Competitive Markets for Personal Data

Simone Galperti   Tianhao Liu   Jacopo Perego

August 2025

Motivation

Consumers supply a crucial input for modern economy: their **personal data**

Yet, they often have **limited control** over how and by whom their data is used:

- This may lead to inefficiencies and inequality

New legislation gives consumers more control over their data (GDPR, CCPA, ...)

- Lays foundations upon which **data markets** could emerge

What properties would these markets have, and how should they be designed to promote desirable outcomes?

This Paper

We contribute to a recent literature on the welfare properties of **data markets**

Choi et al 19, Ichihashi 21, Bergemann, Bonatti, Gan 22, Acemoglu et al. 22, Chen 24, [...]

This Paper

We contribute to a recent literature on the welfare properties of **data markets**

Choi et al 19, Ichihashi 21, Bergemann, Bonatti, Gan 22, Acemoglu et al. 22, Chen 24, [...]

1. A model of a **competitive data market** in which:
   - Consumers own their data and can sell it to intermediaries, which use it to inform merchants, who sell products to consumers

This Paper

We contribute to a recent literature on the welfare properties of **data markets**

Choi et al 19, Ichihashi 21, Bergemann, Bonatti, Gan 22, Acemoglu et al. 22, Chen 24, [...]

1. A model of a **competitive data market** in which:
   – Consumers own their data and can sell it to intermediaries, which use it to inform merchants, who sell products to consumers

2. We identify a **novel externality** leading to a market failure
   – Its presence depends on how the intermediary uses consumer data
   – *Equilibirum is efficient* ***iff*** *intermediary chooses "partitional information"*

(e.g., full disclosure)

This Paper

We contribute to a recent literature on the welfare properties of **data markets**

Choi et al 19, Ichihashi 21, Bergemann, Bonatti, Gan 22, Acemoglu et al. 22, Chen 24, [...]

1. A model of a **competitive data market** in which:

- Consumers own their data and can sell it to intermediaries, which use it to inform merchants, who sell products to consumers

2. We identify a **novel externality** leading to a market failure

- Its presence depends on how the intermediary uses consumer data
- *Equilibirum is efficient **iff** intermediary chooses "partitional information"*

(e.g., full disclosure)

3. Propose three **remedies** to market failure: Data taxes, data union, Lindahl pricing

Related Work

Exploit progress in info-design to microfound components of data economy:

- How does an intermediary use the data? **Information design** (Bergmann-Morris '19, Kamenica '19)
- What's the value of data for an intermediary? (GLP '23)

Related Work

Exploit progress in info-design to microfound components of data economy:

- How does an intermediary use the data? **Information design** (Bergmann-Morris '19, Kamenica '19)
- What's the value of data for an intermediary? (GLP '23)

Limited understanding of **data markets**:

- "Correlation" externality Choi et al ('19), BBG ('22), Acemoglu et al. ('22)
- Our inefficiency not due to exogenous correlation, but to platform's role as info intermediary (indeed, no intermediation $\Rightarrow$ no externality)

Related Work

Exploit progress in info-design to microfound components of data economy:

- How does an intermediary use the data? **Information design** (Bergmann-Morris '19, Kamenica '19)
- What's the value of data for an intermediary? (GLP '23)

Limited understanding of **data markets**:

- "Correlation" externality Choi et al ('19), BBG ('22), Acemoglu et al. ('22)
- Our inefficiency not due to exogenous correlation, but to platform's role as info intermediary (indeed, no intermediation $\Rightarrow$ no externality)

More broadly, we contribute to the growing literature on the economics of platforms, data, & privacy Bergemann and Ottaviani '21, Baley and Veldkamp '24

model

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω in the population

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω in the population

Each consumer has an outside option $r \sim F_\omega$ (fully supported on $\mathbb{R}$)

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω in the population

Each consumer has an outside option $r \sim F_\omega$ (fully supported on $\mathbb{R}$)

Each consumer owns a **data record** that fully reveals her corresponding ω

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω in the population

Each consumer has an outside option $r \sim F_\omega$ (fully supported on $\mathbb{R}$)

Each consumer owns a **data record** that fully reveals her corresponding ω

Two periods: 1. Data are traded 2. Data are used

Period 1: Competitive Data Markets

Period 1: Competitive Data Markets

The consumers and the platform trade data records at prices $p = (p(\omega))_{\omega \in \Omega}$, which they take as given

Period 1: Competitive Data Markets

The consumers and the platform trade data records at prices $p = (p(\omega))_{\omega \in \Omega}$, which they take as given

The demand side:

- Platform demands **database** $q = (q(\omega))_{\omega \in \Omega}$, for which it pays $\sum_\omega q(\omega) p(\omega)$

Period 1: Competitive Data Markets

The consumers and the platform trade data records at prices $p = (p(\omega))_{\omega \in \Omega}$, which they take as given

The demand side:

– Platform demands **database** $q = (q(\omega))_{\omega \in \Omega}$, for which it pays $\sum_\omega q(\omega)p(\omega)$

The supply side:

– If a type-ω consumer sells her record to the platform, she is paid $p(\omega)$ and is later intermediated with merchant

– If she doesn't her record, she enjoys her outside option $r \sim F_\omega$

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It commits to signal for merchant about each ω in its database
- Given signal, the merchant chooses action $a \in A$ (e.g., price/variety/quality)
- Together, a and ω determine period-2 payoffs:

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It commits to signal for merchant about each ω in its database
- Given signal, the merchant chooses action $a \in A$ (e.g., price/variety/quality)
- Together, a and ω determine period-2 payoffs:

Consumer's:	$u(a, \omega)$	e.g., trading surplus
Merchant's:	$\pi(a, \omega)$	e.g., profits from trade
Platform's:	$v(a, \omega)$	e.g., revenues

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It commits to signal for merchant about each ω in its database
- Given signal, the merchant chooses action $a \in A$ (e.g., price/variety/quality)
- Together, a and ω determine period-2 payoffs:

Consumer's:	$u(a, \omega)$	e.g., trading surplus
Merchant's:	$\pi(a, \omega)$	e.g., profits from trade
Platform's:	$v(a, \omega)$	e.g., revenues

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It commits to signal for merchant about each ω in its database
- Given signal, the merchant chooses action $a \in A$ (e.g., price/variety/quality)

Remark: Info design problem equivalent to a linear program: (BM '16)

$$V(q) = \max_{x:\Omega \to \Delta(A)} \sum_{\omega, a} v(a,\omega) x(a|\omega) q(\omega)$$

$$\text{s.t. } \forall a, a': \sum_{\omega} \Big(\pi(a,\omega) - \pi(a',\omega)\Big) x(a|\omega) q(\omega) \geq 0$$

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It commits to signal for merchant about each ω in its database
- Given signal, the merchant chooses action $a \in A$ (e.g., price/variety/quality)

Remark: Info design problem equivalent to a linear program: (BM '16)

$$V(q) = \max_{x:\Omega \to \Delta(A)} \sum_{\omega, a} v(a,\omega) x(a|\omega) q(\omega)$$

$$\text{s.t. } \forall a, a' \colon \sum_{\omega} \Big(\pi(a,\omega) - \pi(a',\omega)\Big) x(a|\omega) q(\omega) \geq 0$$

(ID problem, but with endogenous prior)

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It commits to signal for merchant about each ω in its database
- Given signal, the merchant chooses action $a \in A$ (e.g., price/variety/quality)

Remark: Info design problem equivalent to a linear program: (BM '16)

$$V(q) = \max_{x:\Omega \to \Delta(A)} \sum_{\omega, a} v(a,\omega) x(a|\omega) q(\omega)$$

$$\text{s.t. } \forall a, a': \sum_{\omega} \Big(\pi(a,\omega) - \pi(a',\omega)\Big) x(a|\omega) q(\omega) \geq 0$$

(ID problem, but with endogenous prior)

Equilibrium

Definition

A profile (p^*, q^*, x^*)

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

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

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

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

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

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

(b). Given q^*, x^* solves platform's info-design problem

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

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

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

(b). Given q^*, x^* solves platform's info-design problem

(c). Given p^* and x^*, a type-ω consumer with outside option r sells record if

$$r \leq p^*(\omega) + \sum_a x^*(a|\omega) u(a, \omega) \triangleq r^*_\omega$$

(d). Data markets clear, i.e. $\quad q^*(\omega) = \bar{q}(\omega) F_\omega(r^*_\omega) \qquad \forall \omega$

Discussion

All results extend to large class of **information-intermediation problems**:

- Multiple agents (e.g., competing merchants)
- Arbitrary downstream (finite) games (e.g., a second-price auctions)
- More than information: Platform can take enforceable action (e.g., set reservation price, charge merchants for information, etc.)

Leading applications: Online marketplaces and advertisement auctions

Discussion

All results extend to large class of **information-intermediation problems**:

- Multiple agents (e.g., competing merchants)
- Arbitrary downstream (finite) games (e.g., a second-price auctions)
- More than information: Platform can take enforceable action (e.g., set reservation price, charge merchants for information, etc.)

Leading applications: Online marketplaces and advertisement auctions

Substantive assumptions we made:

- A competitive data market
- Platform is a "gate keeper" alt see BB '23
- A data record bundles "access" and information alt see ALV '22

Discussion

All results extend to large class of **information-intermediation problems**:

- Multiple agents (e.g., competing merchants)
- Arbitrary downstream (finite) games (e.g., a second-price auctions)
- More than information: Platform can take enforceable action (e.g., set reservation price, charge merchants for information, etc.)

Leading applications: Online marketplaces and advertisement auctions

Substantive assumptions we made:

- A competitive data market
- Platform is a "gate keeper" alt see BB '23
- A data record bundles "access" and information alt see ALV '22

main results

Goal

Our Main Question: Is the **competitive** data market efficient?

We study whether equilibria maximize the welfare of the participants in the data market, i.e., the platform and the consumers

Goal

Our Main Question: Is the **competitive** data market efficient?

We study whether equilibria maximize the welfare of the participants in the data market, i.e., the platform and the consumers

Denote their welfare by $\mathcal{W}(q, x)$

Goal

Our Main Question: Is the **competitive** data market efficient?

We study whether equilibria maximize the welfare of the participants in the data market, i.e., the platform and the consumers

Denote their welfare by $\mathcal{W}(q, x)$

Definition discussion

An allocation (q, x) is **efficient** if it solves

$$\max_{q,x} \mathcal{W}(q, x)$$

$$\text{s.t.}\, q \leq \bar{q} \quad \text{and} \quad x \text{ solves platform's info-design problem given } q$$

Data Externality

Main Insight: This economy may feature an **externality**

Data Externality

Main Insight: This economy may feature an **externality**

- When a consumer sells her record,
  - $\Rightarrow$ She marginally affects $q(\omega)$...
  - $\Rightarrow$ ...which may change how the platform uses the data, namely x_q,...
  - $\Rightarrow$ ...which in turn affects other consumers' payoffs

This externality makes equilibria inefficient

Externality is **endogenous**: Its presence depends on how platform use data

Data Externality

Main Insight: This economy may feature an **externality**

- When a consumer sells her record,
  - $\Rightarrow$ She marginally affects $q(\omega)$...
  - $\Rightarrow$ ...which may change how the platform uses the data, namely x_q,...
  - $\Rightarrow$ ...which in turn affects other consumers' payoffs

This externality makes equilibria inefficient

Externality is **endogenous**: Its presence depends on how platform use data

(Aside:) Not related with "externality" from GLP '23, which are "priced in" by the competitive market

Data Externality

Goal: Characterize which equilibria feature externality and which don't

Data Externality

Goal: Characterize which equilibria feature externality and which don't

An externality exists if the platform's optimal mechanism x^* locally depends on q

Data Externality

Goal: Characterize which equilibria feature externality and which don't

An externality exists if the platform's optimal mechanism x^* locally depends on q

Proposition

An equilibrium (p^*, q^*, x^*) is efficient only if

$$\sum_{a,\omega'} q^*(\omega') u(a, \omega') \frac{\partial x^*(a|\omega')}{\partial q^*(\omega)} = 0 \qquad \forall \omega$$

If welfare function $\mathcal{W}(q, x_q)$ is concave in q, such a condition is also sufficient

Data Externality

Goal: Characterize which equilibria feature externality and which don't

An externality exists if the platform's optimal mechanism x^* locally depends on q

Proposition

An equilibrium (p^*, q^*, x^*) is efficient only if

$$\sum_{a,\omega'} q^*(\omega') u(a, \omega') \frac{\partial x^*(a|\omega')}{\partial q^*(\omega)} = 0 \qquad \forall \omega$$

If welfare function $\mathcal{W}(q, x_q)$ is concave in q, such a condition is also sufficient

So what? E.g., what ways of using consumer data leads to (in)efficiency?

A Typology of Recommendation Mechanisms

A Typology of Recommendation Mechanisms

Definition

A recommendation mechanism is **partitional** if $x : \Omega \to A$.

A Typology of Recommendation Mechanisms

Definition

A recommendation mechanism is **partitional** if $x : \Omega \to A$.

Such a mechanism recommends a single action for each cell in a partition of Ω

Notable examples:

Full Disclosure. For each ω, it recommends only an action the merchant is willing to choose under complete information

No Disclosure. For each ω, it recommends the same action

A Typology of Recommendation Mechanisms

Definition

A recommendation mechanism is **partitional** if $x : \Omega \to A$.

Such a mechanism recommends a single action for each cell in a partition of Ω

Notable examples:

- **Full Disclosure**. For each ω, it recommends only an action the merchant is willing to choose under complete information
- **No Disclosure**. For each ω, it recommends the same action

A non-partitional mechanism requires the platform to **randomize**

Result 1: Full Disclosure Leads to Efficiency

Result 1: Full Disclosure Leads to Efficiency

Proposition

If full disclosure is optimal at an interior $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Result 1: Full Disclosure Leads to Efficiency

Proposition

If full disclosure is optimal at an interior $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Proof Sketch:

- If full disclosure is optimal at interior q, it is optimal at all q

Result 1: Full Disclosure Leads to Efficiency

Proposition

If full disclosure is optimal at an interior $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Proof Sketch:

- If full disclosure is optimal at interior q, it is optimal at all q
- Thus, x_q constant in $q \Rightarrow$ (1) No data externality; (2) welfare function is concave $\Rightarrow$ Efficiency

Result 1: Full Disclosure Leads to Efficiency

Proposition

If full disclosure is optimal at an interior $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Proof Sketch:

- If full disclosure is optimal at interior q, it is optimal at all q
- Thus, x_q constant in $q \Rightarrow$ (1) No data externality; (2) welfare function is concave $\Rightarrow$ Efficiency

Corollary

Fix an equilibrium (p^*, q^*, x^*). If x^* is a full-disclosure mechanism, the equilibrium is efficient.

Result 1: Full Disclosure Leads to Efficiency

Proposition

If full disclosure is optimal at an interior $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Proof Sketch:

- If full disclosure is optimal at interior q, it is optimal at all q
- Thus, x_q constant in $q \Rightarrow$ (1) No data externality; (2) welfare function is concave $\Rightarrow$ Efficiency

Corollary

Fix an equilibrium (p^*, q^*, x^*). If x^* is a full-disclosure mechanism, the equilibrium is efficient.

Thus, inefficiency requires platform to **withhold information** from merchant

Result 2: Random Mechanisms Lead to Inefficiency

Proposition

Let (p^*, q^*, x^*) be an equilibrium. If x^* is **non-partitional**, the equilibrium is inefficient (generically in u)

Result 2: Random Mechanisms Lead to Inefficiency

Proposition

Let (p^*, q^*, x^*) be an equilibrium. If x^* is **non-partitional**, the equilibrium is inefficient (generically in u)

Intuition. x^* must randomize at some ω and randomization depends on q (at least one IC constraint binds). $\sum_{a,\omega'} q^*(\omega')u(a,\omega')\frac{\partial x_{q^*}(a|\omega')}{\partial q^*(\omega)} \neq 0 \Rightarrow$ Externality

Result 2: Random Mechanisms Lead to Inefficiency

Proposition

Let (p^*, q^*, x^*) be an equilibrium. If x^* is **non-partitional**, the equilibrium is inefficient (generically in u)

Intuition. x^* must randomize at some ω and randomization depends on q (at least one IC constraint binds). $\sum_{a,\omega'} q^*(\omega')u(a,\omega')\frac{\partial x_{q^*}(a|\omega')}{\partial q^*(\omega)} \neq 0 \Rightarrow$ Externality

The converse holds under additional conditions (satisfied in apps such as BBM'15)

Proposition

Let [$\mathcal{W}(q, x_q)$ be concave in q] and (p^*, q^*, x^*) be an equilibrium.
If x^* is **partitional** [and $q^* \in \text{int}(R_i)$ for some i], the equilibrium is efficient

an application

Mediated Price Discrimination

So far, minimal assumptions on the intermediation problem

Let's specialize setting to a canonical application: Price discrimination *à la* BBM

Mediated Price Discrimination

So far, minimal assumptions on the intermediation problem

Let's specialize setting to a canonical application: Price discrimination *à la* BBM

- Let $\omega \in \mathbb{R}_{++}$ be the consumer's WTP for merchant's product
- Let a denote price the merchant sets for his product
- Players payoffs are

$$\begin{aligned}
&\text{Consumer's:} \quad && u(a,\omega) = \max\{\omega - a, 0\} \\
&\text{Merchant's:} \quad && \pi(a,\omega) = a \, \mathbb{1}(\omega \geq a) \\
&\text{Platform's:} \quad && v(a,\omega) = \gamma_u \, u(a,\omega) + \gamma_\pi \, \pi(a,\omega)
\end{aligned}$$

Mediated Price Discrimination

$$v(a, \omega) = \gamma_u \, u(a, \omega) + \gamma_\pi \, \pi(a, \omega)$$

Mediated Price Discrimination

$$v(a,\omega) = \gamma_u \; u(a,\omega) + \gamma_\pi \; \pi(a,\omega)$$

Proposition

Fix any equilibrium (q^*, p^*, x^*).

- If $\gamma_u < \gamma_\pi$, x^* is **full disclosure**, and, thus, the equilibrium is efficient.
- If $\gamma_u > \gamma_\pi$, the equilibrium is efficient if and only if x^* is **no disclosure**.

Mediated Price Discrimination

$$v(a,\omega) = \gamma_u \, u(a,\omega) + \gamma_\pi \, \pi(a,\omega)$$

Proposition

Fix any equilibrium (q^*, p^*, x^*).

- If $\gamma_u < \gamma_\pi$, x^* is **full disclosure**, and, thus, the equilibrium is efficient.
- If $\gamma_u > \gamma_\pi$, the equilibrium is efficient if and only if x^* is **no disclosure**.

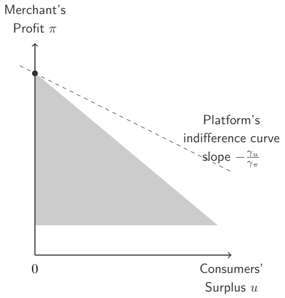
That is, any "non-trivial" use of information by the platform will lead to inefficiencies

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$

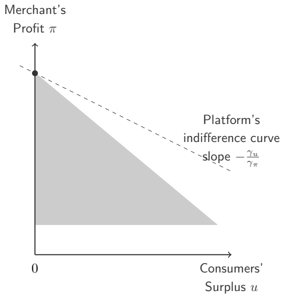
Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



Intuition for Efficiency

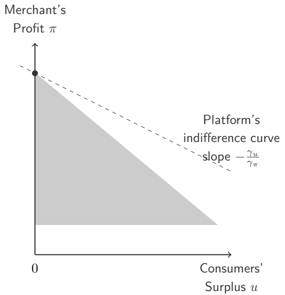
If $\gamma_u < \gamma_\pi$



- At all q, **full disclosure** is optimal
- Merchant extracts surplus from all consumers

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



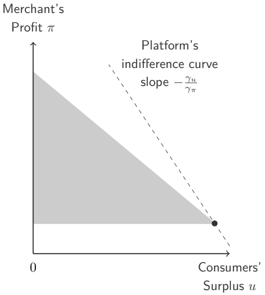
- At all q, **full disclosure** is optimal
- Merchant extracts surplus from all consumers
- Therefore, $x^*(a, \omega)$ does not depend on q
- Therefore, no externality! All equilibria are constrained efficient

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$

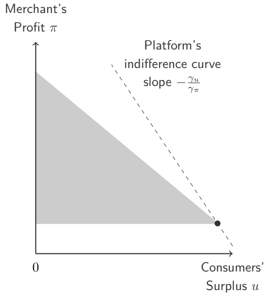
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



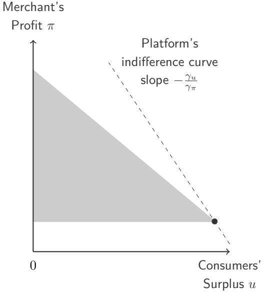
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



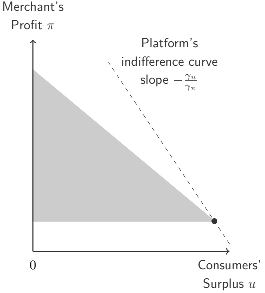
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



Intuition for Inefficiency

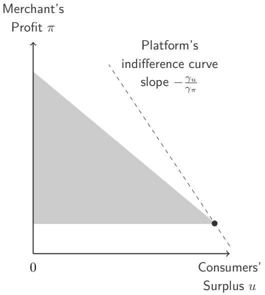
If $\gamma_u > \gamma_\pi$



- Platform **withholds information** from merchant

Intuition for Inefficiency

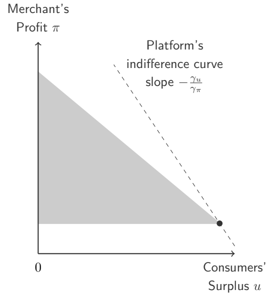
If $\gamma_u > \gamma_\pi$



- Platform **withholds information** from merchant
- Pooling types required to keep merchant indifferent at some as

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



- Platform **withholds information** from merchant
- Pooling types required to keep merchant indifferent at some as
- Thus, x_q depends on q
- Thus, $\sum_{\omega'} q(\omega') \frac{\partial U(\omega', x^*)}{\partial q^*(\omega)}$ can be non-zero: data externality

Bigger Picture

Application highlights when there is **no conflict of interest** btw platform and merchant $\Rightarrow$ full disclosure is optimal $\Rightarrow$ data market is efficient

Special case: if platform is the merchant $\Rightarrow$ no intermediation

Bigger Picture

Application highlights when there is **no conflict of interest** btw platform and merchant $\Rightarrow$ full disclosure is optimal $\Rightarrow$ data market is efficient

Special case: if platform is the merchant $\quad\Rightarrow\quad$ no intermediation

Thus the source of the inefficiency is the role platforms play as **information intermediaries**

- Platforms typically balance conflicting interests, which they rarely resolve with full disclosure — otw, no info-design literature! :)
- Instead, they often garble the data they have collected

Bigger Picture

Application highlights when there is **no conflict of interest** btw platform and merchant $\Rightarrow$ full disclosure is optimal $\Rightarrow$ data market is efficient

Special case: if platform is the merchant $\Rightarrow$ no intermediation

Thus the source of the inefficiency is the role platforms play as **information intermediaries**

- Platforms typically balance conflicting interests, which they rarely resolve with full disclosure — otw, no info-design literature! :)
- Instead, they often garble the data they have collected

This paper shows why garbling can lead to a failure of the first-welfare theorem in a competitive data market

Remedies

How to fix this market failure?

The paper also explores three alternative market designs to fix this market failure:

1. Introducing a **data union**
2. Implementing **data taxes**
3. Making data markets more **complete**

data union

Data Unions

Recent policy proposals for the data economy (Posner-Weyl 18; Bergemann et al 23)

A data union would represent consumers by managing data on their behalf

We offer some theoretical support to these policy proposals

Data Union

How does a data union work?

- Consumers can participate in the union
- If they do, they relinquish their data to the union
- Union sells some of this data to the platform

  Consumers retain reservation utility unless record is sold to platform

- With the proceeds of sale, union compensates all participating consumers (to incentivize their participation)
- Union maximizes welfare of participating consumers

Data Union

Formally, the data union problem is:

$$
\begin{aligned}
\max_{(p,q,x)} \quad & \sum_{\omega} p(\omega)\bar{q}(\omega) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega}(\bar{q}(\omega) - q(\omega))\bar{r} \\
\text{such that} \quad & q \leq \bar{q}, \\
\text{and} \quad & \textstyle\sum_{\omega} p(\omega)\bar{q}(\omega) = V(q), \\
\text{and} \quad & x \text{ solves } \mathcal{P}_q, \\
\text{and} \quad & p(\omega) + \tfrac{q(\omega)}{\bar{q}(\omega)} \textstyle\sum_a u(a,\omega)x(a|\omega) + \left(1 - \tfrac{q(\omega)}{\bar{q}(\omega)}\right)\bar{r} \geq \bar{r}.
\end{aligned}
$$

Data Union

Proposition

Equilibria of the data-union economy are constrained efficient and maximize consumers' welfare (and vice versa), regardless of the platform's objective

Some intuition:

Data union coordinates consumers by deciding which records to sell and how to compensate them

By doing so, data union acts as a substitute for the competitive market and avoids market failure

data taxes

Data Taxes

Enrich competitive economy by introducing a simple **data tax**:

- When selling her record, consumer pays tax $\tau(\omega) \in \mathbb{R}$ to the govt

When properly designed, data taxes force consumers to internalize effects that selling their records create on economy

Data Taxes

Enrich competitive economy by introducing a simple **data tax**:

- When selling her record, consumer pays tax $\tau(\omega) \in \mathbb{R}$ to the govt

When properly designed, data taxes force consumers to internalize effects that selling their records create on economy

Proposition

Let (q°, x°) be a constrained-efficient allocation. There exists a profile of taxes τ^*, of prices p^*, and of consumer choices ζ^*, such that $(p^*, \zeta^*, q^\circ, x^\circ)$ is an equilibrium of the economy with taxation τ^* and the government does not run a deficit.

Sketch of the Proof

Let allocation (q°, x°) be constrained efficient

Let p^* be a supergradient of $V(q^\circ)$

Define $\boxed{\tau^*(\omega) \triangleq p^*(\omega) + \sum_a x^\circ(a|\omega)u(a,\omega) - \bar{r}}$

Notice that $U^*(\omega) - \tau^*(\omega) \equiv \bar{r}$

Therefore, all consumers indifferent $\rightsquigarrow$ choose ζ^* to implement q° $\square$

more-complete markets

Lindahl Economy

We let price of data depend not only on its type (i.e., ω) but also on its "intended use" (i.e., a)

> Platform and the consumer trade on **how** record will be used—i.e., which fee a platform will recommend to the merchant

Lindahl Economy

We let price of data depend not only on its type (i.e., ω) but also on its "intended use" (i.e., a)

> Platform and the consumer trade on **how** record will be used—i.e., which fee a platform will recommend to the merchant

This adapts to our setting the standard approach for modeling economies with externalities (Arrow 1969, Laffont 1976)

Lindahl Economy

We let price of data depend not only on its type (i.e., ω) but also on its "intended use" (i.e., a)

> Platform and the consumer trade on **how** record will be used—i.e., which fee a platform will recommend to the merchant

This adapts to our setting the standard approach for modeling economies with externalities (Arrow 1969, Laffont 1976)

This is reminiscent of GDPR: "*The **specific purposes** for which personal data are used should be determined at the time of the collection*"

Lindahl Economy

A market for each (a, ω), where ω-records can be traded for use a at price $p(a, \omega)$

Our equilibrium definition extends naturally to this richer economy

In particular, timing is the same

Lindahl Economy

A market for each (a, ω), where ω-records can be traded for use a at price $p(a, \omega)$

Our equilibrium definition extends naturally to this richer economy

In particular, timing is the same

Proposition

Equilibria of this economy are (unconstrained) efficient and maximize consumers' welfare, regardless of platform's incentives

conclusion

Summary

1. A framework to study competitive markets for personal data
2. Identify novel inefficiency leading this otherwise perfectly competitive market to fail

   Show how inefficiency critically depends on platform's role as an information intermediary

3. Propose three alternative market designs that fix inefficiency: data unions, data taxes, richer data prices

Efficiency Benchmark

To illustrate market failure, a less demanding efficiency benchmark is desirable:

Efficiency Benchmark

To illustrate market failure, a less demanding efficiency benchmark is desirable:

1. We require x° to be optimal given q° **for the platform**

   If not, detect inefficiency driven by platform lack of commitment in period 1

   (main results extend to "unconstrained" efficiency)

Efficiency Benchmark

To illustrate market failure, a less demanding efficiency benchmark is desirable:

1. We require x° to be optimal given q° **for the platform**

   If not, detect inefficiency driven by platform lack of commitment in period 1

   (main results extend to "unconstrained" efficiency)

2. We exclude merchant's payoff from $W(q, x)$

   If not, detect inefficiency driven by platform not fully internalizing merchant's payoff (main results extend to "aggregate" welfare)

Efficiency Benchmark

To illustrate market failure, a less demanding efficiency benchmark is desirable:

1. We require x° to be optimal given q° **for the platform**

   If not, detect inefficiency driven by platform lack of commitment in period 1

   (main results extend to "unconstrained" efficiency)

2. We exclude merchant's payoff from $W(q, x)$

   If not, detect inefficiency driven by platform not fully internalizing merchant's payoff (main results extend to "aggregate" welfare)

Bonus: In eqm, platform makes not profits. Thus, $W(q^*, x^*)$ equals consumer welfare. Thus, any constrained-efficient eqm maximizes consumer welfare

back

Efficiency of the Data Market

The **data market** in which platform and consumers participate is competitive:

Is it efficient? I.e., does it maximize the welfare of its participants?

Efficiency of the Data Market

The **data market** in which platform and consumers participate is competitive:

Is it efficient? I.e., does it maximize the welfare of its participants?

Definition

An allocation (q°, x°) is **efficient** if it solves

$$\begin{aligned} W^\circ = \quad & \max_{q,x} \quad V(q) + \sum_\omega q(\omega)U(\omega, x) + \sum_\omega \Big(\bar{q}(\omega) - q(\omega)\Big)\bar{r} \\ & \text{s.t.} \quad q \leq \bar{q} \;\text{ and }\; x \text{ solves platform' problem } \mathcal{P}_q \end{aligned}$$

Efficiency of the Data Market

The **data market** in which platform and consumers participate is competitive:

Is it efficient? I.e., does it maximize the welfare of its participants?

Definition

An allocation (q°, x°) is **efficient** if it solves

$$W^\circ = \max_{q,x} \quad V(q) + \sum_\omega q(\omega)U(\omega, x) + \sum_\omega \Big(\bar{q}(\omega) - q(\omega)\Big)\bar{r}$$

$$\text{s.t.} \quad q \leq \bar{q} \text{ and } x \text{ solves platform' problem } \mathcal{P}_q$$

To illustrate failure in data market, this less-demanding benchmark is desirable

(We also study "social" welfare and "unconstrained" efficiency discussion)