Competitive Markets for Personal Data

Simone Galperti   Tianhao Liu   Jacopo Perego

October 2024

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 (Bergemann et al. '23)

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

Model. A stylized **competitive economy** where

- Consumers own their data and can sell it to a platform
- Platform uses this data to interact consumers with a merchant

This Paper

Model. A stylized **competitive economy** where

- Consumers own their data and can sell it to a platform
- Platform uses this data to interact consumers with a merchant

Main Results

1. Identify **novel inefficiency** leading this perfectly competitive market to fail

This Paper

Model. A stylized **competitive economy** where

- Consumers own their data and can sell it to a platform
- Platform uses this data to interact consumers with a merchant

Main Results

1. Identify **novel inefficiency** leading this perfectly competitive market to fail

- When platform withholds information from merchant, consumers' decisions to sell data create externalities

This Paper

Model. A stylized **competitive economy** where

- Consumers own their data and can sell it to a platform
- Platform uses this data to interact consumers with a merchant

Main Results

1. Identify **novel inefficiency** leading this perfectly competitive market to fail
   - When platform withholds information from merchant, consumers' decisions to sell data create externalities
2. Propose three solutions to this market failure:
   - Data unions; Data taxes; "Lindahl" pricing for the data

Related Work

Model rooted in a GE tradition but leverages on progress in info-design literature, which offers microfoundation for key components of a data economy:

- E.g., how data is used (BBM '15); How data is valued (GLP '23)

Related Work

Model rooted in a GE tradition but leverages on progress in info-design literature, which offers microfoundation for key components of a data economy:

- E.g., how data is used (BBM '15); How data is valued (GLP '23)

We contribute to a recent literature that studies **data markets**:

- "Learning" 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 building on GLP '23

Related Work

Model rooted in a GE tradition but leverages on progress in info-design literature, which offers microfoundation for key components of a data economy:

- E.g., how data is used (BBM '15); How data is valued (GLP '23)

We contribute to a recent literature that studies **data markets**:

- "Learning" 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 building on GLP '23

More broadly, we contribute to the growing literature on the economics of platforms, data, & privacy Jones and Tonetti '20, Hidir and Vellodi '21, Chen '22

a stylized model

A Competitive Data Economy

One merchant, one (representative) platform, a unit mass of consumers

A Competitive Data Economy

One merchant, one (representative) platform, a unit mass of consumers

Each **consumer** has unit demand for merchant's product with a WTP of $\omega \in \Omega$

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

A Competitive Data Economy

One merchant, one (representative) platform, a unit mass of consumers

Each **consumer** has unit demand for merchant's product with a WTP of $\omega \in \Omega$

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

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

A Competitive Data Economy

One merchant, one (representative) platform, a unit mass of consumers

Each **consumer** has unit demand for merchant's product with a WTP of $\omega \in \Omega$

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

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

Two periods: 1. Data markets are open 2. Product market is open

Period 1: Data Markets

Period 1: 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: 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: 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

Notation: $\zeta(\omega) \in [0, 1]$ probability type-ω consumer sells her record

Period 1: 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

  Notation: $\zeta(\omega) \in [0, 1]$ probability type-ω consumer sells her record

- If type-ω consumer doesn't sell her record, she gets reservation utility $\bar{r}$

Period 2: Product Market

Given acquired database q, platform acts as **information designer**: (as in BBM)

- It sends merchant signal about each consumer in database
- Given signal, merchant charges each consumer a fee a
- Given a, type-ω consumer purchases product if $\omega \geq a$

Period 2: Product Market

Given acquired database q, platform acts as **information designer**: (as in BBM)

- It sends merchant signal about each consumer in database
- Given signal, merchant charges each consumer a fee a
- Given a, type-ω consumer purchases product if $\omega \geq a$

The **payoffs** in period 2 are:

Consumer's: $u(a, \omega) = \max\{\omega - a, 0\}$

Merchant's: $\pi(a, \omega) = a \, \mathbb{1}(\omega \geq a)$

Platform's: $v(a, \omega) = \gamma_u \, u(a, \omega) + \gamma_\pi \, \pi(a, \omega)$

Period 2: Product Market

Given acquired database q, platform acts as **information designer**: (as in BBM)

- It sends merchant signal about each consumer in database
- Given signal, merchant charges each consumer a fee a
- Given a, type-ω consumer purchases product if $\omega \geq a$

Info-design problem equiv to platform choosing mechanism $x : \Omega \to \Delta(A)$ s.t.

$$\begin{aligned} V(q) = &\max_{x:\Omega\to\Delta(A)} \sum_{\omega,a} v(a,\omega)x(a|\omega)q(\omega) \\ &\text{s.t. } \forall a, a': \sum_{\omega} \Big(\pi(a,\omega) - \pi(a',\omega)\Big)x(a|\omega)q(\omega) \geq 0 \end{aligned} \tag{$\mathcal{P}_q$}$$

(canonical ID problem, but with endogenous q)

Equilibrium

Definition

A profile (p^*, ζ^*, q^*, x^*)

Equilibrium

Definition

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

Equilibrium

Definition

A profile (p^*, ζ^*, 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^*, ζ^*, 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 the platform's problem $\mathcal{P}_{q^*}$ in the second-period

Equilibrium

Definition

A profile (p^*, ζ^*, 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 the platform's problem $\mathcal{P}_{q^*}$ in the second-period

(c). Given p^* and x^*, ζ^* solves consumers' problem, i.e.,

$$\zeta^*(\omega) \in \arg\max_{z \in [0,1]} z\Big(p^*(\omega) + \underbrace{\sum_a x^*(a|\omega) u(a, \omega)}_{U(\omega, x^*)}\Big) + (1 - z)\bar{r}$$

Equilibrium

Definition

A profile (p^*, ζ^*, 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 the platform's problem $\mathcal{P}_{q^*}$ in the second-period

(c). Given p^* and x^*, ζ^* solves consumers' problem, i.e.,

$$\zeta^*(\omega) \in \arg\max_{z \in [0,1]} z \Big(p^*(\omega) + \underbrace{\sum_a x^*(a|\omega) u(a, \omega)}_{U(\omega, x^*)} \Big) + (1 - z)\bar{r}$$

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

Discussion

Substantive assumptions:

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

Main insights extend to more general **intermediation problems**:

- Multiple agents (e.g., competing merchants).
- Arbitrary payoffs (e.g., second-price auctions)
- Beyond information design (e.g., platform also sets reserve price)

Leading applications: online marketplaces & online ad auctions

efficiency

Efficiency Benchmark

Does data market "efficiently" allocate records btw consumers and platform?

Efficiency Benchmark

Does data market "efficiently" allocate records btw consumers and platform?

Definition

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

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

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

Efficiency Benchmark

Does data market "efficiently" allocate records btw consumers and platform?

Definition

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

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

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

Results extend to "social" welfare and "unconstrained" efficiency discussion

inefficiency of the data economy

“Social” Cost and Benefit of Data Records

“Social” Cost and Benefit of Data Records

What’s the **social cost** of allocating a ω-record to the platform’s database?

– It’s $\bar{r} \geq 0$

“Social” Cost and Benefit of Data Records

What’s the **social cost** of allocating a ω-record to the platform’s database?

– It’s $\bar{r} \geq 0$

What’s the **social benefit** of allocating a ω-record to the platform’s database?

“Social” Cost and Benefit of Data Records

What’s the **social cost** of allocating a ω-record to the platform’s database?

- It’s $\bar{r} \geq 0$

What’s the **social benefit** of allocating a ω-record to the platform’s database?

- Fix any $q \leq \bar{q}$. Consider the following maximization problem:

$$W(q) \triangleq \max_{x:\Omega\to\Delta(A)} \quad V(q) + \sum_{\omega} q(\omega)U(\omega, x)$$
$$\text{s.t.} \qquad x \text{ solves } \mathcal{P}_q$$

“Social” Cost and Benefit of Data Records

What’s the **social cost** of allocating a ω-record to the platform’s database?

– It’s $\bar{r} \geq 0$

What’s the **social benefit** of allocating a ω-record to the platform’s database?

– Fix any $q \leq \bar{q}$. Consider the following maximization problem:

$$\begin{aligned} W(q) \;\triangleq\; & \max_{x:\Omega \to \Delta(A)} && V(q) + \sum_{\omega} q(\omega) U(\omega, x) \\ & \text{s.t.} && x \text{ solves } \mathcal{P}_q \end{aligned}$$

and denote by $\Psi_q \subset \mathbb{R}_+^{\Omega}$ the **supergradients** of $W(q)$ (a.s. a singleton)

"Social" Cost and Benefit of Data Records

What's the **social cost** of allocating a ω-record to the platform's database?

- It's $\bar{r} \geq 0$

What's the **social benefit** of allocating a ω-record to the platform's database?

- Fix any $q \leq \bar{q}$. Consider the following maximization problem:

$$W(q) \triangleq \max_{x:\Omega\to\Delta(A)} \quad V(q) + \sum_{\omega} q(\omega)U(\omega, x)$$
$$\text{s.t.} \quad x \text{ solves } \mathcal{P}_q$$

and denote by $\Psi_q \subset \mathbb{R}_+^{\Omega}$ the **supergradients** of $W(q)$ (a.s. a singleton)

$\psi_q(\omega)$ is change in $W(q)$ from adding a ω-record to $q \rightsquigarrow$ **social benefit**

Constrained-Efficient Allocations

Using these two concepts, we characterize constrained-efficient allocations

Proposition

An allocation (q, x) is constrained efficient **if and only if** x solves $\mathcal{P}_q$ and there is a $\psi \in \Psi_q$ s.t.

- If $q(\omega) > 0$, then $\psi(\omega) \geq \bar{r}$
- If $q(\omega) < \bar{q}(\omega)$, then $\psi(\omega) \leq \bar{r}$

Constrained-Efficient Allocations

Using these two concepts, we characterize constrained-efficient allocations

Proposition

An allocation (q, x) is constrained efficient **if and only if** x solves $\mathcal{P}_q$ and there is a $\psi \in \Psi_q$ s.t.

- If $q(\omega) > 0$, then $\psi(\omega) \geq \bar{r}$
- If $q(\omega) < \bar{q}(\omega)$, then $\psi(\omega) \leq \bar{r}$

Intuition: Planner's problem is concave, "FOC" is necessary and sufficient

“Private” Cost and Benefit of Data Records

“Private” Cost and Benefit of Data Records

Fix an equilibrium (p^*, ζ^*, q^*, x^*)

The **“private” benefit** for a type-ω consumer when she sells her record is

$$G^*(\omega) \triangleq p^*(\omega) + \underbrace{U(\omega, x^*)}_{\sum_a x^*(a,\omega) u(a,\omega)}$$

“Private” Cost and Benefit of Data Records

Fix an equilibrium (p^*, ζ^*, q^*, x^*)

The **“private” benefit** for a type-ω consumer when she sells her record is

$$G^*(\omega) \triangleq p^*(\omega) + \underbrace{U(\omega, x^*)}_{\sum_a x^*(a,\omega)u(a,\omega)}$$

In equilibrium, it must hold that

- If $q^*(\omega) > 0$, then $G^*(\omega) \geq \bar{r}$
- If $q^*(\omega) < \bar{q}(\omega)$, then $G^*(\omega) \leq \bar{r}$

"Private" Cost and Benefit of Data Records

Fix an equilibrium (p^*, ζ^*, q^*, x^*)

The **"private" benefit** for a type-ω consumer when she sells her record is

$$G^*(\omega) \triangleq p^*(\omega) + \underbrace{U(\omega, x^*)}_{\sum_a x^*(a,\omega)u(a,\omega)}$$

In equilibrium, it must hold that

- If $q^*(\omega) > 0$, then $G^*(\omega) \geq \bar{r}$
- If $q^*(\omega) < \bar{q}(\omega)$, then $G^*(\omega) \leq \bar{r}$

Thus, an eqm is efficient if and only if the social (ψ_{q^*}) and private (G^*) benefit of data records are sufficiently "aligned"

A Data Externality (intuition)

Thus, equilibrium **private benefit** of selling ω-record:

$$G^*(\omega) = p^*(\omega) + U(\omega, x^*)$$

A Data Externality (intuition)

Thus, equilibrium **private benefit** of selling ω-record:

$$G^*(\omega) = p^*(\omega) + U(\omega, x^*)$$

The **social benefit** of selling ω-record is:

$$\psi_{q^*}(\omega) = \frac{\partial V(q^*)}{\partial q^*(\omega)} + U(\omega, x^*) + \underbrace{\sum_{\omega'} q(\omega') \frac{\partial U(\omega', x^*)}{\partial q^*(\omega)}}_{\text{change in the trading surplus of all consumers}}$$

A Data Externality (intuition)

Thus, equilibrium **private benefit** of selling ω-record:

$$G^*(\omega) = p^*(\omega) + U(\omega, x^*)$$

The **social benefit** of selling ω-record is:

$$\psi_{q^*}(\omega) = \frac{\partial V(q^*)}{\partial q^*(\omega)} + U(\omega, x^*) + \underbrace{\sum_{\omega'} q(\omega') \frac{\partial U(\omega', x^*)}{\partial q^*(\omega)}}_{\substack{\text{change in the trading} \\ \text{surplus of all consumers}}}$$

In equilibrium, platform's zero-profit requires $\dfrac{\partial V(q^*)}{\partial q^*(\omega)} = p^*(\omega)$ GLP23

A Data Externality (intuition)

Thus, equilibrium **private benefit** of selling ω-record:

$$G^*(\omega) = p^*(\omega) + U(\omega, x^*)$$

The **social benefit** of selling ω-record is:

$$\psi_{q^*}(\omega) = p^*(\omega) + U(\omega, x^*) + \underbrace{\sum_{\omega'} q(\omega') \frac{\partial U(\omega', x^*)}{\partial q^*(\omega)}}_{\text{change in the trading surplus of all consumers}}$$

In equilibrium, platform's zero-profit requires $\dfrac{\partial V(q^*)}{\partial q^*(\omega)} = p^*(\omega)$ GLP23

A Data Externality (intuition)

Thus, equilibrium **private benefit** of selling ω-record:

$$G^*(\omega) = p^*(\omega) + U(\omega, x^*)$$

The **social benefit** of selling ω-record is:

$$\psi_{q^*}(\omega) = p^*(\omega) + U(\omega, x^*) + \underbrace{\sum_{\omega'} q(\omega') \frac{\partial U(\omega', x^*)}{\partial q^*(\omega)}}_{\substack{\text{a non pecuniary} \\ \textbf{data externality}}}$$

In equilibrium, platform's zero-profit requires $\dfrac{\partial V(q^*)}{\partial q^*(\omega)} = p^*(\omega)$ GLP23

Main Characterization

This externality depends on how platform uses the data, and thus on its objective

Recall: platform's objective is $v(a,\omega) = \gamma_u \, u(a,\omega) + \gamma_\pi \, \pi(a,\omega)$

Main Characterization

This externality depends on how platform uses the data, and thus on its objective

Recall: platform's objective is $v(a,\omega) = \gamma_u \; u(a,\omega) + \gamma_\pi \; \pi(a,\omega)$

Proposition

- If $\gamma_u < \gamma_\pi$, all equilibria are constrained efficient and thus consumers' welfare is maximal
- If $\gamma_u \geq \gamma_\pi$, an r exists such that all equilibria of the economy are inefficient

Main Characterization

This externality depends on how platform uses the data, and thus on its objective

Recall: platform's objective is $v(a,\omega) = \gamma_u \, u(a,\omega) + \gamma_\pi \, \pi(a,\omega)$

Proposition

- If $\gamma_u < \gamma_\pi$, all equilibria are constrained efficient and thus consumers' welfare is maximal
- If $\gamma_u \geq \gamma_\pi$, an r exists such that all equilibria of the economy are inefficient

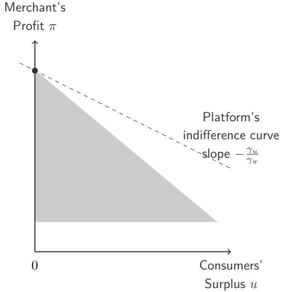
Equilibrium efficient when platform cares more about merchant $\rightsquigarrow$ **Why?**

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$

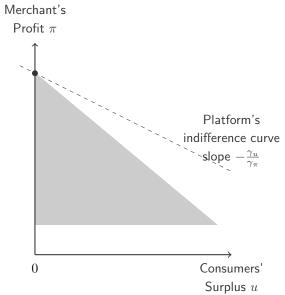
Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



Intuition for Efficiency

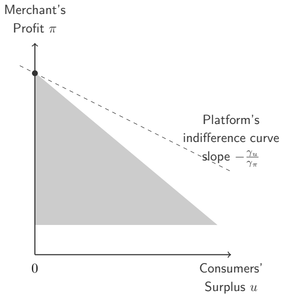
If $\gamma_u < \gamma_\pi$



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

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



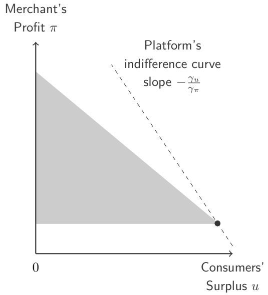
- At all q, **full disclosure** is optimal
- Merchant extracts surplus from all consumers
- Therefore, $x^*(a, \omega)$ does not depend on q
- Therefore, no externality! All equilibria are constrained efficient

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$

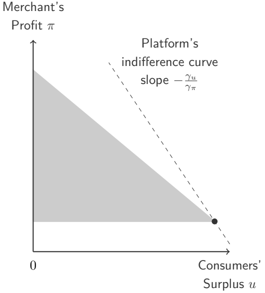
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



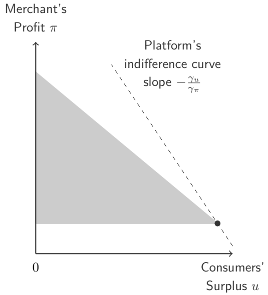
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



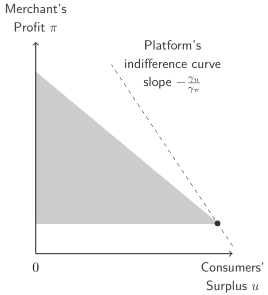
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



Intuition for Inefficiency

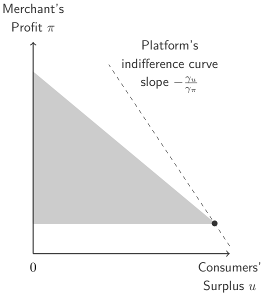
If $\gamma_u > \gamma_\pi$



– Platform **withholds information** from merchant

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



- Platform **withholds information** from merchant
- Optimal mechanism x^* depends on q
- Thus, $\sum_{\omega'} q(\omega') \frac{\partial U(\omega', x^*)}{\partial q^*(\omega)}$ can be non-zero
- Example: think of lowest-type consumer

example

A Simple Example to Illustrate

Suppose:

- $\gamma_u > \gamma_\pi = 0$, i.e. platform only cares about consumers' surplus
- Only two types of consumers: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Suppose $\bar{r} < \frac{1+\gamma_u}{2}$ so that some trade is efficient

A Simple Example to Illustrate

Suppose:

- $\gamma_u > \gamma_\pi = 0$, i.e. platform only cares about consumers' surplus
- Only two types of consumers: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Suppose $\bar{r} < \frac{1+\gamma_u}{2}$ so that some trade is efficient

There is a **unique** constrained-efficient allocation (q°, x°):

- All low-type consumers sell: $q^\circ(1) = \bar{q}(1)$
- Only some high-type consumers sell: $q^\circ(2) = \bar{q}(1) < \bar{q}(2)$
- Platform provides no info to merchant, who charges lowest fee to all consumers in database: $x^\circ(a = 1|\omega) = 1, \ \forall \omega$

A Simple Example to Illustrate

Suppose:

- $\gamma_u > \gamma_\pi = 0$, i.e. platform only cares about consumers' surplus
- Only two types of consumers: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Suppose $\bar{r} < \frac{1+\gamma_u}{2}$ so that some trade is efficient

There is a **unique** constrained-efficient allocation (q°, x°):

- All low-type consumers sell: $q^\circ(1) = \bar{q}(1)$
- Only some high-type consumers sell: $q^\circ(2) = \bar{q}(1) < \bar{q}(2)$
- Platform provides no info to merchant, who charges lowest fee to all consumers in database: $x^\circ(a = 1|\omega) = 1, \ \forall\omega$

A Simple Example to Illustrate

Claim: If $\gamma_u < \bar{r}$, all equilibria are inefficient $\rightsquigarrow$ no trade (Corollary 1)

A Simple Example to Illustrate

Claim: If $\gamma_u < \bar{r}$, all equilibria are inefficient $\rightsquigarrow$ no trade (Corollary 1)

It can be shown that $p^*(\omega) \leq \gamma_u$. This implies that:

- Low-type consumers do not want to sell their records, $q^*(1) = 0$

  Why? $U^*(1) = p^*(1) \leq \gamma_u < \bar{r}$

  Do not internalize positive externality that selling their record generate for high-type consumers

A Simple Example to Illustrate

Claim: If $\gamma_u < \bar{r}$, all equilibria are inefficient $\rightsquigarrow$ no trade (Corollary 1)

It can be shown that $p^*(\omega) \leq \gamma_u$. This implies that:

- Low-type consumers do not want to sell their records, $q^*(1) = 0$

  Why? $U^*(1) = p^*(1) \leq \gamma_u < \bar{r}$

  Do not internalize positive externality that selling their record generate for high-type consumers

- Hence, high-type consumer do not want to sell either

  Why? $U^*(2) = p^*(2) \leq \gamma_u < \bar{r}$

A Simple Example to Illustrate

Claim: If $\gamma_u < \bar{r}$, all equilibria are inefficient $\rightsquigarrow$ no trade (Corollary 1)

It can be shown that $p^*(\omega) \leq \gamma_u$. This implies that:

- Low-type consumers do not want to sell their records, $q^*(1) = 0$

  Why? $U^*(1) = p^*(1) \leq \gamma_u < \bar{r}$

  Do not internalize positive externality that selling their record generate for high-type consumers

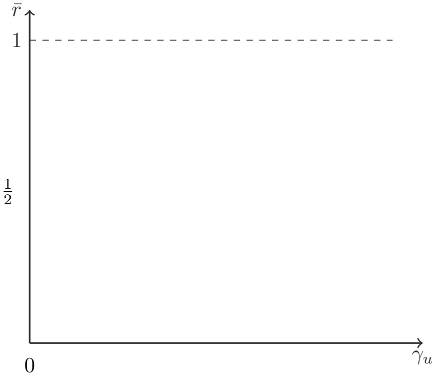
- Hence, high-type consumer do not want to sell either

  Why? $U^*(2) = p^*(2) \leq \gamma_u < \bar{r}$

- Market unravels $\rightsquigarrow$ No trade $\rightsquigarrow$ Inefficiency

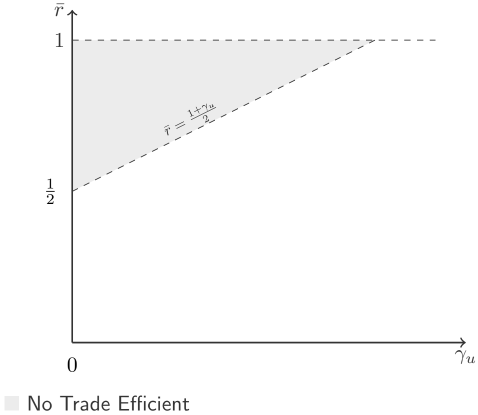
Example

Complete equilibrium characterization for this example:



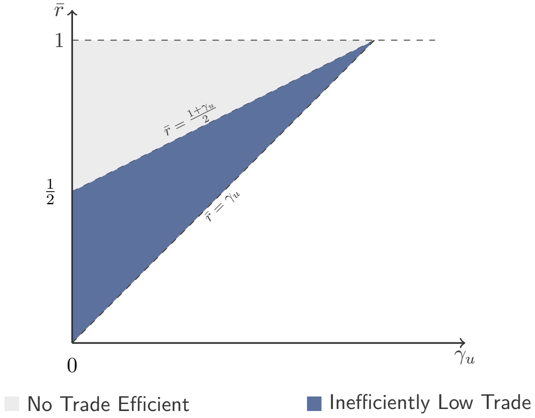
Example

Complete equilibrium characterization for this example:



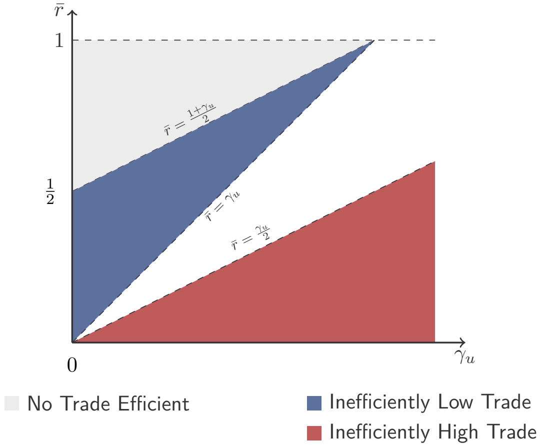
Example

Complete equilibrium characterization for this example:



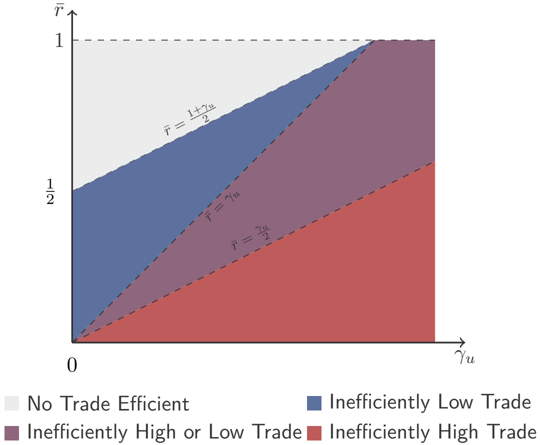
Example

Complete equilibrium characterization for this example:



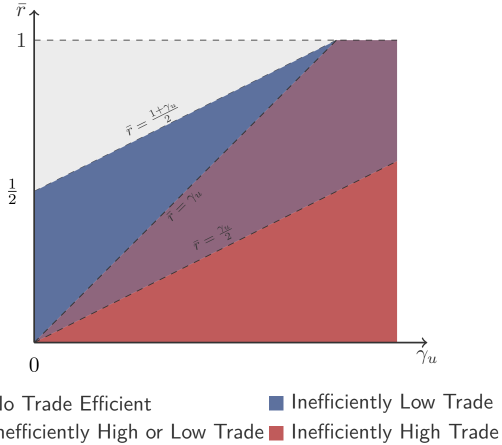
Example

Complete equilibrium characterization for this example:



Example

Complete equilibrium characterization for this example:



generalizations

A More General Approach

General information-intermediation problems:

- n agents (e.g., competing merchants).
- arbitrary payoffs $\pi_i(a, \omega)$, $v(a, \omega)$ (e.g., revenue maximization in auctions)
- beyond information design (e.g., platform also sets reserve price)

These problems preserve $\mathcal{P}_q$'s LP structure $\Rightarrow V(q)$ has similar properties

Leading applications: online marketplaces & online ad auctions

Conjecture

If full-disclosure is uniquely optimal at all q's, all competitive equilibria are efficient. Otherwise, market can fail.

Stepping Back

Information intermediaries play ubiquitous role in digital markets

They often balance interests of conflicting parties (sellers-buyers, drivers-riders)

They do so by optimally withholding some information from the agents

This paper illustrates how this practice can lead to market failure remedies

Remedies

How to fix this market failure?

We explore three alternative market designs:

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:

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

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 stylized framework to study competitive markets for personal data

   Rooted in GE tradition but leveraging recent progress in info-design

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

Competitive Markets for Personal Data

Simone Galperti
UCSD

Tianhao Liu
Columbia

Jacopo Perego
Columbia

Thank You!

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