Competitive Markets for Personal Data

Simone Galperti
UC, San Diego

Jacopo Perego
Columbia University

December 22, 2022

Preliminary Draft

Abstract

Personal data is an essential input for the digital economy. Yet, individuals often have limited control over its use and are rarely compensated for it. What inefficiencies does this status quo generate? Which market institutions could improve upon it? We study the competitive equilibria of an economy where platforms acquire consumers' data and use it to intermediate consumers and sellers. We find that granting consumers control over their data can backfire: It can lead to lower social welfare than when consumers are simply expropriated of their data. Even in the best-case scenario, allowing consumers to choose to whom they sell their data can only reach second-best efficiency. We show that to achieve full efficiency a large set of markets would need to open, which allows platforms and consumers to contract on how their data will be used. We comment on the difficulties of decentralizing in a practical way these efficient outcomes.

JEL Classification Numbers: C72, D82, D83

Keywords: Personal Data, Markets, Competitive Equilibrium, Externalities

We are thankful to Paolo Siconolfi and 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

Personal data has become a critical input in the business of many firms. For example, it allows marketers to access individual consumers using detailed information about their tastes, and to recommend them the right product at the right time. This has fueled a multi-billion dollars online-advertisement industry. In similar ways, personal data underlies the functioning of many e-commerce platforms, social-media networks, and, more in general, any recommendation-based platform system. Despite the important role that personal data play in the modern economy, however, we have not witnessed a corresponding development of large and competitive data markets where this personal data is traded. Today, the vast majority of personal data is either non-tradable, as it is a proprietary asset owned by large quasi-monopolistic firms, or it is traded in markets that lack transparency and barely involve the consumers (see, Federal Trade Commission, 2014).

The current arrangement of data markets can be a source of inefficiencies. First, since trade between firms is artificially restrained, markets cannot guarantee that personal data is used by the firms that can make the most profitable use of it. Second, since consumers have limited control over who owns their data, they cannot escape the externalities associated with how it is used. Third, since consumers are imperfectly compensated for their data, their incentive to supply better data may be inefficiently low. In recent years, many scholars have expressed concerns regarding the functioning of these markets and reasoned about which market institutions could be promoted to increase efficiency and fairness (e.g., Posner and Weyl (2018) and Seim et al. (2022)).

This paper formalizes these concerns and studies the extent to which different market arrangements affect the welfare of its participants. We model an economy with buyers, sellers, and intermediaries called platforms. Each buyer is uniquely identified by a "data record," which serves a dual purpose: It contains information that allows its owner to "access" the corresponding buyer (e.g., the record may contain this buyer's IP address, telephone number, or e-mail address) as well as information about her preferences (e.g., her age, gender, and education).[1] Platforms acquire a database of records to intermediate the corresponding buyers with a given set of sellers. Specifically, the platform acts as an information designer: It provides information about each buyer to its sellers to influence their product offers, which then affect the payoff of all participants in the economy.

[1] This definition of a data record is reminiscent of "marketing lists," which are one of the main information products traded in data-brokerage markets (see Federal Trade Commission (2014)).

We begin by studying an economy where buyers are passive: They have been expropriated of their data and platforms can trade it without their consent. We show that this boils down to a rather standard exchange economy and study its competitive equilibria. In equilibrium, each data record is allocated to the platform that uses it in the most profitable way. The converse is also true: any platform-optimal data allocation can be supported as an equilibrium. However, these equilibria ignore the external effects that platforms create on buyers and, as such, are generally inefficient.

Next, we enrich the previous economy by granting buyers ownership over their data: A buyer chooses whether to sell her record and to which platform. Perhaps surprisingly, we find that empowering buyers can backfire. Specifically, there can be equilibria whose welfare is lower than that the previous economy achieves. The reason is that, while buyers can now be directly compensated for the value their data create for the platforms, they cannot be fully compensated for the value their data create for other buyers. By selling her record, a buyer can change a platform's database and, thus, the way it is used: This can affect the payoff of other buyers. We show how this market failure relates to the "pooling externalities" identified in Galperti, Levkun, and Perego (2023) (henceforth GLP).

In the special case of an economy where all platforms behave as direct sellers—rather than intermediaries with complex objectives—we show that this particular source of inefficiency disappears and the equilibrium does a fairly good job at efficiently allocating data records. And yet, this economy is still inefficient The reason is that, once a platform has acquired a database, it can use it as it wants. The way data records are used is not contractible and this can create an inefficiency, which is akin to moral hazard.

To address these issues, we further enrich our economy following the work of Arrow (1969) and Laffont (1976). We create markets where buyers can trade not only the ownership of the data records but also how these records are to be used by the platform. We show that competitive equilibria in this economy exist, are efficient, and can be characterized as solutions to a grand information-design problem. The shortcoming of this economy is its realism: to achieve efficiency, a large number of markets need to exist. We comment on the challenges of decentralizing these equilibria with simple institutions.

In conclusion, we believe that our paper contributes to the burgeoning literature on data markets in two ways. First, we introduce a framework to study data markets that is both classical and tractable. It is classical because it is rooted in the general-equilibrium tradition. It is tractable because it exploits the recent progress of the information-design literature.[2] Second,

[2]Relatedly, we show that the equilibria of some of the economies that we consider are solutions to a "grand"

we analyze some of the market institutions that have been proposed in practice—chiefly, granting buyers control over their data and a compensation for their use—and study their effects on welfare and possible shortcomings.

Related Literature. **TBW**

2 Model

This section introduces the main building blocks of the economies that we will analyze in this paper. For concreteness, we present the model in the context of the e-commerce industry. There are buyers (*she*), sellers (*he*), and platforms (*it*). The platforms intermediate the interactions between buyers and sellers using data records. Our analysis applies more broadly to other types of intermediaries that information to influence the behavior of other agents.

Denote by K the set of sellers and assume they are partitioned among a set of I intermediaries. Let $K_i \subseteq K$ be the sellers who participate in platform i. The partition $\{K_i\}_{i \in I}$ is exogenously given. Each seller $k \in K_i$ chooses an action in A^k, which can be interpreted as the price, quality, or variety of his product. We denote the set of action profiles of the sellers active on platform i by $A_i := \prod_{k \in K_i} A^k$. All these sets are assumed to be finite.

There is a continuum population of buyers. Each buyer demands at most one of the products sold by the sellers (unit demand). Her preference is pinned down by a payoff type $\omega \in \Omega$. We denote by $\bar{q}(\omega) \geq 0$ the mass of buyers of type ω and assume that Ω is finite. A buyer of type ω obtains a payoff $g_i(a_i, \omega) \in \mathbb{R}$ when platform i's sellers choose action profile a_i. In this case, given a_i and ω, platform i and seller k obtain (real-valued) payoffs of $u_i(a_i, \omega)$ and $\pi_k(a_i, \omega)$, respectively.

Each buyer is uniquely identified by a *data record*, which will be the commodity traded in the economies that we consider. A data record can be owned either by a platform or by the corresponding buyer. A data record serves a dual purpose. First, it contains information that allows a platform to "access" this buyer (e.g., the record may contain this her IP address, telephone number, or e-mail address). Access means that the platform can intermediate the interaction between this buyer and its sellers K_i. Second, the record contains information about this buyer's type ω (e.g., her age, gender, and education). We make two simplifying assumptions:

information-design problem. This suggests that the many tools that have been developed by the literature in recent years to characterize these problems (see, e.g., Kamenica (2019) and Bergemann and Morris (2019) for a review) could be repurposed to further characterize equilibria of data markets.

First, we assume that ownership of a record is exclusive. Namely, no two platforms can own the same record. Second, the record fully reveals the buyer's payoff type ω. Because of the latter, we call ω records those data records that identify buyers whose payoff type is ω.

A platform is an intermediary that provides its sellers with information about the buyers. Specifically, suppose that platform i owns a database of records $q_i \in \mathbb{R}_+^\Omega$. Sellers in K_i know q_i and can sell to the corresponding buyers. However, sellers cannot tell these buyers apart. They have no information about them except for the frequencies implied by q_i. The platform acts as an information designer: It conveys information about each buyer's record to its sellers so as to influence their actions $a_i = (a^k)_{k \in K_i}$. To do so, the platform commits to an information structure that maps the set of types Ω into random signals. By standard arguments (e.g., see Bergemann and Morris (2016)), this information-design problem can be represented as choosing a recommendation mechanism $x_i : A_i \times \Omega \to \mathbb{R}_+$ subject to incentive-compatibility and feasibility constraints:

$$
\begin{aligned}
\mathcal{P}_i : \qquad & \max_{x_i : A_i \times \Omega \to \mathbb{R}_+} \sum_{\omega \in \Omega, a_i \in A_i} u_i(a_i, \omega) x_i(a_i, \omega) \\
& \text{s.t. for all } k \in K_i, \text{ and } a^k, \hat{a}^k \in A^k \\
& \sum_{\omega,\, a^{-k} \in \prod_{k \neq k' \in K_i} A^{k'}} \left(\pi_k(a^k, a^{-k}, \omega) - \pi_k(\hat{a}^k, a^{-k}, \omega) \right) x_i(a^k, a^{-k}, \omega) \geq 0 \qquad (1) \\
& \text{and for all } \omega \in \Omega, \\
& \sum_{a_i \in A_i} x_i(a_i, \omega) \leq q_i(\omega) \qquad (2)
\end{aligned}
$$

We call x_i obedient if it satisfies (1). The obedience constraint requires that the sellers find it optimal to follow their recommended action conditional on the information it conveys. The resource constraint in (2), instead, requires that some recommendation be sent for every record in the database. For each i and q_i, fix once and for all $x^*_{q_i}$—a solution to $\mathcal{P}_i$—and denote the platform's payoff given database q_i by

$$
U_i(q_i) := \sum_{a_i, \omega} x^*_{q_i}(a_i, \omega) u_i(a_i, \omega).
$$

By standard arguments, $U_i(q_i)$ is continuous and concave in q_i and represents the platform's preference over databases. The payoff of ω buyers whose record belongs to q_i is given by:

$$
G_{i\omega}(q_i) := \sum_{a_i} \frac{x^*_{q_i}(a_i, \omega)}{q_i(\omega)} g_i(a_i, \omega). \qquad (3)
$$

In this expression, $x^*_{q_i}(a_i, \omega)/q_i(\omega)$ is the conditional probability that recommendation a_i is sent given ω. Finally, note that unlike U_i, the buyer's payoff $G_{i\omega}$ may fail to be continuous in q_i, as $x^*_{q_i}$ can be discontinuous in q_i.

We postpone the discussion of the model to Section 6. Here, we briefly comment on the role of data records, discuss an important special case, and offer a few examples.

Data Records. In this paper, data records are *rival goods* and we study markets where these goods are traded. For this to be the case, we made two important assumptions. First, we assumed that ownership of a record is necessary to access the corresponding buyer. Implicitly, we consider situations where accessing a specific buyer without her data record is prohibitively costly due to the large size of the population. Second, we assumed that ownership is exclusive: At most one platform can own a record at each given time. Both assumptions are strong and may not be descriptive of all data markets. Yet, our data records are reminiscent of "marketing lists," which are the main information product traded in the data-brokerage industry (see Federal Trade Commission (2014)). Moreover, both assumptions can be partially weakened, as discussed in Section 6. In the same section, we will discuss the other stylized assumptions that we made.

A Special Case of the Model. Our model accommodates the special case of a "platform" that sells directly to buyers. To do so, fix i and assume $|K_i| = 1$ and $u(a_i, \omega) = \pi_k(a_i, \omega)$ for all a_i and ω. In this case, the incentives of i and k are perfectly aligned. Thus, obedience constraints (1) are trivially satisfied and the problem of the platform becomes a decision problem, rather than an intermediation one. Equivalently, we can think of this platform as a seller with direct, i.e. non-intermediated, access to the buyers. An important special case of our model (see Section 4) is the one where *all* platforms are sellers in the sense just explained. We will refer to this as the no-intermediation case. Another special case that is worth mentioning is that of a platform with $K_i = \varnothing$. We interpret this platform as a pure data-broker. $\triangle$

An Example for the Payoffs. As an example, suppose sellers sell an homogeneous product at zero marginal cost and compete on price. Let ω be the buyer's willingness to pay for the product and a^k the price set by seller k. Then, $g_i(a_i, \omega) = \max_{k \in K_i}\{\omega - a^k, 0\}$. This payoff incorporates an outside option (in this case 0) that the buyer can exercise upon observing the sellers' action. Seller k's payoff is $\pi_k(a_i, \omega) = a^k$ if it charges the lowest price (ties broken at random) and zero otherwise. Finally, the platforms' payoff $u_i(a_i, \omega)$ is proportional to the sum if the seller's profits $\sum_{k \in K_i} \pi_k(a_i, \omega)$, as if the platform charged a transaction fee on sellers' profits. $\triangle$

3 A Baseline Economy

The goal of this paper is to study how different market designs affect the welfare of the participants of the economy. We begin our analysis with a simple market organization. We consider an exchange economy where only platforms can trade data records. Buyers have been expropriated of their records and cannot participate in this market. This is a salient benchmark to study as it resembles the current arrangement of many real-world data-brokerage markets, where data records are collected without the consent or even the knowledge of the corresponding buyers (see Federal Trade Commission (2014) and Seim et al. (2022)).

We refer to this baseline economy as $\mathcal{E}_1$. In this economy, platforms are endowed with some of the data records and their preferences are quasilinear in money. Taking prices as given, the problem of platform i is then equivalent to choosing a database q_i to solve

$$\max_{q_i \in \mathbb{R}_+^{\Omega}} U_i(q_i) - \sum_{\omega} p(\omega) q_i(\omega), \tag{4}$$

where $p(\omega)$ denotes the market price for ω records.

A *data allocation* in economy $\mathcal{E}_1$ is a profile $q \in \mathbb{R}_+^{\Omega \times I}$, specifying a database $q_i \in \mathbb{R}_+^{\Omega}$ for each platform i. A data allocation is feasible if $\sum_i q_i(\omega) \leq \bar{q}(\omega)$ for all ω. A competitive equilibrium is defined as follows:

Definition 1 (Equilibrium for $\mathcal{E}_1$). *A price vector $p^* \in \mathbb{R}^{\Omega}$ and a feasible data allocation $q^* \in \mathbb{R}_+^{\Omega \times I}$ constitute an equilibrium of economy $\mathcal{E}_1$ if:*

1. *For each platform i, q_i^* maximizes (4) taking p^* as given;*
2. *All markets clear: for all $\omega \in \Omega$, $p^*(\omega)\big(\bar{q}(\omega) - \sum_i q_i^*(\omega)\big) = 0$.*

The market clearing condition simply requires that either there is no excess supply of ω records or, if there is, the price of ω records must be zero.

Economy $\mathcal{E}_1$ is a standard exchange economy, with the only peculiarity that the goods being traded are data records. Equilibria of this economy exist and they maximize the unweighted sum of platforms' payoffs. That is, if (p^*, q^*) is an equilibrium of $\mathcal{E}_1$, the data allocation q^* solves the following social-planner problem:

$$\begin{aligned} \mathcal{TB} \quad : \quad & \max_{q \in \mathbb{R}_+^{\Omega \times I}} \sum_i U_i(q_i) \\ & \text{s.t.} \sum_i q_i(\omega) \leq \bar{q}(\omega) \text{ for all } \omega. \end{aligned}$$

In equilibrium, markets optimally allocate the data records among the platforms that make the most profitable use of them. By quasilinearity, this also implies that equilibria are Pareto optimal. The converse is also true: Any solution q^{**} to $\mathcal{TB}$ can be supported as an equilibrium of this economy $\mathcal{E}_1$. These two results are a standard instance of the first and second Welfare Theorems for an appropriately defined notion of welfare, namely $\mathcal{TB}$. However, it is evident that, while the equilibrium allocation is optimal from the point of view of the platforms, it ignores the external effects that platforms create on buyers by using their data. Indeed, $\mathcal{TB}$ stands for "third-best" efficient to indicate the fact that equilibria in this economy are, in general, not "socially" efficient. We will come back to this point in the next sections. Before that, our first result summarizes the aforementioned properties.

Proposition 1. *Economy $\mathcal{E}_1$ admits an equilibrium. Moreover, the equilibrium data allocation q^* is third-best efficient, namely, it solves $\mathcal{TB}$. Conversely, for any q^{**} that solves $\mathcal{TB}$, there is a price p^{**} such that (p^{**}, q^{**}) is an equilibrium.*

To prove this result, we take advantage of the special linear structure of this economy, which is inherited by the fact that each $\mathcal{P}_i$ is an information-design problem. Indeed, by combining each platform's $\mathcal{P}_i$, we obtain a "grand" information-design problem $\mathcal{P}$:

$$\begin{aligned}
\mathcal{P} \quad : \quad & \max_{x_1,\dots,x_I} \sum_i \sum_{a_i,\omega} u_i(a_i,\omega) x_i(a_i,\omega) \\
& \text{s.t. for all } \omega, \ \sum_i \sum_{a_i} x_i(a_i,\omega) \leq \bar{q}(\omega) \qquad (5) \\
& \text{and for all } i, k, a_i^k, \hat{a}_i^k \\
& \sum_{\omega\in\Omega, a_i^{-k}} \left(\pi_k((a_i^k, a_i^{-k}),\omega) - \pi_k((\hat{a}_i^k, a_i^{-k}),\omega)\right) x_i((a_i^k, a_i^{-k}),\omega) \geq 0.
\end{aligned}$$

This grand information-design problem $\mathcal{P}$ is equivalent to $\mathcal{TB}$ in the sense that every solution $(x_i^{**})_{i\in I}$ to $\mathcal{P}$ implies a solution $(q_i^{**})_{i\in I}$ to $\mathcal{TB}$, and vice versa. In problem $\mathcal{P}$, a social planner chooses the recommendation mechanism on behalf of each platform to maximize the sum of their payoffs. A solution $(x_i^{**})_{i\in I}$ to $\mathcal{P}$ pins down, not only how data records are allocated—namely, $q_i^{**}(\omega) := \sum_a x_i^{**}(a,\omega)$ for all ω and i— but also how they are used by each platform. In the proof of Proposition 1, we show that a solution to $\mathcal{P}$ exists and that the implied data allocation solves $\mathcal{TB}$. Moreover, we show that, as it is standard, the dual variables of constraints (5) in $\mathcal{P}$ constitute a price vector that supports q^{**} as a competitive equilibrium of the economy.

In summary, $\mathcal{P}$ constitutes a tractable and compact way to determine how data are allocated, how they are used, and (through its dual) what their market price will be. The fact that $\mathcal{P}$

is an information-design problem can be useful in itself. It suggests that the many tools that have been developed by the literature in recent years to characterize these problems (see, e.g., Kamenica (2019) and Bergemann and Morris (2019) for a review) can be applied to study the equilibrium of data markets.

To illustrate this, we can further characterize the equilibrium. First, the equilibrium price p^* is tightly related to the values that each platform i individually derives from using each record in its database q_i. Denote this value by $v^*_{q_i}(\omega)$, which is the dual variable of constraint (2) in problem $\mathcal{P}_i$. These values and how they depend on the database q_i have been analyzed by GLP. We have the following result.

Corollary 1. *Let (q^*, p^*) be an equilibrium of economy $\mathcal{E}_1$. For every i, there is an individual value $v^*_{q_i^*}(\omega)$ that equals $p^*(\omega)$.*

This means that, as it is standard in these simple economies, all platforms that use ω records must derive the same value from them, even if they use them in different ways and for different objectives. Second, the price of data records depends on their aggregate supply $\bar{q}$ in a simple fashion. Since $\mathcal{P}$ is an information design problem, it follows from Proposition 3 in GLP together with Proposition 1 above that the scarcer ω records are (i.e., the lower $\bar{q}(\omega)$ is), the higher their equilibrium price $p^*(\omega)$.

4 Giving Buyers Ownership of Their Data Records

The previous section established that equilibria of $\mathcal{E}_1$ maximize the sum of the platforms' payoff. However, they may still be inefficient from a broader societal perspective. This is because the use of data records can (and, in general, will) directly affect the buyers' payoffs in ways that platforms do not internalize. In this section, we enrich economy $\mathcal{E}_1$ with new institutions that aim at correcting these inefficiencies. We show that, while these institutions may help improve efficiency, they do not entirely resolve the problem and, in fact, could worsen it.

We consider an economy, called $\mathcal{E}_2$, where each buyer has ownership over her data records and decides to which platform to sell it (if any). Since all buyers of the same type ω are identical, we will treat them as a single representative buyer who is endowed with the entire stock of ω records, namely $\bar{q}(\omega)$. We will refer to this agent as the "representative ω buyer." As before, there is a market for each type of data record. Unlike before, records are supplied endogenously by the buyers.

This economy is a natural benchmark to study, as it captures some of the salient institutions (chiefly, buyers have property rights and are compensated for their data) that some authors have advocated for as tools that could improve the efficiency of data markets (see, e.g., Posner and Weyl, 2018; Seim et al., 2022).

In economy $\mathcal{E}_2$, a *data allocation* is a profile $q \in \mathbb{R}_+^{\Omega+\Omega\times I}$, which is richer than before as it includes data that buyers may decide to keep for themselves. Specifically, a data allocation specifies a database $q_0 \in \mathbb{R}_+^{\Omega}$, where each $q_0(\omega)$ denotes the quantity of ω records that the representative ω buyer keeps for herself; for each i, it also specify a database $q_i \in \mathbb{R}_+^{\Omega}$, where $q_i(\omega)$ denotes the quantity of ω records allocated to platform i. We assume that each record that the representative buyer keeps for herself gives her a unit payoff of $r_\omega \geq 0$, which we interpret as the value of preserving her privacy. Since $r_\omega \geq 0$, it is without loss of generality to focus attention on data allocations that are feasible in the strict sense that, for all ω, $q_0(\omega) + \sum_i q_i(\omega) = \bar{q}(\omega)$ (i.e. no data is disposed of).

In this economy, the platform's problem is identical to that described in Section 3. Namely, platform i chooses q_i, where $q_i(\omega)$ represents the demanded quantity of ω records. Denoting by $p(\omega)$ the market price for ω records, platform i chooses q_i to solve the problem in (4), just as before.

The novelty is that in this economy the supply of records is endogenous. The representative ω buyer chooses how many of her records to keep for herself and how many to sell and to which platforms. We denote this choice by $q_\omega = (q_\omega^0(\omega), q_\omega^1(\omega), \ldots, q_\omega^I(\omega)) \in \mathbb{R}_+^{1+I}$, which has the following interpretation: $q_\omega^0(\omega)$ denotes the quantity her records that she keeps for herself; $q_\omega^i(\omega)$ denotes the quantity of ω records she suppply to platform i.

For a given data allocation $q \in \mathbb{R}_+^{\Omega+\Omega\times I}$, we denote by $q_\omega \in \mathbb{R}_+^{1+I}$ the dimensions corresponding to ω records.[3] Then, the total payoff of the representative ω buyer can be written as:

$$U_\omega(q_\omega, (q_{\omega'})_{\omega'\neq\omega}) = r_\omega q_\omega^0(\omega) + \sum_i q_\omega^i(\omega) G_{i\omega}(q_\omega^i(\omega), (q_{\omega'}^i(\omega'))_{\omega'\neq\omega}).$$

The first term on the right-hand side is the utility of self-consumption. The second term is the sum of the "external effects" that are imposed by the platforms on this representative buyer (recall equation (3)). Buyers' preferences are assumed to be quasilinear in money and, thus, the problem of a representative buyer can be written as follows:

$$\max_{q_\omega \in \mathbb{R}_+^{1+I}} U_\omega(q_\omega, (q_{\omega'})_{\omega'\neq\omega}) + p(\omega) \sum_i q_\omega^i(\omega) \tag{6}$$

[3] We use the notation q_ω to specifically refer to the choice of the representative ω buyer. Given a data allocation q, market clearing will require that $q_i(\omega) = q_\omega^i(\omega)$ and, therefore, $q = (q_\omega)_{\omega\in\Omega}$.

In this equation, the second term is the price $p(\omega)$ that this buyer gets for each data record she sells to a platform. Note that, importantly, this buyer has no control over the quantity of records that other buyers sell to the platforms. This means that, in this economy, the market allocation is not determined exclusively by a market mechanism (i.e., it is not entirely mediated by prices). In other words, the strategic interactions among buyers matter and the competitive equilibrium notion is that of Nash-Walras (Ghosal and Polemarchakis (1997)).

Definition 2 (Equilibrium for $\mathcal{E}_2$)**.** *A nontrivial price vector $p^* = (p^*(\omega))_{\omega\in\Omega}$ and a feasible data allocation $q^* = (q_0^*, (q_i^*)_{i\in I}) \in \mathbb{R}_+^{\Omega+\Omega\times I}$ constitute an equilibrium for economy $\mathcal{E}_2$ if:*

1. *For each ω, $q_\omega^* \in \mathbb{R}_+^{1+I}$ maximizes (6) taking as given prices p^* and $(q_i^*(\omega'))_{i\in I,\omega'\neq\omega}$.*
2. *For each platform i, q_i^* maximizes (4) taking taking as given prices p^*.*
3. *All markets clear: for all ω, $p^*(\omega)\Big(\bar{q}(\omega) - \sum_i q_i^*(\omega) - q_0^*(\omega)\Big) = 0$.*

What are the welfare properties of this equilibrium? Since we allowed buyers to participate in the markets for data records, it is natural to wonder whether the equilibrium allocation maximizes the sum of platforms' *and* buyers' payoffs. For any data allocation q, let us denote this notion of welfare by $W(q) = \sum_i U_i(q) + \sum_\omega U_\omega(q)$.[4] We say that a feasible allocation that maximizes welfare W is "second-best" efficient. Namely, it solves

$$
\begin{aligned}
\mathcal{SB} \quad : \quad & \max_{q\in\mathbb{R}_+^{\Omega+\Omega\times I}} \sum_{i\in I} U_i(q_i) + \sum_{\omega\in\Omega} U_\omega(q) \\
& \text{s.t. for all } \omega \\
& q_0(\omega) + \sum_i q_i(\omega) = \bar{q}(\omega)
\end{aligned}
$$

Are equilibria in $\mathcal{E}_2$ second-best efficient? In general, the answer is negative. Since platforms act as intermediaries, the presence of nontrivial obedience constraints (1) makes the optimal mechanism $x_{q_i}^*$ "non-separable" in ω. This means that the recommended actions for ω records $x^*(\cdot,\omega)$ depend on the actions recommended for ω' records, and vice versa. This implies that the payoff $G_{i\omega}(q_i)$ that ω buyers experiences when platform i uses her data is affected by $q_i(\omega')$, the quantity of records other than ω. Even though buyers are allowed to trade their data records, there are no markets where ω buyers can purchase claims regarding the quantity of ω' records that platform i should buy. Intuitively, this can lead to inefficiencies.

[4] Note that in this calculation we ignore the sellers' welfare. We explain the reason for this choice at the end of Section 6.

Proposition 2. *Let (p^*, q^*) be an equilibrium of economy $\mathcal{E}_2$. The sum of equilibrium payoffs of platforms and buyers is lower than that achieved by the solution to $\mathcal{SB}$. Moreover, the equilibrium welfare $W(q^*)$ in $\mathcal{E}_2$ may be lower than that induced by any equilibrium of $\mathcal{E}_1$.*

The first part of this result formalizes the sense in which this economy is inefficient. This inefficiency appears related to that of economies with incomplete markets a la Greenwald and Stiglitz (1986) and Geanakoplos and Polemarchakis (1986). The second part of this result shows that, perhaps counter-intuitively, allowing buyers to trade their records can backfire and lower the welfare of the economy below the levels of $\mathcal{E}_1$. This is reminiscent of older results in general equilibrium showing that financial innovation in economies with incomplete markets may make agents worse off (see Hart (1975) and Elul (1995)).

We present a stylized example of an economy that illustrates that giving buyers control over their data can backfire. That is, there is an equilibrium in $\mathcal{E}_2$ whose welfare is lower than that of $\mathcal{E}_1$. The economy in the example is meant to be simple rather than realistic. It is based on Bergemann et al. (2015) and builds on results from Galperti et al. (2023).

Example. There is a single platform $I = \{1\}$ with a single seller $K_1 = \{k\}$. There are two types of buyer $\Omega = \{1, 2\}$, and ω denotes her willingness to pay for the seller's product. Let $\bar{q}(2) > \bar{q}(1)$. Let $A^k = \Omega$ and note that $A_1 = A^k$. The seller maximizes profits, i.e., $\pi_k(a_1, \omega) = a_1 \mathbb{1}(\omega \geq a_1)$. The platform maximizes buyers' surplus, i.e, $u_1(a_1, \omega) = \max\{\omega - a_1, 0\}$. Let $g_1(a_1, 1) = \max\{1 - a_1, 0\}$ and $g_1(a_1, 2) = \alpha \max\{2 - a_1, 0\}$. Suppose that low types care more about their privacy than the high types, namely $r_1 = \beta$ and $r_2 = 0$. We assume $\alpha > \beta > 1$.

The equilibrium allocation q^* of $\mathcal{E}_1$ is trivial and unique: the platform obtains all the data. By Proposition 1, the (unique) equilibrium price vector can be computed from the dual of $\mathcal{P}$, which in this example equals $\mathcal{P}_1$. These prices are $p^*(1) = 1$ and $p^*(2) = 0$ (see Appendix E in GLP). The platform chooses $x^*_{\bar{q}}$ as to create two segments: The first one pools all low-type buyers and $\bar{q}(1)$ high-type buyers. For this segment, it recommends the seller to charge $a_1 = 1$; The second segment contains all the remaining $(\bar{q}(2) - \bar{q}(1))$ high-type buyers. For this segment, the platform's recommendation is to charge $a_1 = 2$. Given this, welfare is as follows: The low-type buyers earn nothing; a share $\bar{q}(1)$ of high-type buyers earn α; the platform earns $\bar{q}(1)$. Therefore, the sum of buyers and platform's payoff is $W(q^*) = (1 + \alpha)\bar{q}(1)$.

Now consider $\mathcal{E}_2$. As a preliminary observation, note that $p^*(1)$ that results from the dual of $\mathcal{P}_1$ also equals the platform's willingness to pay for records of type 1 (see GLP). Moreover, we know from the previous paragraph that $p^*(1) = 1$ if $q^*(1) < q^*(2)$ and otherwise $p^*(1) \leq$

1 by the scarcity principle (see GLP, Proposition 3). Since $r_1 > 1$, it follows that in any equilibrium (q^{**}, p^{**}) of $\mathcal{E}_2$ we must have $q_0^{**}(1) = \bar{q}(1)$ and so $q_1^{**}(1) = 0$—that is, low-type buyers hold on to their data. Given this, the optimal mechanism for the platform given q_1^{**} is trivial and leads to zero surplus for each buyer of type 2, which implies $p^{**}(2) = 0$. At this price, buyers of type 2 are indifferent between selling and keeping their records, so the allocation $q_1^*(2) = \bar{q}(2)$ and $q_0^*(2) = 0$ solve the buyer's problem. To complete the equilibrium characterization, we need to find $p^{**}(1)$ that supports this candidate q^{**} leading to market clearing. To this end, we can let $p^{**}(1) = 1 + \frac{\beta}{2}$. Given this, we have established the existence of an equilibrium and that any equilibrium leads to the following welfare: The low-type buyers get $\beta\bar{q}(1)$; the high-type buyers get 0; the platform gets 0. Therefore, $W(q^{**}) = \beta\bar{q}(1)$ and it is strictly lower than $W(q^*)$.

The reason why giving buyers control over their data reduces welfare is simple. Type-1 records help the platform generate a positive surplus for type-2 buyers using the mechanisms like the $x_{\bar{q}}^*$ described above. However, the markets in $\mathcal{E}_2$ do not allow buyers of type 2 to compensate buyers of type 1 for the positive externality that they exert, at least not at a level that would exceed their intrinsic desire to hold on to their data (i.e., r_1). As a result, the platform is denied the use of such data, which hurts type-2 buyers more than it benefits type-1 buyers. $\triangle$

Before proceeding, it is important to note that another potential issue with economy $\mathcal{E}_2$ is that equilibrium existence is not guaranteed in general. The reason is that the component of the buyer's payoff $G_{i\omega}(q_i)$ is not continuous in q_i, since $x_{q_i}^*$ can change discontinuously in q_i.

These issues that characterize economy $\mathcal{E}_2$ could be addressed in two ways. One is to further enrich the economy by opening new markets, a solution we will consider in the next section. Another is to consider a special case of this economy. This special case is instructive as it will help us understand that the source of the inefficiency is deeper than what one may think. It is instructive also because it will reveal which market should be opened in the next section to achieve efficiency.

The No-Intermediation Case. Recall from the very end of Section 2 that a special case of a platform is the one where $|K_i| = 1$ and $u_i = \pi_k$. This platform has a single seller with perfectly aligned incentives. Therefore, the obedience constraints (1) drop. In other words, the platform is the seller and faces a decision (rather than an intermediation) problem. For the next result, we specialize our model to the case where all the platforms are in this situation. We refer to this as the *no-intermediation* case.

In this special case, the platform's optimal mechanism $x_{q_i}^*$, i.e., the one that solves $\mathcal{P}_i$, is "separable" in the sense that the optimal action for ω, $x^*(\cdot,\omega)$, does not depend on the optimal action for ω', $x^*(\cdot,\omega')$. Therefore, it does not depend on the quantity of ω' records in the database q_i.[5] As a consequence, the representative buyer's payoff U_ω only depends on her own choice q_ω and not on the choices of other buyers $(q_{\omega'})_{\omega'\neq\omega}$. In other words, in the no-intermediation case, the payoff of the representative ω buyer is not affected by which records of type other than ω a platform acquires. The only way in which platform i affects the payoff of this buyer is through the extensive margin $q_i(\omega)$—how many ω records platform i acquires. This margin is mediated by a market in which this buyer can participate. Intuitively, this should make equilibria of $\mathcal{E}_2$ efficient in the sense of $\mathcal{SB}$. The next result formalizes this intuition.

Corollary 2. *In the no-intermediation case, any equilibrium data allocation of economy $\mathcal{E}_2$ is second-best efficient, namely, it solves $\mathcal{SB}$.*

When platforms do not act as intermediaries, the prices in $\mathcal{E}_2$ are enough to make platforms internalize the effects that their use of the data generates on buyers. Two final comments are in order.

First, while we can Corollary 2 as "good news," it comes with the caveat that, in many modern digital markets, platforms have often complex objectives that leads them to optimally withhold some information from the sellers (see, e.g., Xu and Yang, 2022). As discussed in GLP, this usually leads to "pooling externalities," which would then generate the problems described in Proposition 2.[6]

Second, even in the stylized case of the no-intermediation economy, the equilibrium only reaches what we called "second-best" efficiency. At this point, the reader may wonder what could prevent the economy from doing better than this. We address this question in the next section.

5 Trading Externalities of Data Use

The economy discussed in the previous section, even under the simplifications introduced by the no-intermediation case, cannot do better than second-best. The reason is that, once a plat-

[5] To see this note that, conditional on ω, the platform chooses $x^*(\cdot,\omega)$ to maximize $\sum_{a_i\in A_i} u_i(a_i,\omega)x_i(a_i,\omega)$ subject to $\sum_{a_i\in A_i} x_i(a_i,\omega) = q_i(\omega)$. This problem is independent of ω' and, therefore, of $q_i(\omega')$.

[6] Pooling externalities are likely not the only way in which the problems described in Proposition 2 may arise. We conjecture that "learning externalities" would generate inefficiencies that would further contribute to lowering social welfare below second-best. These externalities are discussed in Choi et al. (2019), Bergemann et al. (2022), Acemoglu et al. (2021), and Ichihashi (2021).

form has acquired its database, it can use it as it wants. That is, given q_i, mechanism $x^*_{q_i}$ is chosen to maximizes platform's payoff only. Since the way data records are used is not contractible, this can create inefficiency (which is akin to moral hazard). The representative ω buyer may be even willing to give a platform more data if only it used them differently. Unfortunately, such a contract cannot be written and there are no markets where claims for how data records should be used are traded.

In this section, we further enrich our economy following standard ways of modeling competitive economies with externalities (e.g., Arrow (1969) and Laffont (1976)).[7] We will refer to this economy as $\mathcal{E}_3$. Specifically, we assume that buyers are allowed to trade the way their records are used. As before, the representative ω buyer choose the quantity $q^0_\omega(\omega)$ of ω records that she wants to keep for herself. Unlike before, she also chooses the quantity of records that she is willing to sell to platform i and their intended use, denoted $\hat{x}_i(a_i,\omega)$ for all a_i. Likewise, platform i chooses $x_i(a_i,\omega)$, which is the quantity of ω records she is willing to acquire and use for recommendation a_i.

In this economy, it is useful to distinguish between a data allocation, which is as before a $q \in \mathbb{R}_+^{\Omega+\Omega\times I}$, and a data use, which is a profile of recommendation mechanisms $(x_i)_{i\in I}$. Recall, however, that it is without loss to focus on allocations that are feasible in the sense that $q_0(\omega) + \sum_i q_i(\omega) = \bar{q}(\omega)$. Then, a data use $(x_i)_{i\in I}$ pins down an allocation q (namely, $q_i(\omega) = \sum_{a_i} x_i(a_i,\omega)$ for all i,ω) as long as the former is feasible, i.e., if $\sum_{i,a_i} x_i(a_i,\omega) \le \bar{q}(\omega)$. In the following, we will sometimes talk about a feasible data use $(x_i)_{i\in I}$ leaving implicit the data allocation it implies.

The price system in this economy is richer than the one of the previous sections. As before, there is a price $p(\omega)$ for each "physical" ω record traded on the market. In addition, for all i, a_i and ω, there is a price $p_i(a_i,\omega)$ for when the platform i uses record ω to induce recommendation a_i. This price $p_i(a_i,\omega)$ should be interpreted as a unit transfer that goes from platform i to the representative ω buyer.

In this economy, the problem of platform i is

$$\begin{aligned} \max_{x_i:A_i\times\Omega\to\mathbb{R}_+} & \sum_{a_i,\omega} \big(u_i(a_i,\omega) - p(\omega) - p_i(a_i,\omega)\big)x_i(a_i,\omega) \\ & \text{s.t. for all } k \in K_i, \text{ and } a^k, \hat{a}^k \in A^k \\ & \sum_{\omega\in\Omega, a^{-k}\in A_i^{-k}} \big(\pi_k(a^k,a^{-k},\omega) - \pi_k(\hat{a}^k,a^{-k},\omega)\big)x_i(a^k,a^{-k},\omega) \ge 0 \end{aligned} \tag{7}$$

Note that this problem boils down to (4) when $p_i(a_i,\omega) = 0$ for all a_i and ω.

[7] See also, Bonnisseau et al. (2022).

The problem of the representative ω buyer is

$$\max_{q^0_\omega(\omega),(\hat{x}_i(a_i,\omega))_{i\in I,a_i\in A_i}} \big(r_\omega - p(\omega)\big)q^0_\omega(\omega) + \sum_{i,a_i}\big(g_i(a_i,\omega) + p(\omega) + p_i(a_i,\omega)\big)\hat{x}_i(a_i,\omega). \tag{8}$$

Note that even when $p_i(a_i,\omega) = 0$ for all a_i, ω, and i, this problem does not boil down to (6). The difference is that, through the choice of $\hat{x}_i$, the buyer can determine the external effect G_i that platform i creates on her.

We now define the (Lindahl) equilibrium for economy $\mathcal{E}_3$.

Definition 3 (Equilibrium for $\mathcal{E}_3$)**.** *A nontrivial price system* $(p^*,(p^*_i)_{i\in I})$*, a feasible data allocation* $q^* \in \mathbb{R}^{\Omega+\Omega\times I}_+$*, and a data use* $(x^*_i)_{i\in I}$ *form an equilibrium of economy* $\mathcal{E}_3$ *if:*

1. *The data allocation and the data use are consistent, namely* $q^*_i(\omega) = \sum_{a_i} x^*_i(a_i,\omega)$ *for all* ω *and* i*.*
2. *For each* i*,* x^*_i *maximizes (7) taking the price system as given.*
3. *For each* ω*,* $q^*_0(\omega)$ *and* $(x^*_i(a_i,\omega))_{i\in I,a_i\in A_i}$ *maximize (8) taking the price system as given.*
4. *Markets clear: for all* ω*,* $p^*(\omega)\Big(\bar{q}(\omega) - \sum_i q^*_i(\omega) - q^*_0(\omega)\Big) = 0$*.*

Given the (admittedly unrealistic) large number of markets that have been introduced in this economy, it is natural to wonder whether an equilibrium achieves or, in fact, surpasses second-best efficiency. Indeed, this economy reaches "first-best" efficiency: not only databases are allocated to maximize the sum of payoffs of buyers and platforms (as in $\mathcal{SB}$) but are also used in a socially optimal way (unlike in $\mathcal{SB}$). Formally, let us define the following social-planner problem:

$$\begin{aligned}
\mathcal{FB}: \quad & \max_{(x_i)_{i\in I}} \sum_{i,\omega,a_i}\Big(u_i(a_i,\omega) + g_i(a_i,\omega)\Big)x_i(a_i,\omega) + \sum_\omega r_\omega\Big(\bar{q}(\omega) - \sum_{i,a_i} x_i(a_i,\omega)\Big)\\
& \text{s.t. for all } i,\ k \in K_i, \text{ and } a^k, \hat{a}^k \in A^k,\\
& \quad \sum_{\omega,a^{-k}\in A^{-k}_i}\big(\pi_k(a^k,a^{-k},\omega) - \pi_k(\hat{a}^k,a^{-k},\omega)\big)x_i(a^k,a^{-k},\omega) \geq 0\\
& \text{and for all } \omega,\\
& \sum_{i,a_i} x_i(a_i,\omega) \leq \bar{q}(\omega).
\end{aligned} \tag{9}$$

Note that unlike in $\mathcal{SB}$, the social planner can choose $(x_i)_{i\in I}$ to maximize the sum of platforms' and buyers' payoffs. We refer to $\mathcal{FB}$ as the benchmark for first-best efficiency.

Proposition 3. *The economy $\mathcal{E}_3$ admits an equilibrium. Moreover, any equilibrium is first-best efficient, namely, its data use solves $\mathcal{FB}$. Conversely, for any data use $(x_i^{**})_{i\in I}$ that solves $\mathcal{FB}$, there is a price system $(p^{**}, (p_i^{**})_{i\in I})$ that supports the data use (and its implied data allocation) as an equilibrium for $\mathcal{E}_3$.*

In economy $\mathcal{E}_3$, platforms internalize all the external effects that their choice of x_i creates on the buyers. This leads to first-best efficiency. The other direction of the result exploits the special structure of the social planner's problem in $\mathcal{FB}$. Just like $\mathcal{P}$, $\mathcal{FB}$ is a "grand" information-design problem, where a social planner chooses the mechanism x_i on behalf of each platform to maximize the sum of the payoffs of buyers and platforms, subject to obedience and feasibility. This linear program admits a solution. Moreover, in the associated dual, the dual variables corresponding to the feasibility constraints (9) determine p^*. The additional step needed consists in retrieving $(p_i^*(a_i, \omega_j))_{\forall i,j,a_i}$. This can be done simply by setting $p_i^*(a_i, \omega_j) = -g_i(a_i, \omega_j)$. This makes buyers indifferent by fully compensating them for the use of their data records and makes the platforms fully internalize the externalities of how they use data. This simple specification of p_i^* works due to the linearity of both the platforms' and the representative buyers' objectives, which is a consequence of the underlying information-design structure of the platforms' problems.

From a methodological point of view, $\mathcal{P}$ and $\mathcal{FB}$ offer tractable ways to compute equilibrium prices and allocations of their respective economies. These are information design problems and many specific tools exist to analyze them. Their direct comparison is useful to better understand how changing the way markets are organized (e.g., who has property rights and which trades can be executed) affects the price of data. For example, one wonders under what conditions it is true that empowering buyers leads to higher prices for their data. This is a direction for future research.

Similarly, one may wonder how increasing r_ω, perhaps because buyers care intrinsically more about their privacy, will affect equilibrium price $p^*(\omega)$. It is easy to construct examples of non-trivial intermediation problems where increasing r_ω for some ω can reduce $p^*(\omega')$ for some $\omega' \neq \omega$. This can never happen for non-intermediation problems.

Finally, and in our opinion most importantly, one needs to recognize that economy $\mathcal{E}_3$ requires an unrealistic number of markets to be open and a high level of finesse. It is natural to ask whether there are practical ways to decentralize (perhaps partially) economy $\mathcal{E}_3$. Perhaps, these solutions may not achieve first-best efficiency but could still improve on the economy $\mathcal{E}_2$ discussed in Section 4. These questions remain open for future research.

6 Discussion

As explained in Section 2, this paper treats data records as *rival goods*. Two assumptions are key for this: the ownership of a record is exclusive and it is necessary in order to access the corresponding buyer. These assumptions can be partially relaxed, as discussed below.

The first assumption is exclusivity. It can be partially relaxed by allowing multiple platforms to own the same record as long as the corresponding buyer will only contemplate an offer chosen at random. Since buyers are a continuum, this version of the model is mathematically equivalent to the one we analyzed so far. To see this, suppose platform i and i' mutually own a given quantity of ω records. Suppose these buyers independently toss a fair coin to decide whether to consider a_i or $a_{i'}$. Then, this situation is equivalent to one where i and i' have exclusive ownership of half of these records. A richer model of mutual ownership would require that i and i' compete for the same buyers, even conditional having purchased their records. This would mean that the optimal mechanism for i may depend on the one chosen by i'. This setting may change some of the results in this paper and is a natural avenue for future research.

The second assumption is that ownership of a record is necessary for accessing the corresponding buyer. This paper deals with situations where accessing a buyer without knowing her identifier is prohibitively costly for a platform. For example, a platform may have a hard time reaching a buyer with an offer if it does not know her telephone number or email address. We can partially relax this assumption by giving buyers an outside option (which may depend on which platform own their data). An outside option captures the idea that a buyer may receive untargeted offers from platforms that do not own her records. In this case, the platform that owns her record will need to promise this buyer an expected payoff that is higher than her outside option in order to successfully intermediate her. In the baseline model, this outside option is normalized to zero, or it is already incorporated in the payoffs of the platform, sellers, and buyers through their implicit response to the sellers' offers. A richer model would require that these outside options are endogenous. This would imply that the optimal mechanism for a platform depends on the outside options offered by other platforms. As before, this setting may also change some of the results in this paper. We believe that this is another natural avenue for future research.

Our model makes other stylized assumptions, which we find less important. First, we assumed that data records are fully informative of the buyers' preferences. It is straightforward to allow for records that are only partially informative (see GLP, Section 4). Second, we assumed that the sellers on platform i are different from those on platform i' (i.e., $\{K_i\}$ partitions K). However, we can easily accommodate the case where the same seller participates on multiple platforms at the cost of additional notation. Importantly, note that because of exclusivity, sellers on different platforms do not compete with each other. Third, we assumed that sellers are exogenously allocated to platforms. This assumption can be weakened simply by adding an (exogenous) participation constraint for the seller (as done in GLP, Appendix B). Finally, we assumed platforms act as information designers. More generally, we can allow the platform to take contractible actions as a function of ω (e.g., monetary transfers).

Finally, this paper focused on the welfare of buyers and platforms. We ignored the sellers' profits for two reasons. First, by assumptions, sellers do not directly participate in the market for data.[8] Instead, they are "agents" of the platforms. It is well known that, when the principal and its agents have conflicting interests, the optimal mechanism will not in general maximize the agents' profits. This is an additional source of inefficiency in this model. However, this inefficiency is rather standard and, more importantly, has nothing to do with data markets. As such, it seems of second-order importance for this paper. Second, our model is flexible enough that it accommodates platforms whose objective is to maximize sellers' profits, i.e., $u_i(a_i, \omega) = \sum_{k \in K_i} \pi_k(a_i, \omega)$. For example, this is the case of a platform that earns transaction fees on every dollar that sellers make. Under this assumption, our welfare results would extend to incorporate the sellers' profits.

References

Acemoglu, D., A. Makhdoumi, A. Malekian, and A. Ozdaglar (2021): "Too Much Data: Prices and Inefficiencies in Data Markets," *Americal Economic Journal: Microeconomics*, Forthcoming.

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

[8] As explained at the end of Section 2, if they did, they should be modeled as platforms and, thus, their payoff would be accounted for by our measure of welfare.

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).

Bonnisseau, J.-M., E. del Mercato, and P. Siconolfi (2022): "Existence of an Equilibrium in Arrowian Markets for Consumption Externalities," *Documents de travail du Centre d'Economie de la Sorbonne*.

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.

Elul, R. (1995): "Welfare Effects of Financial Innovation in Incomplete Markets Economies with Several Consumption Goods," *Journal of Economic Theory*, 65, 43–78.

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*, Forthcoming.

Geanakoplos, J. and H. Polemarchakis (1986): "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations when the Asset Market is Incomplete," in *Essays in Honor of Kenneth Arrow*, ed. by W. Heller, R. Starr, and D. Starrett, Cambridge University Press, vol. 3, 65–95.

Ghosal, S. and H. M. Polemarchakis (1997): "Nash–Walras equilibria," *Research in Economics*, 51, 31–40.

Greenwald, B. and J. E. Stiglitz (1986): "Externalities in Economies with Imperfect Information and Incomplete Markets," *Quarterly Journal of Economics*, Vol. 101, No. 2, pp. 229–264.

Hart, O. D. (1975): "On the optimality of equilibrium when the market structure is incomplete," *Journal of Economic Theory*, 11, 418–443.

Ichihashi, S. (2021): "The Economics of Data Externalities," *Journal of Economic Theory*.

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.

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*.

A Proofs

A.1 Proposition 1

Step 1: We begin by showing that $\mathcal{TB}$ and $\mathcal{P}$ are equivalent. Consider any solution $(q_1^*, \ldots, q_I^*)$ of problem $\mathcal{TB}$. For each q_i, there is an associated optimal $x^*_{q_i}$ that solves $\mathcal{P}_i$.[9] It follows that

$$\begin{aligned}
\sum_i \sum_{\omega, a_i} u_i(a_i, \omega) x^*_{q_i^*}(a_i, \omega) &= \sum_i U_i(q_i^*) \\
&\geq \sum_i U_i(q_i) \qquad \text{for all } (q_1, \ldots, q_i) \text{ such that } \sum_\omega q_i(\omega) \leq \bar{q}(\omega)\, \forall \omega \\
&\geq \sum_i \sum_{\omega, a_i} u_i(a_i, \omega) x_i(a_i, \omega),
\end{aligned}$$

for all $(x_1, \ldots, x_I)$ such that each x_i satisfies the obedience constraints (1) and $\sum_{i,a_i} x_i(a_i, \omega) \leq \bar{q}(\omega)$ for all ω. It follows that the implied $(x^*_{q_1^*}, \ldots, x^*_{q_I^*})$ is a solution of $\mathcal{P}$.

Conversely, consider any solution $(x_1^*, \ldots, x_I^*)$ of $\mathcal{P}$. For every i and ω, define $q_i^*(\omega) = \sum_{a_i} x_i^*(a_i, \omega)$. The same inequalities as before imply that

$$\sum_i U_i(q_i^*) \geq \sum_i U_i(q_i) \qquad \text{for all } (q_1, \ldots, q_i) \text{ such that } \sum_\omega q_i(\omega) \leq \bar{q}(\omega)\, \forall \omega.$$

Therefore, the so defined $(q_1^*, \ldots, q_I^*)$ is a solution of $\mathcal{TB}$.

Step 2: Next, we argue that $\mathcal{P}$ and its dual $\mathcal{D}$ have a solution. The existence of an optimal solution to $\mathcal{P}$ follows from the same argument showing the existence of an optimal solution of each $\mathcal{P}_i$ (Footnote 9). This implies that the dual of $\mathcal{P}$ also has an optimal solution. In particular,

[9] The existence of $x^*_{q_i}$ follows from the boundeness of u_i and the non-emptiness of the constraint set defined by (1). To see the latter, given q_i, construct a Bayesian game between the sellers on platform i using the payoff functions $(\pi_i^k)_{k \in K_i}$ and the common prior $\mu_i(\omega) = \frac{q_i(\omega)}{\sum_{\omega'} q_i(\omega')}$ for all ω. This game has at least one BNE, which induces a conditional distribution $y_i(\omega) \in \Delta(A_i)$ for all ω. If we now define $x_i(a_i, \omega) = y_i(a_i|\omega) q_i(\omega)$ for all a_i, ω, the resulting x_i satisfies both constraints in $\mathcal{P}_i$.

let p^* be the optimal dual variables corresponding to the constraint (5). By strong duality, we have that

$$\sum_{\omega} p^* \bar{q}(\omega) = \sum_i \sum_{\omega,a_i} u_i(a_i,\omega) x_i^*(a_i,\omega).$$

Step 3: We now show that p^* and $(x_1^*, \ldots, x_I^*)$—hence, the induced $(q_1^*, \ldots, q_I^*)$—form an equilibrium of economy $\mathcal{E}_1$. To this end, note that each platform i's problem (4) can be equivalently expressed as

$$\max_{x_i} \sum_{\omega,a_i} u_i(a_i,\omega) x_i(a_i,\omega) - \sum_{\omega} p^*(\omega) \sum_{a_i} x_i(a_i,\omega)$$

subject to constraint (1). We need to show that x_i^* maximizes this problem. To see this, note that the value of the dual of the maximization problem faced by platform i is constant and equal to zero (because the right-hand-side of constraint (1) involves all zeros). Hence, by strong duality, x_i^* maximizes the platform's problem if

$$\sum_{\omega,a_i} u_i(a_i,\omega) x_i^*(a_i,\omega) - \sum_{\omega} p^*(\omega) \sum_{a_i} x_i^*(a_i,\omega) = 0.$$

In particular, by weak duality, we must have

$$\sum_{\omega,a_i} u_i(a_i,\omega) x_i^*(a_i,\omega) - \sum_{\omega} p^*(\omega) \sum_{a_i} x_i^*(a_i,\omega) \leq 0. \tag{A.1}$$

Therefore, summing over i, we must have

$$\sum_i \left\{ \sum_{\omega} p^*(\omega) \sum_{a_i} x_i^*(a_i,\omega) - \sum_{\omega,a_i} u_i(a_i,\omega) x_i^*(a_i,\omega) \right\} \leq \sum_{\omega} p^*(\omega) \bar{q}(\omega) - \sum_i \sum_{\omega,a_i} u_i(a_i,\omega) x_i^*(a_i,\omega)$$

because $(x_1^*, \ldots, x_I^*)$ satisfies constraint (5). However, the right-hand side equals zero. Since the left-hand side is a sum over non-negative terms, it follows that all inequalities in (A.1) must hold with equality, as desired.

Finally, market clearing holds because by complementary slackness we must have

$$p^*(\omega) \Big(\sum_{i,a_i} x_i^*(a_i,\omega) - \bar{q}(\omega) \Big) = 0, \quad \forall \omega.$$

A.2 Proof of Corollary 1

Step 1: We start by writing the dual $\mathcal{D}$ of $\mathcal{P}$. To this end, for every i, $k \in K_i$, and $a^k, \hat{a}^k \in A^k$, define a scalar $\lambda_i^k(\hat{a}^k|a^k) \geq 0$. Using this, for every i, ω, and $a_i \in A_i$, define

$$t_i(a_i,\omega;\lambda_i) = \sum_{k \in K_i} \sum_{\hat{a}^k \in A^k} \big(\pi_k(a^k, a^{-k}, \omega) - \pi_k(\hat{a}^k, a^{-k}, \omega)\big) \lambda_i^k(\hat{a}^k|a^k). \tag{A.2}$$

Letting $p \in \mathbb{R}_+^{\Omega}$, we then have

$$\begin{aligned}\mathcal{D} \quad : \quad & \min_{p,\lambda} \sum_{\omega} p(\omega)\bar{q}(\omega) \\ & \text{s.t. for all } \omega, i, a_i \\ & p(\omega) \geq u_i(a_i,\omega) + t_i(a_i,\omega;\lambda_i).\end{aligned}$$

Denote the (generically unique) solution of $\mathcal{D}$ by (p^*, λ^*).

Step 2: We now write the dual $\mathcal{D}_i$ of $\mathcal{P}_i$ given q_i^* induced by x_i^* that is part of the optimal solution of $\mathcal{P}$. Letting $v \in \mathbb{R}_+^{\Omega}$, we then have

$$\begin{aligned}\mathcal{D} \quad : \quad & \min_{v,\lambda_i} \sum_{\omega} v(\omega)q_i^*(\omega) \\ & \text{s.t. for all } \omega, a_i \\ & v(\omega) \geq u_i(a_i,\omega) + t_i(a_i,\omega;\lambda_i).\end{aligned}$$

Step 3: We now argue that (p^*, λ_i^*) must be a solution of each $\mathcal{D}_i$ and therefore p^* is equal to some solution $v^*_{q_i^*}$ of $\mathcal{D}_i$ for all i. This is because (p^*, λ_i^*) satisfies all the constraints of $\mathcal{D}_i$. Therefore, we must have

$$\sum_{\omega} p^*(\omega)q_i^*(\omega) - \sum_{\omega} v^*_{q_i^*}(\omega)q_i^*(\omega) \geq 0. \tag{A.3}$$

Moreover, by strong duality,

$$\sum_{\omega} v^*_{q_i^*}(\omega)q_i^*(\omega) = \sum_{\omega,a_i} u_i(a_i,\omega)x_i^*(a_i,\omega).$$

Therefore,

$$\begin{aligned}\sum_i \Big\{ \sum_{\omega} p^*(\omega)q_i^*(\omega) - \sum_{\omega} v^*_{q_i^*}(\omega)q_i^*(\omega) \Big\} \quad &= \quad \sum_i \Big\{ \sum_{\omega} p^*(\omega)q_i^*(\omega) - \sum_{\omega,a_i} u_i(a_i,\omega)x_i^*(a_i,\omega) \Big\} \\ &= \quad \sum_{\omega} p^*(\omega)\bar{q}(\omega) - \sum_i \sum_{\omega,a_i} u_i(a_i,\omega)x_i^*(a_i,\omega) = 0,\end{aligned}$$

where the second equality uses market clearing (or equivalently complementary slackness) and the last equality follows from strong duality for $\mathcal{P}$ and $\mathcal{D}$. It follows that each inequality in (A.3) must hold with equality, which means that p^* is part of an optimal solution of each $\mathcal{D}_i$.

A.3 Proof of Proposition 2

Recall that a data allocation is a vector $q \in \mathbb{R}_+^{\Omega+\Omega\times I}$. This includes: a database $q_0 \in \mathbb{R}_+^{\Omega}$, where each $q_0(\omega)$ denotes the quantity of ω records for the representative ω buyer; and, for each i a database $q_i \in \mathbb{R}^{\Omega}_{+}$, where $q_i(\omega)$ denotes the quantity of ω records for platform i. Recall that a data allocation q is feasible if, for all ω, $\sum_{\ell=0}^{I} q(\omega) = \bar{q}(\omega)$.

Let (p^*, q^*) be an equilibrium of $\mathcal{E}_2$. Fix $i \in I$ and $q_i \in \mathbb{R}^{\Omega}_{+}$. Condition 2 in Definition 2 implies that:

$$U_i(q_i^*) - U_i(q_i) \geq \sum_{\omega} p^*(\omega)\Big(q_i^*(\omega) - q_i(\omega)\Big).$$

Summing this expression over i, we obtain that, for every feasible data allocation q,

$$\sum_i U_i(q_i^*) - \sum_i U_i(q_i) \geq \sum_{\omega} p^*(\omega)\Big(\sum_i q_i^*(\omega) - \sum_i q_i(\omega)\Big). \tag{A.4}$$

Now fix ω and consider the problem of the representative buyer in Equation (6). Recall that a deviation for this buyer consists of choosing $q_\omega \in \mathbb{R}^{1+I}_{+}$ where $q^0_\omega(\omega)$ and $q^i_\omega(\omega)$ is the quantity of ω records that this representative buyer assigns to herself and the platform i, respectively. A deviation is feasible if $\sum_{\ell=0}^{I} q^{\ell}_\omega(\omega) = \bar{q}(\omega)$. Condition 1 from Definition 2 implies that, for every feasible deviation $q_\omega \in \mathbb{R}^{1+I}_{+}$ (i.e., one such that $\sum_{\ell=0}^{I} q^{\ell}_\omega(\omega) = \bar{q}(\omega)$), the following holds

$$\begin{aligned} & r_\omega q_0^*(\omega) + \sum_i q_i^*(\omega) G_{i\omega}(q_i^*) + p^*(\omega)\sum_i q_i^*(\omega) \geq \\ & r_\omega q^0_\omega(\omega) + \sum_i q^i_\omega(\omega) G_{i\omega}\Big(q^i_\omega(\omega), (q_i^*(\omega'))_{\omega' \neq \omega)}\Big) + p^*(\omega)\sum_i q^i_\omega(\omega) \end{aligned}$$

Summing over ω gives that, for every profile of feasible buyers' deviations $(q_\omega)_{\omega \in \Omega} \in \mathbb{R}^{\Omega + \Omega \times I}_{+}$:

$$\begin{aligned} & \sum_\omega U_\omega(q^*) - \sum_\omega \Big(r_\omega q^0_\omega(\omega) + \sum_i q^i_\omega(\omega) G_{i\omega}\Big(q^i_\omega(\omega), (q_i^*(\omega'))_{\omega' \neq \omega}\Big)\Big) \\ \geq \; & \sum_\omega p^*(\omega)\Big(\sum_i q^i_\omega(\omega) - \sum_i q_i^*(\omega)\Big) \end{aligned}$$

Note that a profile a feasible buyers' deviations is a feasible data allocation and vice versa. Fix any feasible data allocation $q \in \mathbb{R}^{\Omega+\Omega\times I}_{+}$ and denote by $q_\omega \in \mathbb{R}^{1+I}_{+}$ the part that concerns ω records. That is, $q = (q_\omega)_{\omega \in \Omega}$. Summing the previous expression with equation (A.4), we obtain that,

$$\begin{aligned} & \sum_i U_i(q_i^*) + \sum_\omega U_\omega(q^*) - \sum_i U_i(q_i) - \sum_\omega U_\omega(q_\omega, (q^*_{\omega'})_{\omega' \neq \omega}) \\ \geq \; & \sum_\omega p^*(\omega)\Big(\sum_i q_i^*(\omega) - \sum_i q_i^*(\omega) - \sum_i q_i(\omega) + \sum_i q^i_\omega(\omega)\Big) = 0. \end{aligned}$$

The last equality holds because of feasibility. Therefore, the equilibrium data allocation q^* satisfies

$$W(q^*) := \sum_i U_i(q_i^*) + \sum_\omega U_\omega(q^*) \geq \sum_i U_i(q_i) + \sum_\omega U_\omega(q_\omega, (q^*_{\omega'})_{\omega' \neq \omega)}) \tag{A.5}$$

for every feasible data allocation $q \in \mathbb{R}_+^{\Omega+\Omega\times I}$.

By $\mathcal{SB}$, a data allocation q^{**} is second-best efficient if it is feasible and, for all $q \in \mathbb{R}_+^{\Omega+\Omega\times I}$, satisfies

$$W(q^{**}) := \sum_i U_i(q_i^{**}) + \sum_\omega U_\omega(q^{**}) \geq \sum_i U_i(q_i) + \sum_\omega U_\omega(q).$$

We conclude that $W(q^{**}) \geq W(q^*)$ and therefore the equilibrium welfare is weakly worse than second-best efficient.

A.4 Proof of Corollary 2

Recall that in the no-intermediation case, $G_{i\omega}(q_\omega^i(\omega), (q_i(\omega'))_{\omega'\neq\omega}) = \bar{G}_{i\omega}$, for any i, ω, and data allocation $q \in \mathbb{R}_+^{\Omega+\Omega\times I}$. This implies that $U_\omega(q_\omega, (q_{\omega'}^*)_{\omega'\neq\omega}) = U_\omega(q_\omega)$. Under this condition, equation A.5 becomes

$$W(q^*) := \sum_i U_i(q_i^*) + \sum_\omega U_\omega(q^*) \geq \sum_i U_i(q_i) + \sum_\omega U_\omega(q_\omega).$$

for all feasible $q \in \mathbb{R}_+^{\Omega+\Omega\times I}$. Note that this is also the definition of q^{**}, the data allocation that solves $\mathcal{SB}$. Therefore, $W(q^*) = W(q^{**})$, i.e. the equilibrium data allocation is second-best efficient. $\square$

A.5 Proof of Proposition 3

Step 1: We first prove that an equilibrium allocation $((x_i^*)_{i\in I}, (q_\omega^*)_{\omega\in\Omega})$ must solve $\mathcal{FB}$. By platform i's maximization, we have that

$$\begin{aligned}\sum_{a_i,\omega} u_i(a_i,\omega)x_i^*(a_i,\omega) - \sum_{a_i,\omega} u_i(a_i,\omega)x_i(a_i,\omega) \quad \geq \quad & \sum_{a_i,\omega} \big(p^*(\omega) + p_i^*(a_i,\omega)\big)x_i^*(a_i,\omega) \\ & - \sum_{a_i,\omega} \big(p^*(\omega) + p_i^*(a_i,\omega)\big)x_i(a_i,\omega)\end{aligned}$$

for all x_i that satisfy (1). Therefore, the same holds summing over i on both sides:

$$\begin{aligned}\sum_{i,a_i,\omega} u_i(a_i,\omega)x_i^*(a_i,\omega) - \sum_{i,a_i,\omega} u_i(a_i,\omega)x_i(a_i,\omega) \quad \geq \quad & \sum_{i,a_i,\omega} \big(p^*(\omega) + p_i^*(a_i,\omega)\big)x_i^*(a_i,\omega) \\ & - \sum_{i,a_i,\omega} \big(p^*(\omega) + p_i^*(a_i,\omega)\big)x_i(a_i,\omega)\end{aligned}$$

for all x_i that satisfy (1) for all i and (9).

Similarly, by the maximization of the representative ω buyer, we get

$$
\begin{aligned}
r_\omega q^*_\omega(\omega) + \sum_{i,a_i} g_i(a_i,\omega)x^*_i(a_i,\omega) - r_\omega q_\omega(\omega) - \sum_{i,a_i} g_i(a_i,\omega)x_i(a_i,\omega) &\geq \\
p^*(\omega)q^*_\omega(\omega) - \sum_{i,a_i} p^*_i(a_i,\omega)x^*_i(a_i,\omega) - p^*(\omega)q_\omega(\omega) + \sum_{i,a_i} p^*_i(a_i,\omega)x_i(a_i,\omega)
\end{aligned}
$$

for all $q_\omega(\omega)$ and x_i. Therefore, the same holds summing over ω on both sides:

$$
\begin{aligned}
\sum_\omega r_\omega q^*_\omega(\omega) + \sum_{i,a_i,\omega} g_i(a_i,\omega)x^*_i(a_i,\omega) - \sum_\omega r_\omega q_\omega(\omega) - \sum_{i,a_i,\omega} g_i(a_i,\omega)x_i(a_i,\omega) &\geq \\
\sum_\omega p^*(\omega)q^*_\omega(\omega) - \sum_{i,a_i,\omega} p^*_i(a_i,\omega)x^*_i(a_i,\omega) - \sum_\omega p^*(\omega)q_\omega(\omega) + \sum_{i,a_i,\omega} p^*_i(a_i,\omega)x_i(a_i,\omega).
\end{aligned}
$$

If we now combine the inequalities for the platforms and the representative buyers, we obtain that for all $(q_\omega)_{\omega\in\Omega}$ and x_i that satisfy (1) for all i and (9)

$$
\begin{aligned}
&\sum_{i,a_i,\omega} \big(u_i(a_i,\omega) + g_i(a_i,\omega)\big)x^*_i(a_i,\omega) + \sum_\omega r_\omega q^*_\omega(\omega) \\
&- \sum_{i,a_i,\omega} \big(u_i(a_i,\omega) + g_i(a_i,\omega)\big)x_i(a_i,\omega) - \sum_\omega r_\omega q_\omega(\omega)
\end{aligned}
$$

is greater than or equal to

$$
\begin{aligned}
&\sum_\omega p^*(\omega)\Big(\sum_{i,a_i} x^*_i(a_i,\omega) + q^*_\omega(\omega)\Big) - \sum_\omega p^*(\omega)\Big(\sum_{i,a_i} x_i(a_i,\omega) + q_\omega(\omega)\Big) \\
= \ &\sum_\omega p^*(\omega)\bar{q}(\omega) - \sum_\omega p^*(\omega)\Big(\sum_{i,a_i} x_i(a_i,\omega) + q_\omega(\omega)\Big) \geq 0,
\end{aligned}
$$

where the equality follows from market clearing and the inequality follows from feasibility.

Step 2: We now prove that for any allocation $((x^{**}_i)_{i\in I}, (q^{**}_\omega)_{\omega\in\Omega}))$ that solves $\mathcal{F}B$, there is a price system $(p^{**}, (p^{**}_i)_{i\in I})$ that together with the allocation constitutes an equilibrium of $\mathcal{E}_3$. First of all, $\mathcal{F}B$ has an optimal solution for the same reason that $\mathcal{P}$ has one. Second, we can define $p^{**}_i(a_i,\omega) = -g_i(a_i,\omega)$ for all i, a_i, ω, so that each representative ω buyer is indifferent across all possible $\hat{x}_i(\cdot,\omega)$ and we can therefore assume to choose $\hat{x}_i(\cdot,\omega) = x^{**}_i(\cdot,\omega)$ for all i.

Next, we let p^{**} be the optimal dual variables associated with constraint (9) for every ω that solve the dual of $\mathcal{FB}$. In particular, this dual involves choosing $p \in \mathbb{R}^\Omega_+$ and $\lambda^k(\hat{a}^k|a^k) \geq 0$ for every i, $k \in K_i$, and $a^k, \hat{a}^k \in A^k$ to minimize

$$
\sum_\omega p(\omega)\bar{q}(\omega)
$$

subject to for all ω

$$p^{**}(\omega) \geq \max\Big\{r_\omega, \max_{i,a_i}\{u_i(a_i,\omega) + g_i(a_i,\omega) + t_i(a_i,\omega;\lambda_i)\}\Big\},$$

where t_i was defined in (A.2). Therefore, if $p^{**}(\omega) > r_\omega$, then $q^{**}_\omega(\omega)$ must be zero by complementary slackness. But in this case, given $p^{**}(\omega)$ it is also optimal for the representative ω buyer to sell all her data and choose $\hat{q}_\omega(\omega) = 0$. If instead $p^{**}(\omega) = r_\omega$, then $q^{**}_\omega(\omega)$ can be positive and we can let $\hat{q}_\omega(\omega) = q^{**}_\omega(\omega)$ because the representative ω buyer is indifferent between keeping and selling any record. Combining these observations, we have that $q^{**}_\omega(\omega)$ and $(x^{**}_i)_{i\in I}$ solve each representative buyer's maximization problem given the defined price system.

It remains to check the platforms' maximization. Using the suggested price system, platform i's problem becomes to choose x_i to maximize

$$\begin{aligned}&\sum_{a_i,\omega}\big(u_i(a_i,\omega) + g_i(a_i,\omega) - p^{**}(\omega)\big)x_i(a_i,\omega)\\ &\qquad\text{s.t. for all } k\in K_i, \text{ and } a^k, \hat{a}^k \in A^k\\ &\sum_{\omega\in\Omega, a^{-k}\in A_i^{-k}}\big(\pi_k(a^k,a^{-k},\omega) - \pi_k(\hat{a}^k,a^{-k},\omega)\big)x_i(a^k,a^{-k},\omega) \geq 0\end{aligned}$$

This is a linear program whose dual must have a constant value equal to zero. Therefore, we only need to show that x^{**}_i achieves a value of zero. To this end, note that by weak duality, for every i,

$$\sum_\omega p^{**}(\omega)\sum_{a_i} x^{**}_i(a_i,\omega) - \sum_{a_i,\omega}\big(u_i(a_i,\omega) + g_i(a_i,\omega)\big)x^{**}_i(a_i,\omega) \geq 0,$$

which implies that

$$\sum_i\Big\{\sum_\omega p^{**}(\omega)\sum_{a_i} x^{**}_i(a_i,\omega) - \sum_{a_i,\omega}\big(u_i(a_i,\omega) + g_i(a_i,\omega)\big)x^{**}_i(a_i,\omega)\Big\} \geq 0. \tag{A.6}$$

Now, by complementary slackness, if $q^{**}_\omega(\omega) > 0$, then $p^{**}(\omega) = r_\omega$. Also, by strong duality, we have

$$\begin{aligned}\sum_{i,a_i,\omega}\big(u_i(a_i,\omega) + g_i(a_i,\omega)\big)x^{**}_i(a_i,\omega) + \sum_\omega r_\omega q^{**}_\omega(\omega) &= \sum_\omega p^{**}(\omega)\bar{q}(\omega)\\ &= \sum_\omega p^{**}(\omega)\Big(q^{**}_\omega(\omega) + \sum_{i,a_i} x^{**}_i(a_i,\omega)\Big),\end{aligned}$$

where the last equality uses complementary slackness for constraint (9). Therefore, we conclude that the left-hand side of (A.6) equals zero. Since this side is the sum of non-negative terms, each term must be zero, which implies that x^{**}_i maximizes platform i's objective given the suggested price system for all i. This completes the proof.