s.t.\ \sum_a z(a) \leq 1} \sum_a z(a)\big(p^*(a,\omega) + u(a,\omega)\big) + (1 - \sum_a z(a))r(\omega).$$

(c). Markets clear. That is, for all ω and a, $x^*(a|\omega)q^*(\omega) = \zeta^*(a,\omega)\bar{q}(\omega)$.

Notice that, relative to the notion of consistency from Definition 1, the consumer's problem and market clearing condition have been adapted to allow for the richer nature of the consumer's problem. Moreover, we do not explicitly require that x^* maximizes the platform's payoff (condition (b) in Definition 1), as it is already embedded in (3).

Before presenting the main result of this section, we need to introduce a more stringent benchmark than that of constrained efficiency, introduced in Definition 3.

[16] Indeed, notice that we could equivalently write (2) as (3) with the restriction that $p(a,\omega) = p(\omega)$ for all (a,ω). With such a formulation of the platform's problem, we would have removed condition (c) from Definition 1, as it is already embedded in the platform's problem.

Definition 6. *An allocation* $(q^\dagger, x^\dagger)$ *is* ***efficient*** *if it solves*

$$(\mathcal{FB}): \qquad W^\dagger = \max_{q,x} \quad W(q,x)$$
$$\text{such that} \quad q \leq \bar{q},$$
$$\text{and} \quad \sum_\omega \big(\pi(a,\omega) - \pi(a',\omega)\big) x(a|\omega) q(\omega) \geq 0 \qquad \forall\, a, a' \in A$$

Notice that the planner is no longer constrained to choose an x that is sequentially rational for the platform given q. Instead, the planner can choose x on behalf of the platform, and only needs to ensure such an x is obedient for the merchant. By definition, the aggregate welfare induced by an efficient allocation is weakly higher than that induced by a constrained efficient allocation: $W^\dagger \geq W^\circ$. The next result shows that all equilibria of the Lindahl economy are not only constrained efficient, but achieve this first-order level of efficiency just introduced.

Proposition 5. *Let* (p^*, ζ^*, q^*, x^*) *be an equilibrium of the Lindahl economy. The equilibrium allocation* (q^*, x^*) *is efficient. Therefore, consumer welfare is* $\mathcal{U}(p^*, \zeta^*, q^*, x^*) = W^\dagger \geq W^\circ$. *Conversely, any efficient allocation* $(q^\dagger, x^\dagger)$ *can be supported as an equilibrium of the Lindahl economy.*

The price system in the Lindahl economy is sufficiently rich that the platform and the consumers internalize the effects that their choice have on others. This leads to an allocation that is efficient. The inefficiency highlighted in Section 4 disappears. On top of this, the allocation is more than constrained-efficient. Indeed, the equilibrium mechanism x^* is optimal not just from the perspective of the platform, but from the perspective of the planner. The next example illustrates.

Example of a Lindahl Economy. Let us return to the simplified setting of Section 4.1. In this case, since $\gamma_\pi = 0$, a mechanism x is optimal for the platform if and only if it is also optimal for the planner given any q. Indeed, the platform maximizes $\gamma_u u(a,\omega)$ while the planner maximizes $(1+\gamma_u)u(a,\omega)$. Therefore, the notions of efficient and constrained-efficient allocations coincide, i.e., $W^\circ = W^\dagger = \bar{r} + \bar{q}(1)(1+\gamma_u - 2\bar{r})$. Therefore, as in Section 4.1, the efficient allocation $(q^\dagger, x^\dagger)$ is unique and given by $q^\dagger(\omega) = \bar{q}(1)$ and $x^\dagger(1|\omega) = 1$ for all ω. Let (p^*, ζ^*, q^*, x^*) be an equilibrium of the Lindahl economy, which is defined as follows. First of all, let $(q^*, x^*) = (q^\dagger, x^\dagger)$, i.e., the equilibrium supports the efficient allocation. Second, for all ω, let $\zeta^*(1,\omega) = \frac{\bar{q}(1)}{\bar{q}(\omega)}$ and $\zeta^*(2,\omega) = 0$. Finally, the price system is such that $p(a = 2,\omega) = 0$, for all ω, $p^*(a=1,\omega=1) = \gamma_u + (1-\bar{r})$, and $p^*(a=1,\omega=2) = -(1-\bar{r})$. Next we show this is an equilibrium of the Lindahl economy. Note that given the prices, type-1 consumers strictly prefer to sell with $a = 1$, and type-2 consumers are indifferent between not selling and selling with $a = 1$, so the consumer optimality is satisfied. The platform maximizes $(\gamma_u + 1 - \bar{r})(x(1|2)q(2) - x(1|1)q(1))$ subject to $x(1|1)q(1) \geq x(1|2)q(2)$. Therefore, the platform cannot make a positive payoff, and (q^*, x^*) achieves the maximum of 0 given p^*. $\triangle$

In the example, we can appreciate how $p^*(1, \omega = 1)$ captures the entire positive externality that low-type consumers generate when selling their record. This is $(1 + \gamma_u)$ minus their opportunity cost $\bar{r}$. Another feature of this equilibrium is that the high-type consumers have to pay in order to participate in the platform's mechanism, rather than the opposite. The platform uses this payment to acquire low-type records. That is, it is as if high-type consumers who participate in the platform's mechanism subsidize the participation of low-type consumers. This ensure existence of this equilibrium even when $\gamma_u < \bar{r}$. Notice that the equilibrium exists even if $p^*(1, \omega = 2) < 0$. Despite the negative price, the platform does not have an incentive to acquire an arbitrary quantity of such records because it needs to guarantee the merchant is willing to charge a low fee $a = 1$ to all of them.

Before concluding, we note that, while the Lindahl economy guarantees that all equilibria are efficient, it requires a potentially unrealistic number of markets to be open and a high level of finesse. It is natural to wonder whether there are more practical ways of decentralizing (perhaps only partially) the allocation achieved by the Lindahl economy. This partial decentralization may not achieve constrained efficiency but could still be an improvement on the equilibria of the competitive economy discussed in Section 4. This question remains open for future research.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2022): "Too much data: Prices and inefficiencies in data markets," *American Economic Journal: Microeconomics*, 14, 218–256.

Acquisti, A., C. Taylor, and L. Wagman (2016): "The Economics of Privacy," *Journal of Economic Literature*, 54, 442–92.

Ali, S. N., G. Lewis, and S. Vasserman (2022): "Voluntary Disclosure and Personalized Pricing," *forthcoming, Review of Economic Studies*.

Arrow, K. J. (1969): "The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-Market Allocation," *The Analysis and Evaluation of Public Expenditures: the PPB System*, Joint Economic Committee, Congress of the United States, Washington, D.C., 47–64.

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

Bergemann, D., A. Bonatti, and T. Gan (2022): "The economics of social data," *The RAND Journal of Economics*, 53, 263–296.

Bergemann, D., A. Bonatti, and A. Smolin (2018): "The Design and Price of Information," *American Economic Review*, 108, 1–48.

Bergemann, D., B. Brooks, and S. Morris (2015): "The Limits of Price Discrimination," *American Economic Review*, 105 (3).

Bergemann, D. and S. Morris (2016): "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games," *Theoretical Economics*, 11, 487–522.

——— (2019): "Information Design: A Unified Perspective," *Journal of Economic Literature*, 57(1), pp. 44-95).

Bergemann, D. and M. Ottaviani (2021): "Information markets and nonmarkets," in *Handbook of industrial organization*, Elsevier, vol. 4, 593–672.

Böhm, V. (1975): "On the continuity of the optimal policy set for linear programs," *SIAM Journal on Applied Mathematics*, 28, 303–306.

Bonatti, A. and D. Bergemann (2023): "Data, Competition, and Digital Platforms," *Working Paper*.

Bonatti, A., M. Dahleh, T. Horel, and A. Nouripour (2022): "Selling Information in Competitive Environments," *Working Paper*.

Bonnisseau, J.-M., E. L. del Mercato, and P. Siconolfi (2023): "Existence of an equilibrium in arrowian markets for consumption externalities," *Journal of Economic Theory*, 209, 105638.

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

Federal Trade Commission (2014): *Data Brokers: A Call for Transparency and Accountability*, A Report by the Federal Trade Commission, May.

Galperti, S., A. Levkun, and J. Perego (2023): "The Value of Data Records," *Review of Economic Studies*, rdad044.

Ichihashi, S. (2021): "The economics of data externalities," *Journal of Economic Theory*, 196, 105316.

Kamenica, E. (2019): "Bayesian persuasion and information design," *Annual Review of Economics*, 11, 249–272.

Laffont, J. J. (1976): "Decentralization with Externalities," *European Economic Review*, 359–375.

Posner, E. and E. G. Weyl (2018): *Radical Markets: Uprooting Capitalism and Democracy* *for a Just Society*, Princeton University Press.

Sayedi, A., K. Jerath, and K. Srinivasan (2014): "Competitive Poaching in Sponsored Search Advertising and Its Strategic Impact on Traditional Advertising," *Marketing Science*, September, 33 (4), 586–608.

Seim, K., D. Bergemann, J. Cremer, D. Dinielli, C. C. Groh, P. Heidhues, D. Schaefer, M. Schnitzer, F. M. Scott Morton, and M. Sullivan (2022): "Market Design for Personal Data," *Policy Discussion Paper*, No. 6, Tobin Center for Economic Policy, Yale University, April.

Xu, W. and K. H. Yang (2022): "Informational Intermediation, Market Feedback, and Welfare Losses," *Working Paper*.

Yang, K. H. (2022): "Selling consumer data for profit: Optimal market-segmentation design and its consequences," *American Economic Review*, 112, 1364–93.

Appendix

A Proofs

Lemma A.1. *$V(q)$ is continuous and has constant return to scale, i.e., $V(\beta q) = \beta V(q)$ for $\beta \geq 0$. Moreover, the solution correspondence of $\mathcal{P}_q$ is nonempty-valued, compact-valued, and upper-hemicontinuous.*

Proof of Lemma A.1. Fix q. Note that $\mathcal{P}_q$ can be reformulated as:

$$\begin{aligned} \max_{\chi \geq 0} \quad & \sum_{a,\omega} v(a,\omega)\chi(a,\omega) \\ \text{such that} \quad & \sum_{\omega} \big(\pi(a,\omega) - \pi(a',\omega)\big)\chi(a,\omega) \geq 0 \qquad \forall\, a, a' \in A. \\ \text{and} \quad & \sum_{a} \chi(a,\omega) = q(\omega) \qquad \forall\, \omega \in \Omega \end{aligned}$$

In this problem, the objective is continuous in χ and the feasible set is nonempty (because $\chi(\omega,\omega) = q(\omega)$ is always feasible) and compact. Therefore, the solution correspondence is nonempty- and compact-valued. By Theorem 2 of Böhm (1975), it is also continuous. This directly implies $V(q)$ is continuous.

Since x is a solution if and only if χ is a solution where $\chi(a,\omega) = x(a|\omega)q(\omega)$, we conclude the solution correspondence for $\mathcal{P}_q$ is nonempty-valued, compact-valued, and upper-hemicontinuous.

From the problem formulation, it is easy to see that, for any $\beta > 0$, χ is a solution to $\mathcal{P}_q$ if and only if $\beta\chi$ is a solution to $\mathcal{P}_{\beta q}$. Moreover, $V(0) = 0$ because $\chi = 0$ is the only feasible solution in this case. Therefore, $V(\beta q) = \beta V(q)$ for $\beta \geq 0$. □

Remark A.1. *A constrained efficient allocation solving $\mathcal{SB}$ exists.*

Proof of Remark A.1 The maximization problem in $\mathcal{SB}$ admits a solution because by Lemma A.1, the set of allocations satisfying the two constraints is nonempty and compact. □

Proof of Proposition 1. We first argue that the monopsony value is no higher than the value of the following problem, which is the same as $\mathcal{SB}$ up to a constant:

$$\begin{aligned} \max_{q,x \in \mathbb{R}_+^{\Omega}} \quad & \sum_{a,\omega} \Big(v(a,\omega) + u(a,\omega)\Big) x(a|\omega)q(\omega) - \sum_{\omega} q(\omega)r(\omega) \\ \text{such that} \quad & q(\omega) \leq \bar{q}(\omega), \qquad \forall \omega \in \Omega \\ & x \text{ is a solution to } \mathcal{P} \text{ given } q. \end{aligned} \tag{A.1}$$

Take any consistent (p, q, x, ζ). By (a) of Definition 1, we have:

$$p(\omega)q(\omega) \geq (r(\omega) - \mathbb{E}_x[u(a,\omega)])q(\omega). \tag{A.2}$$

Imposing this condition and dropping constraint (a) in Definition 1, we get that (A.1) is a relaxation of the monopsonist's problem, so its value is weakly higher than the monopsony value.

Next, we show that if (q, x) is an efficient allocation, then (p, q, x, ζ) is an equilibrium of the monopsonist economy with $p(\omega) = r(\omega) - \mathbb{E}_x[u(a,\omega)]$ and $\zeta(\omega) = q(\omega)/\bar{q}(\omega)$. Note that under (p, q, x, ζ), the platform's value is exactly that of (A.1). Since (q, x) is a maximizer of (A.1) and the monopsony value is no higher than that of (A.1), we only need to show (p, q, x, ζ) is consistent. To see this, note that since (q, x) is in the feasible set of $\mathcal{SB}$, we have that (c) in Definition 1 is met, and $0 \leq q(\omega) \leq \bar{q}(\omega)$. Thus, $0 \leq \zeta(\omega) \leq 1$ and (b) in Definition 1 is met. Moreover, since $p(\omega) + \mathbb{E}_x[u(a,\omega)] = r(\omega)$, buyers are indifferent between keeping or selling their data, so (a) in Definition 1 is met. Therefore, we conclude (p, q, x, ζ) is consistent, and thus is an equilibrium of the monopsonist economy.

Finally, we argue that any equilibrium of the monopsony economy (p, q, x, ζ) is efficient with (A.2) binding. To see this, let (q', x') be an efficient allocation, then we have:

$$\begin{aligned} V(q) - \sum p(\omega)q(\omega) &\leq \sum_{a,\omega} \Big(v(a,\omega) + u(a,\omega)\Big) x(a|\omega)q(\omega) - \sum_{\omega} q(\omega)r(\omega) \\ &\leq \sum_{a,\omega} \Big(v(a,\omega) + u(a,\omega)\Big) x'(a|\omega)q'(\omega) - \sum_{\omega} q'(\omega)r(\omega) \end{aligned}$$

where the first inequality follows from (A.2). Following the argument of the previous paragraph, we have that (p', q', x', ζ') is consistent, with $p'(\omega) = r(\omega) - \mathbb{E}_{x'}[u(a,\omega)]$ and $\zeta'(\omega) = q'(\omega)/\bar{q}(\omega)$. Therefore, in order for (p, q, x, ζ) to be an equilibrium, both inequalities need to bind. In order for the first inequality to be binding, (A.2) needs to bind; in order for the second inequality to be binding, (q, x) needs to be efficient. This completes the proof. □

Remark A.2. *Let (p^*, ζ^*, q^*, x^*) be an equilibrium of the competitive economy. The platform's equilibrium payoff is zero, i.e., $\mathcal{V}(p^*, \zeta^*, q^*, x^*) = 0$.*

Proof of Remark A.2. Suppose in a competitive equilibrium (p, q, x, ζ), the platform makes a strictly positive profit. Then take any $\beta > 1$, by choosing βq the platform makes a profit

$$V(\beta q) - \sum p(\omega)\beta q(\omega) = \beta(V(q) - \sum p(\omega)q(\omega)) > V(q) - \sum p(\omega)q(\omega),$$

where Lemma A.1 shows that $V(\beta q) = \beta V(q)$. This contradicts that q solves the platform's problem (2). Therefore, the platform must make zero profit in a competitive equilibrium. □

Proof of Proposition 3. Let (p^*, q^*, x^*, ζ^*) be an equilibrium for the competitive economy. In period 1, consider any alternative database q that the platform could acquire. Definition 4 requires that

$$V(q^*) - V(q) \geq \sum_{\omega} p^*(\omega)\Big(q^*(\omega) - q(\omega)\Big). \tag{A.3}$$

Now consider the problem of a consumer of type ω and any deviation $z \in [0,1]$. Definition 4 requires that

$$\zeta^*(\omega)(p^*(\omega) + \mathbb{E}_{x^*}(u(a,\omega)) + (1-\zeta^*(\omega))r(\omega) \geq z(p^*(\omega) + \mathbb{E}_{x^*}(u(a,\omega)) + (1-z)r(\omega).$$

We can multiply $\bar{q}(\omega)$ on both sides and define $q'(\omega) = z\bar{q}(\omega)$. Manipulating terms, we get

$$\Big(q^*(\omega) - q'(\omega)\Big)\Big(\mathbb{E}_{x^*}(u(a,\omega)) - r(\omega)\Big) \geq -p^*(\omega)\Big(q^*(\omega) - q'(\omega)\Big).$$

Summing over ω,

$$\sum_{\omega}\Big(q^*(\omega) - q'(\omega)\Big)\Big(\mathbb{E}_{x^*}(u(a,\omega)) - r(\omega)\Big) \geq -\sum_{\omega} p^*(\omega)\Big(q^*(\omega) - q'(\omega)\Big). \tag{A.4}$$

Thus, (A.3) and (A.4) jointly imply that for all $q \in \mathbb{R}^{\Omega}_{+}$ such that $q \leq \bar{q}$,

$$V(q^*) - V(q) + \sum_{\omega}\Big(q^*(\omega) - q(\omega)\Big)\Big(\mathbb{E}_{x^*}(u(a,\omega)) - r(\omega)\Big) \geq 0. \tag{A.5}$$

Now suppose $\gamma_\pi > \gamma_u$. We first argue that in this case the unique solution to $\mathcal{P}$ for any q is to choose a mechanism $\hat{x}$ that allows the merchant to engage in first-degree price discrimination. That is, $\hat{x}(a|\omega) = 1$ if $a = \omega$. To see this, recall that Theorem 1 of Bergemann et al. (2015) states that given data set q, by varying x, the feasible set of consumer surplus s_u and merchant's profit s_π is characterized by:[17]

$$s_u \geq 0,\ s_\pi \geq \pi^*,\ s_u + s_\pi \leq w^*.$$

Here, $\pi^* := \max_a \sum_\omega \pi(a,\omega)q(\omega)$ is the monopolistic profit and $w^* := \sum_\omega \omega q(\omega)$ is the total surplus. This feasible set is shown by the shaded area in Figure 2. Now, the platform's problem $\mathcal{P}$ is equivalent to choosing (s_u, s_π) in this feasible set to maximize $\gamma_u s_u + \gamma_\pi s_\pi$. Its indifference curves (denoted by IC_v in Figure 2) are given by the lines with slope $-\gamma_u/\gamma_\pi$. Therefore, when $\gamma_\pi > \gamma_u$, the unique optimal mechanism is full disclosure.

Since by assumption x^* solves $\mathcal{P}$, it must be that $x^* = \hat{x}$. Therefore, we can rewrite (A.5) as:

$$q^* \in \arg\max_{q} \sum_{a,\omega}\Big(v(a,\omega) + u(a,\omega)\Big)\hat{x}(a|\omega)q(\omega) - \sum_{\omega} q(\omega)r(\omega).$$

[17]Formally, given an obedient mechanism x, the consumer's trading surplus and the merchant's profit are defined by $s_u = \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega)$ and $s_\pi = \sum_{a,\omega} \pi(a,\omega)x(a|\omega)q(\omega)$, respectively.

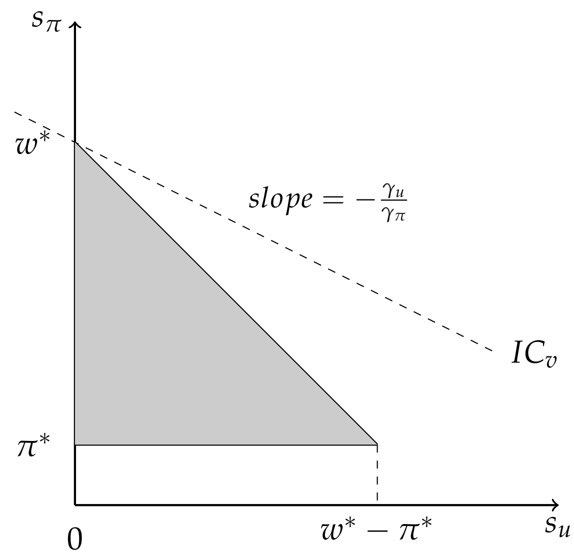

Figure 2: Bergemann et al. (2015)'s Triangle

$$s.t.\ q \leq \bar{q}$$

Since $\hat{x}$ uniquely solves $\mathcal{P}$ for all q, this problem is equivalent to the social planner's. Therefore, the equilibrium allocation (q^*, x^*) is constrained efficient. □

Proof of Proposition 4.

Only If. Let (q^*, x^*) be a solution to the planner's problem $\mathcal{SB}$. First, we argue that (q^*, x^*) is a solution of a relaxed version of the data union's problem. We first discard the consumer's participation constraint from the data union's problem. Then, let us substitute the constraint $\sum_\omega \hat{q}(\omega)p(\omega) = V(q)$ into the data union's objective. By doing so, prices p do not appear in the relaxed problem. Summing up, the relaxed problem is

$$\begin{aligned} \max_{(q,x)} \quad & \sum_{a,\omega} \Big(v(a,\omega) + u(a,\omega)\Big) x(a|\omega) q(\omega) + \sum_{\omega} (\bar{q}(\omega) - q(\omega)) r(\omega) \\ \text{such that} \quad & q \leq \bar{q}, \\ \text{and} \quad & x \text{ solves } \mathcal{P} \text{ at } q. \end{aligned}$$

This problem is exactly the planner's problem. Since (q^*, x^*) is a solution to the relaxed problem, it must yield a value that is weakly higher than the value of the data union's problem, the proof is complete if we find prices p^* such that the participation constraints are satisfied and the data union's budget is balanced, given (q^*, x^*).

To this end, let $p^*(\omega) = \tilde{p}(\omega) + t(\omega)$ with $\tilde{p}(\omega) = \frac{q^*(\omega)}{\bar{q}(\omega)}\Big(r(\omega) - \mathbb{E}_{x^*}(u(a,\omega))\Big)$. We pin down $t(\omega)$ later. If $t(\omega) = 0$, all type-ω consumers would be indifferent between joining the union or not, and in particular, $\zeta^*(\omega) = 1$ is optimal. In this case, the union's budget is:

$$\begin{aligned} G(q^*, x^*) &= V(q^*) - \sum_{\omega} \bar{q}(\omega)\tilde{p}(\omega) \\ &= \sum_{a,\omega} \Big(v(a,\omega) + u(a,\omega)\Big) x^*(a|\omega) q^*(\omega) - \sum_{\omega} q^*(\omega) r(\omega). \end{aligned}$$

Since (q^*, x^*) is constrained efficient, $G(q^*, x^*) \geq 0$. To see this, we add $\sum_\omega \bar{q}(\omega) r(\omega)$ on both sides of this inequality. On the left hand side, we obtain the value of the planner's objective at (q^*, x^*), which must be no smaller than $\sum_\omega \bar{q}(\omega) r(\omega)$ because it is always feasible for the planner.

Since the union cannot earn a profit, we redistribute $G(q^*, x^*)$ back to the consumers in a uniform manner. Specifically, we let $t(\omega) = G(q^*, x^*)$ (recall that $\sum_\omega \bar{q}(\omega) = 1$). Therefore, if $\zeta^*(\omega) = 1$ was optimal under $\tilde{p}(\omega)$, it is still optimal under $p^*(\omega) \geq \tilde{p}(\omega)$.

We thus constructed a profile (p^*, q^*, x^*) that is feasible for the data union. Moreover, since (q^*, x^*) solves the relaxed problem, it also solves the data union's problem.

If Direction. Let (p^*, q^*, x^*) be a solution to the data union's problem. To the contrary suppose it is not constrained efficient. Then take any constrained efficient allocation (q°, x°). We have that:

$$\begin{aligned} &\sum_{a,\omega} \Big(v(a,\omega) + u(a,\omega)\Big) x^*(a|\omega) q^*(\omega) - \sum_{\omega} q^*(\omega) r(\omega) \\ <&\sum_{a,\omega} \Big(v(a,\omega) + u(a,\omega)\Big) x^\circ(a|\omega) q^\circ(\omega) - \sum_{\omega} q^\circ(\omega) r(\omega) \end{aligned} \tag{A.6}$$

By the "only if" direction, we know there exist p° such that $(p^\circ, q^\circ, x^\circ)$ is feasible for the data union. (A.6) implies that:

$$\begin{aligned} &\textstyle\sum_\omega p^*(\omega)\bar{q}(\omega) + \sum_{a,\omega} u(a,\omega) x^*(a|\omega) q^*(\omega) + \sum_\omega (\bar{q}(\omega) - q^*(\omega)) r(\omega) \\ < \;&\textstyle\sum_\omega p^\circ(\omega)\bar{q}(\omega) + \sum_{a,\omega} u(a,\omega) x^\circ(a|\omega) q^\circ(\omega) + \sum_\omega (\bar{q}(\omega) - q^\circ(\omega)) r(\omega) \end{aligned}$$

This contradicts (p^*, q^*, x^*) being a solution to the data union's problem, so we conclude it must be constrained efficient. □

Proof of Proposition 5.

Step 1: Let (p^*, q^*, x^*, ζ^*) be a Lindahl equilibrium. We first prove that (q^*, x^*) must solve $\mathcal{FB}$. Since (q^*, x^*) solves $\mathcal{P}'$, we have that

$$\begin{aligned} &\sum_{a,\omega} v(a,\omega) x^*(a|\omega) q^*(\omega) - \sum_{a,\omega} v(a,\omega) x(a|\omega) q(\omega) \\ \geq &\sum_{a,\omega} p^*(a,\omega) x^*(a|\omega) q^*(\omega) - \sum_{a,\omega} p^*(a,\omega) x(a|\omega) q(\omega) \end{aligned} \tag{A.7}$$

for all (q,x) that satisfies obedience. Similarly, by the maximization problem of type-ω consumers, we get

$$\sum_a u(a,\omega)\zeta^*(a,\omega) + r(\omega)\big(1-\sum_a \zeta^*(a,\omega)\big) - \sum_a u(a,\omega)\zeta(a,\omega) - r(\omega)\big(1-\sum_{a,} \zeta(a,\omega)\big)$$
$$\geq -\sum_a p^*(a,\omega)\zeta^*(a,\omega) + \sum_a p^*(a,\omega)\zeta(a,\omega)$$

for all $\zeta(a,\omega) \in \mathbb{R}^A_+$ such that $\sum_a \zeta(a,\omega) \leq 1$. Summing over consumers of the same type and across type, we get that for all (q,x) such that $q \leq \bar{q}$:

$$\sum_{a,\omega} u(a,\omega)x^*(a|\omega)q^*(\omega) - \sum_\omega r(\omega)q^*(\omega) - \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_\omega r(\omega)q(\omega)$$
$$\geq -\sum_{a,\omega} p^*(a,\omega)x^*(a|\omega)q^*(\omega) + \sum_{a,\omega} p^*(a,\omega)x(a|\omega)q(\omega). \tag{A.8}$$

Equations (A.7) and (A.8) jointly imply that for all (q,x) satisfying feasibility and obedience:

$$\begin{aligned} &\sum_{a,\omega}\big(v(a,\omega)+u(a,\omega)\big)x^*(a|\omega)q^*(\omega) - \sum_\omega r(\omega)q^*(\omega) \\ \geq \quad &\sum_{a,\omega}\big(v(a,\omega)+u(a,\omega)\big)x(a|\omega)q(\omega) - \sum_\omega r(\omega)q(\omega). \end{aligned}$$

Therefore, (q^*,x^*) solves $\mathcal{FB}$.

Step 2: We now prove that for any allocation (q^*,x^*) that solves $\mathcal{F}B$, there is a (p^*,ζ^*) such that (p^*,q^*,x^*,ζ^*) is a Lindahl equilibrium. First of all, notice that $\mathcal{F}B$ admits an optimal solution. Second, we can define $p^*(a,\omega) = r(\omega) - u(a,\omega)$ for all a,ω, so that each ω consumer is indifferent across all possible $\zeta(\cdot,\omega)$ and we can therefore assume to choose ζ^* such that $\zeta^*(\cdot,\omega)\bar{q}(\omega) = x^*(\cdot|\omega)q^*(\omega)$.

We can equivalently rewrite $\mathcal{FB}$ in terms of χ:

$$\begin{aligned} (\mathcal{FB}'): \quad \max_{\chi\in\mathbb{R}^{A\times\Omega}_+} \quad & \sum_{a,\omega}\Big(v(a,\omega)+u(a,\omega)\Big)\chi(a,\omega) + \sum_\omega\Big(\bar{q}(\omega) - \sum_a \chi(a,\omega)\Big)r(\omega) \\ \text{such that} \quad & \sum_a \chi(a,\omega) \leq \bar{q}(\omega), \qquad \forall \omega\in\Omega \\ \text{and} \quad & \sum_\omega \big(\pi(a,\omega) - \pi(\hat{a},\omega)\big)\chi(a,\omega) \geq 0 \qquad \forall\, a,\hat{a}\in A \end{aligned}$$

Since (q^*,x^*) is a first-best efficient allocation, we know $\chi^*(a,\omega) := x^*(a|\omega)q^*(\omega)$ solves $\mathcal{FB}'$. Define $\zeta^*(a,\omega) = \chi^*(a,\omega)/\bar{q}(\omega)$. Since χ^* is an optimal solution to $\mathcal{FB}'$, by strong duality, we know its dual admits an optimal solution $(\mu^*(\omega),\lambda^*(\hat{a}|a))$. Define $p^*(a,\omega) = \mu^*(\omega) + r(\omega) - u(a,\omega)$.

We first argue that given p^*, $\zeta^*(\omega)$ is optimal for type-ω consumers. When $\mu^*(\omega) = 0$, we have $p^*(a,\omega) = r(\omega) - u(a,\omega)$. Thus, type-$\omega$ consumers are indifferent between keeping the data and selling it with any a, so $zeta^*(\cdot,\omega)$ is optimal. When $\mu^*(\omega) > 0$, by complementary slackness, we have that $\sum_a \zeta^*(a,\omega) = 1$. Therefore, no type-$\omega$ consumer keeps the data. Since selling the data with any a gives the consumer a payoff of $\mu^*(\omega) + r(\omega)$. They are indifferent between different a and thus $\zeta^*(\omega)$ is optimal.

Next, we argue that χ^* solves the platform's problem given p^*. We first show that the platform's payoff is non-positive under p^*. To show this, we only need to show the dual problem of the platform's problem is feasible. The dual feasible set is given by:

$$\sum_{\hat{a}}(\pi(\hat{a},\omega) - \pi(a,\omega))\lambda(\hat{a}|a) \geq v(a,\omega) - p^*(a,\omega)$$

$$= v(a,\omega) + u(a,\omega) - \mu^*(\omega) - r(\omega)$$

for all a, ω, with $\lambda \geq 0$. But we know this is feasible because λ^* satisfies these constraints. Given dual feasibility, weak duality implies:

$$\sum_{a,\omega}(v(a,\omega) - p^*(a,\omega))\chi(a,\omega) \leq 0$$

for all χ that is feasible to the platform.

Finally, by strong duality we have:

$$\sum_{a,\omega}\Big(v(a,\omega) + u(a,\omega)\Big)\chi^*(a,\omega) - \sum_{a,\omega}\chi^*(a,\omega)r(\omega) = \sum_{\omega}\mu^*(\omega)\bar{q}(\omega).$$

This implies:

$$\sum_{a,\omega}(v(a,\omega) - p^*(a,\omega) + \mu^*(\omega))\chi^*(a,\omega) = \sum_{\omega}\mu^*(\omega)\bar{q}(\omega).$$

By complementary slackness we know $\sum_{a,\omega}\mu^*(\omega)\chi^*(a,\omega) = \sum_{\omega}\mu^*(\omega)\bar{q}(\omega)$, which implies:

$$\sum_{a,\omega}(v(a,\omega) - p^*(a,\omega))\chi^*(a,\omega) = 0.$$

Therefore, we conclude χ^* solves the platform's problem given p^*.

A.1 Proof of Proposition 2

We start by characterizing prices under which a data allocation q if a solution to the platform's problem (2). Toward this, note that the platform's problem is essentially to choose (q, x) given price p, or equilvalently choosing $\chi(a,\omega) = x(a|\omega)q(\omega)$. Therefore, its dual problem can be formulated as:

$$
\begin{aligned}
(\mathcal{P}'_q): \quad & \min_{\psi,\lambda} \quad \sum_{\omega} \psi(\omega)q(\omega) \\
\text{such that} \quad & \psi(\omega) \geq v(a,\omega) + \sum_{\hat{a}} (\pi(a,\omega) - \pi(\hat{a},\omega))\lambda(\hat{a}|a) \qquad \forall a, \omega \\
\text{and} \quad & \lambda(\hat{a}|a) \geq 0 \qquad \forall \hat{a}, a
\end{aligned}
$$

The next lemma states that ψ is a solution to $\mathcal{P}'_q$ if and only if q is a solution to the platform under price $p = \psi$.[18] Intuitively, this is in light of Remark A.2, which states that in a competitive equilibrium the platform breaks even. Therefore, the price must equal to the marginal value of a data record whenever $q(\omega) > 0$.

Lemma A.2. *A data allocation $0 \leq q \leq \bar{q}$ is a solution to the platform's problem (2) under price p if and only if there exists λ such that (p, λ) is a solution to $\mathcal{P}'_q$.*

Proof. Toward showing this, we first formulate the dual of the platform's problem (2) as:

$$
\begin{aligned}
& \min_{\lambda} \quad 0 \\
\text{such that} \quad & \sum_{\hat{a}} (\pi(\hat{a},\omega) - \pi(a,\omega))\lambda(\hat{a}|a) \geq v(a,\omega) - p(\omega) \qquad \forall a, \omega \qquad \text{(A.9)} \\
\text{and} \quad & \lambda(\hat{a}|a) \geq 0 \qquad \forall \hat{a}, a
\end{aligned}
$$

To show the "only if" direction, suppose $0 \leq q \leq \bar{q}$ solves the platform's problem (2) under price p. Then we must have $V(q) - \sum_{\omega} p(\omega)q(\omega) = 0$. By strong duality, Problem (A.9) is feasible. Take any feasible solution λ, and consider (ψ, λ) where $\psi = p$. Next we argue (ψ, λ) is an optimal solution to $\mathcal{P}'_q$. Suppose not, then since (ψ, λ) is feasible to $\mathcal{P}'_q$, we must have $V(q) < \sum_{\omega} \psi(\omega)q(\omega)$, but this contradicts $V(q) - \sum_{\omega} p(\omega)q(\omega) = 0$.

To show the "if" direction, suppose (p, λ) is an optimal solution to $\mathcal{P}'_q$. This means Problem (A.9) is feasible. Therefore, the platform's optimal payoff is 0 in this competitive economy. By strong duality, we have $V(q) = \sum_{\omega} p(\omega)q(\omega)$. This means that q gives the platform a payoff of 0. Therefore, q is a solution to the platform's problem (2) given price p. □

Next we introduce a correspondence whose fixed points characterize the set of competitive equilibria. Let $P = [-M, M]^{|\Omega|}$ be the space of potential prices, where M is chosen to be large so that any possible equilibrium prices are within that range. Let $Q \times X$ be the space of

[18] The solution of $\mathcal{P}'_q$ is characterized by Proposition 2 of Galperti et al. (2023), where their weights r and $1 - r$ correspond to our γ_π and γ_u, respectively.

feasible data allocations. Taken together, $P \times Q \times X$ is a nonempty, compact, and convex set. Define a correspondence $F : P \times Q \times X \rightrightarrows P \times Q \times X$ such that $(p', q', x') \in F(p, q, x)$ if:

1. x' solves problem $\mathcal{P}_q$.
2. q' solves the consumers' problem given (p, x).[19]
3. p' is such that q solves the platform's problem (2).

Note that (p, q, x) is a competitive equilibrium if and only if it is a fixed point of F. Therefore, a competitive equilibrium exists if F admits a fixed point. Toward this, we first show the following lemma.

Lemma A.3. *F is nonempty-valued, convex-valued, and has a closed graph.*

Proof. We first show that F is nonempty-valued. Fix any (p, q, x). By Lemma A.1, $\mathcal{P}_q$ admits a solution x'; given (p, x), the consumers' problem always has a solution q'; given q, since $\mathcal{P}_q$ admits an optimal solution, by strong duality $\mathcal{P}'_q$ also admits an optimal solution. Lemma A.2 then implies that a price p' under which q solves the platform's problem exists.[20] Therefore, $(p', q', x') \in F(p, q, x)$.

Next we show F is convex-valued. Note that by definition of F, given (p, q, x), the choice of p', q', and x' are independent with each other. Therefore, it is sufficient to check convexity for each dimension. If x' and x'' both solve $\mathcal{P}_q$, clearly any convex combination also solves it; If q' and q'' both solve the consumers' problem, then any convex combination also solves the consumers' problem. To see this, if under (p, x) consumer ω has a strict preference, then $q'(\omega) = q''(\omega)$. if under (p, x) consumer ω is indifferent, then any $q(\omega)$ is optimal; If under both p' and p'', q solves the platform's problem, then for all $\tilde{q} \geq 0$:

$$V(q) - \sum_{\omega} p'(\omega) q(\omega) \geq V(\tilde{q}) - \sum_{\omega} p'(\omega) \tilde{q}(\omega)$$

$$V(q) - \sum_{\omega} p''(\omega) q(\omega) \geq V(\tilde{q}) - \sum_{\omega} p''(\omega) \tilde{q}(\omega)$$

Therefore, for all $\alpha \in [0, 1]$ and $\tilde{q} \geq 0$:

$$V(q) - \sum_{\omega} (\alpha p'(\omega) + (1-\alpha) p''(\omega)) q(\omega) \geq V(\tilde{q}) - \sum_{\omega} (\alpha p'(\omega) + (1-\alpha) p''(\omega)) \tilde{q}(\omega).$$

[19] Formally, we should impose market clearing say $\zeta' = q'/\bar{q}$ solves the consumers' problem. We skip this step to abbreviate notation.

[20] In fact, Proposition 2 of Galperti et al. (2023) identifies an explicit formula for p'. In particular, it can be chosen such that $p' \leq (\gamma_\pi + \gamma_u) \max_{\omega \in \Omega} \omega$. Therefore, if we take $M \geq (\gamma_\pi + \gamma_u) \max_{\omega \in \Omega} \omega$, it is guaranteed that $p' \in P$.

Finally, we argue F has a closed graph. Suppose $(p_n, q_n, x_n) \to (p, q, x)$, $(p'_n, q'_n, x'_n) \to (p', q', x')$, and $(p'_n, q'_n, x'_n) \in F(p_n, q_n, x_n)$. We want to show $(p', q', x') \in F(p, q, x)$. By Lemma A.1, we know the solution correspondence of $\mathcal{P}_q$ is upper-hemicontinuous, so x' is a solution to $\mathcal{P}_q$; To see q' solves the consumers problem, note that for all ω and $z \in [0, \bar{q}(\omega)]$:

$$q'_n(\omega)(p_n(\omega) + \sum_a u(a,\omega)x_n(a|\omega)) + (\bar{q}(\omega) - q'_n(\omega))r(\omega)$$

$$\geq z(p_n(\omega) + \sum_a u(a,\omega)x_n(a|\omega)) + (\bar{q}(\omega) - z)r(\omega)$$

By continuity we get:

$$q'(\omega)(p(\omega) + \sum_a u(a,\omega)x(a|\omega)) + (\bar{q}(\omega) - q'(\omega))r(\omega)$$

$$\geq z(p(\omega) + \sum_a u(a,\omega)x(a|\omega)) + (\bar{q}(\omega) - z)r(\omega)$$

Therefore, q' is optimal for the consumers given (p, x); To see under p', q solves the platform's problem, note that for all $\tilde{q} \geq 0$:

$$V(q_n) - \sum_\omega p'_n(\omega)q_n(\omega) \geq V(\tilde{q}) - \sum_\omega p'_n(\omega)\tilde{q}(\omega)$$

Since V is continuous by Lemma A.1, taking limit we get:

$$V(q) - \sum_\omega p'(\omega)q(\omega) \geq V(\tilde{q}) - \sum_\omega p'(\omega)\tilde{q}(\omega).$$

This completes the proof that $(p', q', x') \in F(p, q, x)$. □

With Lemma A.3, we can apply Kakutani's fixed-point theorem to F and conclude that F admits a fixed point. Therefore, a competitive equilibrium exists.

B Additional Material

B.1 Social Welfare

In the main text, we focused on a notion of welfare that excludes the merchant's profit (see Equation (1)). We take this stance because we want to focus on the trade between the platform and the consumers. Since we do not consider transfers between the merchant and the platform, it is natural that inefficiency can arise if the merchant's profit is taken into account.[21] That said, we show in this section that a parallel result of Proposition 3 holds if we define the efficiency notion to incorporate the merchant's profit.

Specifically, define the social welfare to be:

$$SW(q,x) = \sum_{a,\omega}\Big(v(a,\omega)+u(a,\omega)+\pi(a,\omega)\Big)x(a|\omega)q(\omega) + \sum_{\omega}\Big(\bar{q}(\omega)-q(\omega)\Big)r(\omega). \tag{B.1}$$

The notion of efficiency becomes:

Definition 7. *An allocation (q°, x°) is* ***constrained socially efficient*** *if it solves*

$$\begin{aligned} &\max_{q,x} && SW(q,x) \\ &\textit{such that} && q \le \bar{q}, \\ &\textit{and} && x \textit{ solves } \mathcal{P}_q. \end{aligned}$$

We have the following result, which extends Proposition 3 to this alternative notion of efficiency.

Proposition B.1. *Let (p^*, ζ^*, q^*, x^*) be an equilibrium of the competitive economy. If $\gamma_\pi > \gamma_u$ and, in addition, $r(\omega) \notin [\gamma_\pi\omega, (1+\gamma_\pi)\omega)$ for all ω, the equilibrium allocation (q^*, x^*) is constrained socially efficient. Otherwise, the equilibrium allocation can be socially inefficient.*

Proof. We prove the sufficiency of the proposition here. Let (p^*, ζ^*, q^*, x^*) be a competitive equilibrium. By Proposition 3, we know the equilibrium allocation (q^*, x^*) is constrained efficient. Moreover, following the argument in the proof of Proposition 3, we also know that $x^* = \hat{x}$, where $\hat{x}(\omega|\omega) = 1$ is the full-disclosure mechanism, is the unique optimal mechanism for the platform given any q. Therefore,

$$\begin{aligned} q^* \in &\ \arg\max_{q \le \bar{q}} \ \sum_{a,\omega}\Big(v(a,\omega)+u(a,\omega)\Big)\hat{x}(a|\omega)q(\omega) - \sum_{\omega} r(\omega)q(\omega) \\ = &\ \arg\max_{q\le\bar{q}} \ \sum_{\omega}\Big(\gamma_\pi\omega - r(\omega)\Big)q(\omega). \end{aligned}$$

The solution to this problem is $q^*(\omega) = \bar{q}(\omega)$ if $\gamma_\pi\omega > r(\omega)$, $q^*(\omega) = 0$ if $\gamma_\pi\omega < r(\omega)$, and $q^*(\omega) \in [0,1]$ if $\gamma_\pi\omega = r(\omega)$. The constrained socially efficient allocation (q°, x°)

[21] In the case where the platform can charge a service fee to the merchant, the platform can extract all the merchant's profit. Therefore, the platform's payoff essentially becomes $v(a,\omega)+\pi(a,\omega)$, which is equal to $\gamma_u u(a,\omega) + (1+\gamma_\pi)\pi(a,\omega)$. This is covered by our model, and particularly in this case, the planner's objective is the social welfare.

also features $x^\circ = \hat{x}$. Therefore, the solution of the planner's problem is $q^\circ(\omega) = \bar{q}(\omega)$ if $(1+\gamma_\pi)\omega > r(\omega)$, $q^\circ(\omega) = 0$ if $(1+\gamma_\pi)\omega < r(\omega)$, and $q^\circ(\omega) \in [0,1]$ if $(1+\gamma_\pi)\omega = r(\omega)$. When $r(\omega) \notin [\gamma_\pi\omega, (1+\gamma_\pi)\omega)$ for all ω, the equilibrium allocation (q^*, x^*) is also a solution to the planner's problem, and thus constrained socially efficient. □

Intuitively, if we take into account the merchant's profit, the inefficiency can arise from two sources. The first one is still the pooling externality. When $\gamma_\pi > \gamma_u$, the only optimal mechanism for the platform given any q is full disclosure, as argued in the proof of Proposition 3, so in this case the pooling externality disappears. The second one is a traditional externality. Since the platform does not take into account the merchant's payoff, it refuses to buy data when the price is high, even when trade is still socially optimal.

When the sufficient condition of the proposition is not satisfied, the equilibrium can be inefficient. Next we elaborate the two sources of externality using the example of Section 4.1. We will denote the constrained efficient allocation by (q°, x°). We also denote the equilibrium allocation in Case 1 (inefficiently low trade) by (q^*_L, x^*_L) and in Case 2 (inefficiently high trade) by (q^*_H, x^*_H). These are characterized in Section 4.1.

We first argue that in both cases, the social welfare of the equilibrium, $SW(q^*, x^*)$, is strictly lower than $SW(q^\circ, x^\circ)$. As before, this is originated from the pooling externality. Using the characterizations in Section 4.1, we can directly compute:

$$SW(q^\circ, x^\circ) = \bar{q}(1)(3+\gamma_u) + \bar{r}(\bar{q}(2) - \bar{q}(1)),$$

$$SW(q^*_L, x^*_L) = \bar{r} < SW(q^\circ, x^\circ),$$

$$SW(q^*_H, x^*_H) = \bar{q}(3+\gamma_u) + \bar{r}\max\{0, \bar{q}(2) - \frac{\bar{q}(1)}{\bar{r}}\} < SW(q^\circ, x^\circ).$$

The take is that, even if we measure efficiency using social welfare (Equation (B.1)), the equilibria are still suboptimal compared to the constrained efficient allocation. One may suspect that in Section 4.1, the inefficiency is an artifact that we did not take into account the merchant's profit, but as we highlight here, that is not the case.

In addition to the pooling externality, there is a new source of inefficiency: since in this case we have $(1+\gamma_\pi)\omega > \bar{r} > \gamma_\pi\omega = 0$, even (q°, x°) is not constrained socially efficient. The social welfare is maximized at $q^\bullet(\omega) = \bar{q}(\omega)$ and $x^\bullet(1|1) = 1, x^\bullet(1|2) = \frac{\bar{q}(1)}{\bar{q}(2)}$, which gives a social welfare of

$$SW(q^\bullet, x^\bullet) = \bar{q}(1)(\gamma_u + 1) + 2(\bar{q}(2) - \bar{q}(1)) > SW(q^\circ, x^\circ).$$

Therefore, the constrained efficient allocation is not constrained socially efficient. This additional gap is created by the fact that the profit of the merchant is not taken into account by the platform or the consumers. This is a traditional externality that can arise even without the informational friction discussed in our paper. For instance, consider the case where there is only one type of consumers $\omega = 1$ with $0 < r(1) < 1$. The platform's objective has $\gamma_u > \gamma_\pi = 0$. Then the constrained socially efficient allocation is $q^\bullet(1) = 1$, but the only equilibrium is no trade.

B.2 Complete Equilibrium Characterization for Section 4.1

In this section, we characterize the entire set of equilibria for our application from Section 4.1. We first note that in order for the platform's problem to admit a solution, we must have $p^*(1) \geq 0, p^*(2) \geq 0, p^*(1) + p^*(2) \geq \gamma_u$. Moreover, in order for the platform to trade, we must have $p^*(1) + p^*(2) = \gamma_u$.

Case 1: $2\bar{r} - 1 < \gamma_u < \bar{r}$**.** The unique equilibrium allocation is no trade, i.e., $q^*(\omega) = 0$ for all ω, and it is supported by a price vector p^* satisfying:

$$p^*(1) \in [0, \bar{r}] \qquad \text{and} \qquad p^*(2) \in [\max\{0, \gamma_u - p^*(1)\}, \bar{r}]. \tag{B.2}$$

Next we explain why this is the solution. Note that type-1 consumers are willing to sell only if $p^*(1) \geq \bar{r} > \gamma_u$. However, in this case we cannot have $p^*(1) + p^*(2) = \gamma_u$. Therefore, the unique equilibrium allocation is $q^*(\omega) = 0$. It can be supported by $x^*(\omega|\omega) = 1$. It can be checked that with the prices in (B.2), it is optimal for the consumers not to sell and for the platform not to buy. Any other price will induce some type of consumers to strictly prefer selling.

Case 2: $\gamma_u > 2\bar{r}$**.** There is a unique equilibrium such that: $p^*(1) = \gamma_u$ and $p^*(2) = 0$; $\zeta^*(1) = 1$ and $\zeta^*(2) = \min\{1, \frac{\bar{q}(1)}{\bar{r}\bar{q}(2)}\}$; $q^*(1) = \bar{q}(1)$ and $q^*(2) = \min\{\bar{q}(2), \frac{\bar{q}(1)}{\bar{r}}\}$; $x^*(1|1) = 1$ and $x^*(1|2) = \frac{q^*(1)}{q^*(2)}$. It can be easily checked that this is an equilibrium. To show uniqueness, note that since $\gamma_u > 2\bar{r}$, at least one type has a strict incentive to sell because $p^*(1) + p^*(2) \geq \gamma_u$. If type-1 has a strict incentive, we have $q^*(2) \geq \min\{\bar{q}(2), \frac{\bar{q}(1)}{\bar{r}}\}$ since $p^*(2) \geq 0$, but this requires $p^*(1) = \gamma_u, p^*(2) = 0$; if type-2 has a strict incentive, in order for the platform to be willing to buy, we must have $p^*(1) = \gamma_u, p^*(2) = 0$.

Case 3: $\bar{r} < \gamma_u \leq 2\bar{r}$**.** It can be easily checked that both equilibria of Case 1 and Case 2 continue to be an equilibrium in this case. Next we argue those are all possible equilibria. On one hand, for the equilibria with no trade, the price has to satisfy (B.2), otherwise some type will have a strict incentive to sell. On the other hand, given any equilibrium with trade, we must have $p^*(1) + p^*(2) = \gamma_u$. Moreover, we must have $q^*(1) > 0$, otherwise the platform is not willing to buy $\omega = 2$ at a positive price while type-2 consumers are not willing to sell at 0 price. If $q^*(1) > 0$, we must have $q^*(2) \geq \min\{\bar{q}(2), \frac{q^*(1)}{\bar{r}}\}$ because $p^*(2) \geq 0$ and the platform will choose $x^*(1|1) = 1, x^*(1|2) = \frac{q^*(1)}{q^*(2)}$. Since $q^*(2) > q^*(1)$, in order for the platform to be willing to buy, it must be the case that $p^*(1) = \gamma_u, p^*(2) = 0$. The unique equilibrium with trade then follows.

Case 4: $\gamma_u = \bar{r}$**.** In this case, the equilibria with no trade is the same as Case 1. The equilibria with trade satisfy $p^*(1) = \gamma_u = \bar{r},\ p^*(2) = 0$ with

$$0 < q^*(1) \leq \bar{q}(1),\ q^*(2) = \min\{\bar{q}(2), \frac{q^*(1)}{\bar{r}}\}.$$

It is easy to check these are equilibria. To see these capture all equilibria with trade, we can follow the same argument as Case 3 to derive the unique equilibrium price under trade: $p^*(1) = \gamma_u = \bar{r},\ p^*(2) = 0$. With these prices, since type-1 consumers are indifferent, any $0 \leq q^*(1) \leq \bar{q}(1)$ is optimal for them. $q^*(2)$ is then pinned down by the indifference condition of type-2 consumers.