Competitive Markets for Personal Data

Simone Galperti — UCSD
Tianhao Liu — Columbia
Jacopo Perego — Columbia

December 2023

Motivation

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

Yet, they often have **limited control** over who uses it and are **imperfectly compensated** in return

- Expropriation and barter, common practice in the industry (FTC '15)

This status quo may be source of market failures (Seim et al. '22)

Could a competitive market for data avoid these problems?

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 Result.

- **Inefficiency**, despite competitive economy and property rights
- Inefficiency stems from an **externality** consumers exert on each other, enabled by how platform uses their data — Galperti, Levkun, Perego (2023)

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 Result.

- **Inefficiency**, despite competitive economy and property rights
- Inefficiency stems from an **externality** consumers exert on each other, enabled by how platform uses their data Galperti, Levkun, Perego (2023)

Solutions. Three alternative market designs to avoid inefficiency:

- Introducing **data taxes** on consumers
- Introducing a **data union**
- Making data markets **more complete**

Literature

This paper contributes to a growing literature on data markets

Bergemann, Bonatti 19

Our model is rooted in GE tradition, but leverages recent progress in information-design literature

Bergemann, Morris '19

Literature

This paper contributes to a growing literature on data markets

Bergemann, Bonatti 19

Our model is rooted in GE tradition, but leverages recent progress in information-design literature Bergemann, Morris '19

A data market can fail due to "**learning externalities**" enabled by exogenous correlation in consumers data Bergemann et al. ('22), Acemoglu et al. ('22) Choi et al ('19), Ichihashi ('21)

Literature

This paper contributes to a growing literature on data markets

Bergemann, Bonatti 19

Our model is rooted in GE tradition, but leverages recent progress in information-design literature Bergemann, Morris '19

A data market can fail due to "**learning externalities**" enabled by exogenous correlation in consumers data Bergemann et al. ('22), Acemoglu et al. ('22) Choi et al ('19), Ichihashi ('21)

This paper.

- Emphasizes a novel source of inefficiency Galperti et al. '23
- Explores novel designs for data markets

model

A Simple Data Economy

One merchant, one platform, a unit mass of consumers

A Simple Data Economy

One merchant, one platform, a unit mass of consumers

The **merchant** wants to sell widgets to consumers (zero MC)

Each **consumer** has unit demand for widget and WTP $\omega \in \Omega$ (finite)

Consumer's WTP distributed as $\bar{q} \in \Delta(\Omega)$

A Simple Data Economy

One merchant, one platform, a unit mass of consumers

The **merchant** wants to sell widgets to consumers (zero MC)

Each **consumer** has unit demand for widget and WTP $\omega \in \Omega$ (finite)

Consumer's WTP distributed as $\bar{q} \in \Delta(\Omega)$

Each consumer owns a **data record** that, if sold to **platform**, reveals ω and provides access to her

A Simple Data Economy

One merchant, one platform, a unit mass of consumers

The **merchant** wants to sell widgets to consumers (zero MC)

Each **consumer** has unit demand for widget and WTP $\omega \in \Omega$ (finite)

Consumer's WTP distributed as $\bar{q} \in \Delta(\Omega)$

Each consumer owns a **data record** that, if sold to **platform**, reveals ω and provides access to her

Two periods: 1. Data is traded, 2. Data is used

Period 1: Data Trade

Period 1: Data Trade

Platform and consumers trade the data records

Period 1: Data Trade

Platform and consumers trade the data records

Demand Side:

For each ω-record the platform buys, it pays price $p(\omega)$

Supply Side:

If type-ω consumer sells her record, she is paid price $p(\omega)$

(on top of the "service" offered by platform)

Notation: $\zeta(\omega) \in [0, 1]$ denotes probability of selling

Period 1: Data Trade

Platform and consumers trade the data records

Demand Side:

For each ω-record the platform buys, it pays price $p(\omega)$

Supply Side:

If type-ω consumer sells her record, she is paid price $p(\omega)$

(on top of the "service" offered by platform)

Notation: $\zeta(\omega) \in [0, 1]$ denotes probability of selling

If consumer does not sell, she forgoes platform's "service" and obtain reservation utility $r(\omega)$

Period 2: Data Use

Given acquired database $q \in \mathbb{R}_+^\Omega$, platform acts as **information designer**:

- It sends merchant signal about each consumer in database
- Given signal, merchant charges each consumer a personal fee a
- Given a, consumer chooses whether to purchase the widget

Period 2: Data Use

Given acquired database $q \in \mathbb{R}_+^{\Omega}$, platform acts as **information designer**:

- It sends merchant signal about each consumer in database
- Given signal, merchant charges each consumer a personal fee a
- Given a, consumer chooses whether to purchase the widget

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: Data Use

Given acquired database $q \in \mathbb{R}^{\Omega}_{+}$, platform acts as **information designer**:

- It sends merchant signal about each consumer in database
- Given signal, merchant charges each consumer a personal fee a
- Given a, consumer chooses whether to purchase the widget

I.e., Platform chooses recommendation mechanism $x : \Omega \to \Delta(A)$ to solve

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

(ID problem with endogenous q)

Equilibrium

Definition

A profile (p^*, ζ^*, q^*, x^*)

Equilibrium

Definition

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

Equilibrium

Definition

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

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

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

Equilibrium

Definition

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

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

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

(b). Given q^*, x^* solves the platform's second-period problem $\mathcal{P}_{q^*}$

Equilibrium

Definition

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

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

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

(b). Given q^*, x^* solves the platform's second-period problem $\mathcal{P}_{q^*}$

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

Equilibrium

Definition

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

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

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

(b). Given q^*, x^* solves the platform's second-period problem $\mathcal{P}_{q^*}$

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

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

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

Discussion

A stylized competitive data economy:

- Fully revealing records (see Galperti, Levkun, Perego 23)
- Platform's payoff $v(a,\omega) = \gamma_u \; u(a,\omega) + \gamma_\pi \; \pi(a,\omega)$ (see Xu, Yang 22)
- A single platform + price taking behavior
- Exogenous $r(\omega)$ (e.g., no"showrooming", see Bergemann et al '22)
- Records bundles access and information

  No access without information

efficiency

Efficiency Benchmark

Goal. Study the efficiency of the market in which consumers and platform trade data records

Efficiency Benchmark

Goal. Study the efficiency of the market in which consumers and platform trade data records

We call (q, x) a **data allocation**

Denote by $W(q, x)$ the total welfare of consumers and platform

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

Efficiency Benchmark

Definition

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

$$\begin{aligned} \max_{q,x} \quad & W(q,x) \\ \text{s.t.} \quad & q \leq \bar{q} \;\text{ and }\; x \text{ solves } \mathcal{P}_q. \end{aligned}$$

Efficiency Benchmark

Definition

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

$$\max_{q,x} \quad W(q,x)$$
$$\text{s.t.} \quad q \leq \bar{q} \ \text{ and } \ x \text{ solves } \mathcal{P}_q.$$

Constrained: planner can't force x° on platform $\rightsquigarrow$ a lower bar $\rightsquigarrow$ a cleaner negative result

Efficiency Benchmark

Definition

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

$$\begin{aligned} \max_{q,x} \quad & W(q,x) \\ \text{s.t.} \quad & q \leq \bar{q} \ \text{ and } \ x \text{ solves } \mathcal{P}_q. \end{aligned}$$

Constrained: planner can't force x° on platform $\rightsquigarrow$ a lower bar $\rightsquigarrow$ a cleaner negative result

We ignore merchant from $W(q,x)$ $\rightsquigarrow$ a lower bar $\rightsquigarrow$ a cleaner negative result

Efficiency Benchmark

Definition

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

$$\max_{q,x} \quad W(q,x)$$
$$\text{s.t.} \quad q \leq \bar{q} \ \text{ and } \ x \text{ solves } \mathcal{P}_q.$$

Constrained: planner can't force x° on platform $\rightsquigarrow$ a lower bar $\rightsquigarrow$ a cleaner negative result

We ignore merchant from $W(q,x)$ $\rightsquigarrow$ a lower bar $\rightsquigarrow$ a cleaner negative result

- When $\gamma_\pi < 1$, platform does not fully internalize merchant's payoff $\rightsquigarrow$ a source of inefficiency: not new, just distracting

Efficiency Benchmark

Definition

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

$$\max_{q,x} \quad W(q,x)$$
$$\text{s.t.} \quad q \leq \bar{q} \;\; \text{and} \;\; x \text{ solves } \mathcal{P}_q.$$

Constrained: planner can't force x° on platform $\rightsquigarrow$ a lower bar $\rightsquigarrow$ a cleaner negative result

We ignore merchant from $W(q,x)$ $\rightsquigarrow$ a lower bar $\rightsquigarrow$ a cleaner negative result

- When $\gamma_\pi < 1$, platform does not fully internalize merchant's payoff $\rightsquigarrow$ a source of inefficiency: not new, just distracting
- Appendix shows how main result extends to "social" welfare

equilibrium

Competitive Equilibrium

First, we establish existence

Proposition

An equilibrium (p^*, ζ^*, q^*, x^*) of the competitive economy exists

Competitive Equilibrium

First, we establish existence

Proposition

An equilibrium (p^*, ζ^*, q^*, x^*) of the competitive economy exists

Result is not off-the-shelf, since consumers behavior depends on p *and* x

- Suffice to show that solution correspondence of $\mathcal{P}_q$ is UHC in q

Competitive Equilibrium

Remark. In equilibrium, platform's payoff is zero

Competitive Equilibrium

Remark. In equilibrium, platform's payoff is zero

This implies that, if an equilibrium allocation is constrained efficient, consumer welfare is maximized

Competitive Equilibrium

Remark. In equilibrium, platform's payoff is zero

This implies that, if an equilibrium allocation is constrained efficient, consumer welfare is maximized

Therefore, are equilibrium allocations constrained efficient?

Inefficiency of the Competitive Economy

It depends on how platform uses data, which depends on its objective

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

Inefficiency of the Competitive Economy

It depends on how platform uses data, which depends on its objective

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

Proposition

- If $\gamma_u < \gamma_\pi$, equilibria are constrained efficient and consumers' welfare is maximized
- If $\gamma_u \geq \gamma_\pi$, equilibria can be inefficient (and consumers' welfare can be as low as $R = \sum_\omega r(\omega)\bar{q}(\omega)$)

Inefficiency of the Competitive Economy

It depends on how platform uses data, which depends on its objective

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

Proposition

- If $\gamma_u < \gamma_\pi$, equilibria are constrained efficient and consumers' welfare is maximized
- If $\gamma_u \geq \gamma_\pi$, equilibria can be inefficient (and consumers' welfare can be as low as $R = \sum_\omega r(\omega)\bar{q}(\omega)$)

Equilibrium maximizes consumers welfare when platform cares more about merchant's payoff $\rightsquigarrow$ **Why?**

Inefficiency of the Competitive Economy

It depends on how platform uses data, which depends on its objective

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

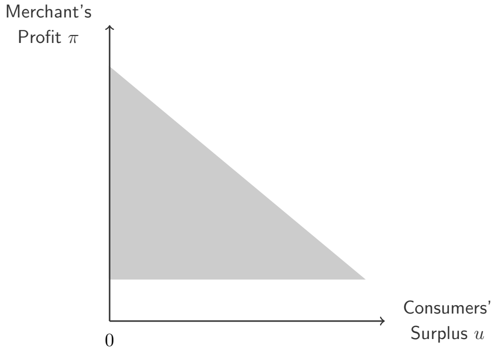
Proposition

- If $\gamma_u < \gamma_\pi$, equilibria are constrained efficient and consumers' welfare is maximized
- If $\gamma_u \geq \gamma_\pi$, equilibria can be inefficient (and consumers' welfare can be as low as $R = \sum_\omega r(\omega)\bar{q}(\omega)$)

Idea. How platform's uses the data can enable externalities

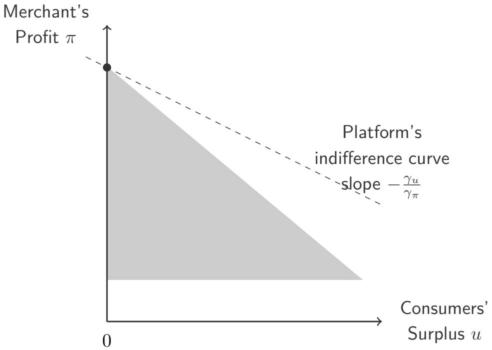
Intuition for Inefficiency

Recall "surplus triangle" from Bergemann et al. (2015):



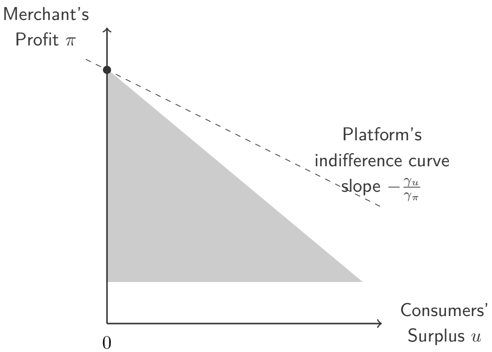
Intuition for Inefficiency

Recall "surplus triangle" from Bergemann et al. (2015):



Intuition for Inefficiency

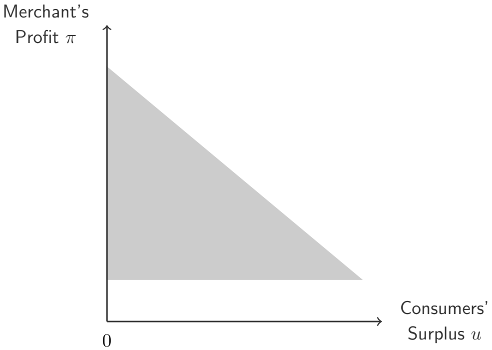
Recall "surplus triangle" from Bergemann et al. (2015):



If $\gamma_u < \gamma_\pi$, optimal x^* involves "**full disclosure**," regardless of q.
Merchant learns consumers' types and extract their surplus

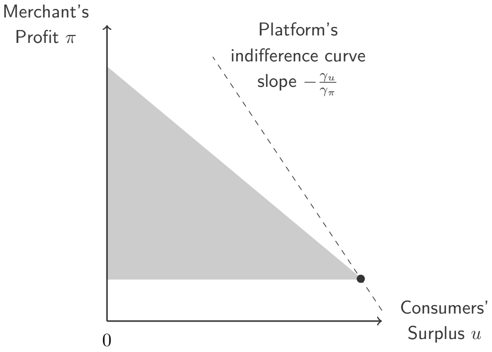
Intuition for Inefficiency

Recall "surplus triangle" from Bergemann et al. (2015):



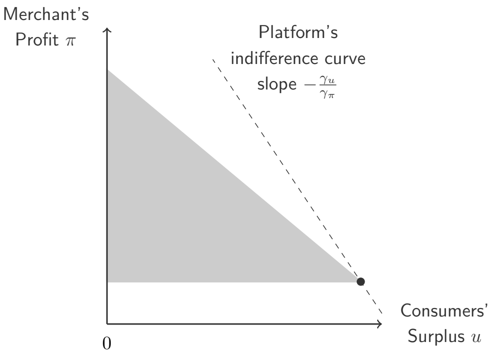
Intuition for Inefficiency

Recall "surplus triangle" from Bergemann et al. (2015):



Intuition for Inefficiency

Recall “surplus triangle” from Bergemann et al. (2015):



If $\gamma_u \geq \gamma_\pi$, optimal x^* involves **pooling**

It prevents merchant from extracting too much surplus

Intuition for Inefficiency

If $\gamma_u \geq \gamma_\pi$, platform pools consumers of diff types to prevent merchant learning their types

The composition of the pool determines merchant's beliefs, thus his fee

If one consumer does not sell her data, she affects pool composition and, thus, other consumers' payoff

Consumers exert **externality** on each other

Intuition for Inefficiency

If $\gamma_u \geq \gamma_\pi$, platform pools consumers of diff types to prevent merchant learning their types

The composition of the pool determines merchant's beliefs, thus his fee

If one consumer does not sell her data, she affects pool composition and, thus, other consumers' payoff

Consumers exert **externality** on each other

Build on Galperti et al (2023): from platform's values to consumers' payoffs

This paper: competitive economy enables this externality to a degree that leads to inefficiency

A Simple Example to Illustrate

Suppose:

- $\gamma_\pi = 0$ (i.e. platform maximizes consumers' surplus)
- Only two types: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Same outside option: $r(\omega) = \bar{r} \in (0, 1)$, for all ω

A Simple Example to Illustrate

Suppose:

- $\gamma_\pi = 0$ (i.e. platform maximizes consumers' surplus)
- Only two types: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Same outside option: $r(\omega) = \bar{r} \in (0, 1)$, for all ω

The constrained-efficient allocation (q°, x°) involves:

- All low-type consumers participate: $q^\circ(1) = \bar{q}(1)$
- Not all high-type consumers participate: $q^\circ(2) = \bar{q}(1) < \bar{q}(2)$
- All participating consumers charged a low fee: $x^\circ(1|\omega) = 1, \ \forall\omega$

A Simple Example to Illustrate

Suppose:

- $\gamma_\pi = 0$ (i.e. platform maximizes consumers' surplus)
- Only two types: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Same outside option: $r(\omega) = \bar{r} \in (0, 1)$, for all ω

The constrained-efficient allocation (q°, x°) involves:

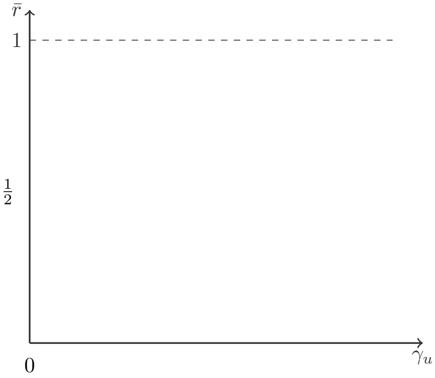
- All low-type consumers participate: $q^\circ(1) = \bar{q}(1)$
- Not all high-type consumers participate: $q^\circ(2) = \bar{q}(1) < \bar{q}(2)$
- All participating consumers charged a low fee: $x^\circ(1|\omega) = 1, \ \forall\omega$

If $\gamma_u < \bar{r}$, any competitive equilibrium has $p^*(1) < \bar{r}$. Thus,

- Low-type consumers do not sell their records $\rightsquigarrow$ neg externality
- Hence, high-type consumer do not want to sell
- Market unravels $\rightsquigarrow$ No trade $\rightsquigarrow$ Inefficiency

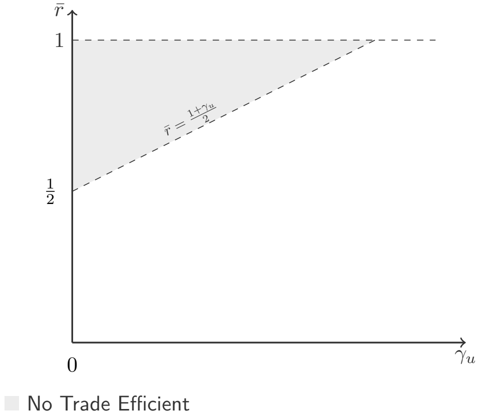
Example

Complete equilibrium characterization for this example:



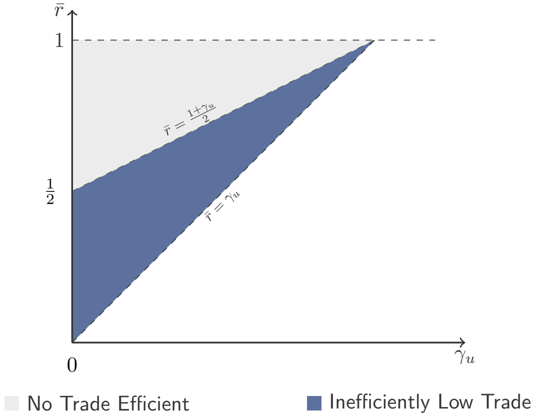
Example

Complete equilibrium characterization for this example:



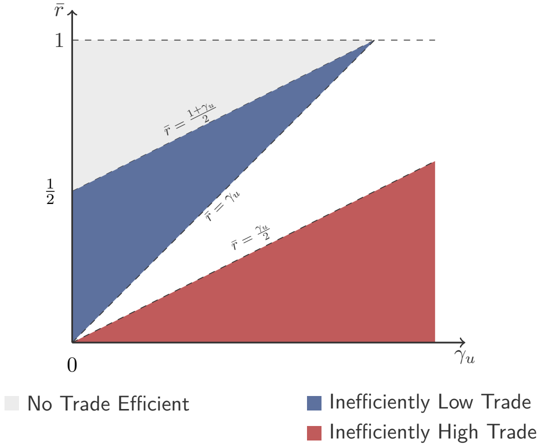
Example

Complete equilibrium characterization for this example:



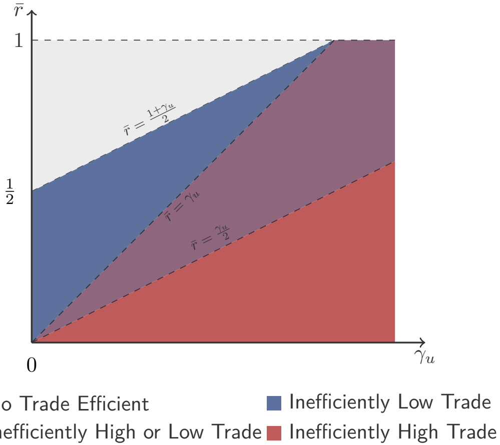
Example

Complete equilibrium characterization for this example:

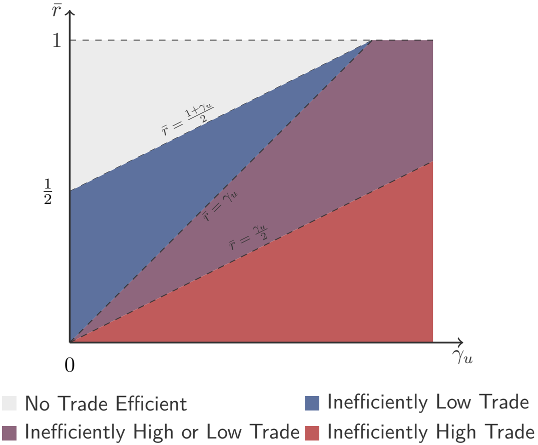


Example

Complete equilibrium characterization for this example:



Complete equilibrium characterization for this example:



A Characterization of Inefficient Equilibria

A formal characterization of inefficient equilibria

A Characterization of Inefficient Equilibria

A formal characterization of inefficient equilibria

Denote by $\psi_q(\omega)$ the marginal **social value** of ω-records given q

- Marginal change in $W(q, x_q^*)$ if we allocate additional record ω to platform's database
- Galperti et al. '23 characterize such values

A Characterization of Inefficient Equilibria

A formal characterization of inefficient equilibria

Denote by $\psi_q(\omega)$ the marginal **social value** of ω-records given q

- Marginal change in $W(q, x_q^*)$ if we allocate additional record ω to platform's database
- Galperti et al. '23 characterize such values

Fixing q and letting the uninformed-merchant fee be a_q:

$$\psi_q(\omega) = \begin{cases} \gamma_\pi \omega & \text{if } \gamma_u < \gamma_\pi \\ (1+\gamma_u)\omega - (1+\gamma_u-\gamma_\pi) a_q \mathbb{1}(\omega \geq a_q) & \text{if } \gamma_u \geq \gamma_\pi \end{cases}$$

A Characterization of Inefficient Equilibria

Proposition

Fix an equilibrium data allocation (q^*, x^*) and let ψ_{q^*} be the associated social values of data. The allocation is constrained efficient if and only if

- $q^*(\omega) > 0 \quad \implies \quad \psi_{q^*}(\omega) \geq r(\omega)$
- $q^*(\omega) < \bar{q}(\omega) \quad \implies \quad \psi_{q^*}(\omega) \leq r(\omega)$

A Characterization of Inefficient Equilibria

Proposition

Fix an equilibrium data allocation (q^*, x^*) and let ψ_{q^*} be the associated social values of data. The allocation is constrained efficient if and only if

- $q^*(\omega) > 0 \quad \implies \quad \psi_{q^*}(\omega) \geq r(\omega)$
- $q^*(\omega) < \bar{q}(\omega) \quad \implies \quad \psi_{q^*}(\omega) \leq r(\omega)$

Interpretation: $r(\omega)$ is the marginal social cost; $\psi_q(\omega)$ is the marginal social value

A Characterization of Inefficient Equilibria

Let $\underline{\omega} := \min_{\omega \in \Omega} \omega$. The characterization implies:

Corollary (A Sufficient Condition for Inefficiency)

Let $\gamma_u \geq \gamma_\pi$ and assume constrained efficiency requires $q \neq 0$.

If $\gamma_u \underline{\omega} < r(\underline{\omega}) < (1 + \gamma_u)\underline{\omega}$, all equilibria are inefficient

A Characterization of Inefficient Equilibria

Let $\underline{\omega} := \min_{\omega \in \Omega} \omega$. The characterization implies:

Corollary (A Sufficient Condition for Inefficiency)

Let $\gamma_u \geq \gamma_\pi$ and assume constrained efficiency requires $q \neq 0$.

If $\gamma_u \underline{\omega} < r(\underline{\omega}) < (1 + \gamma_u)\underline{\omega}$, all equilibria are inefficient

Why? Take any allocation $q \neq 0$.

- $\gamma_u \underline{\omega} < r(\underline{\omega}) \implies$ no trade between the platform and type-$\underline{\omega}$ consumers
- $q \neq 0 \implies a_q > \underline{\omega} \implies \psi_q(\underline{\omega}) = (1 + \gamma_u)\underline{\omega}$
- If $r(\underline{\omega}) < (1 + \gamma_u)\underline{\omega} \implies$ social cost $<$ social value
- $\underline{\omega}$-consumers not selling is inefficient

Summary so far

The equilibrium of the competitive data economy can be inefficient

We characterized this inefficiency in two ways:

1. In terms of the social values of data records:

   We provide a tight characterization

2. In terms of the model primitives:

   We find sufficient conditions for efficiency ($\gamma_u < \gamma_\pi$)

   We find sufficient conditions for inefficiency ($\gamma_u > \gamma_\pi$ and more)

Underlying mechanism: platform's incentives $\rightsquigarrow$ induce some pooling $\rightsquigarrow$ enables externalities $\rightsquigarrow$ creates inefficiency remedies

Remedies

How to fix this market failure?

We explore three alternative market designs:

1. Introducing **data taxes**
2. Introducing **data unions**
3. Making data markets more **complete**

data taxes

Data Taxes

Introduce a simple **data tax** on consumers:

- When selling her record, consumer pays tax / receive subsidy $t(\omega) \in \mathbb{R}$

We show that any constrained-efficient allocation can be supported as an equilibrium of the competitive economy with taxation

Data Taxes

Introduce a simple **data tax** on consumers:

- When selling her record, consumer pays tax / receive subsidy $t(\omega) \in \mathbb{R}$

We show that any constrained-efficient allocation can be supported as an equilibrium of the competitive economy with taxation

Data Taxes

Proposition

Fix any constrained-efficient data allocation (q^*, x^*).

There exists an equilibrium of the competitive economy (p^*, ζ^*, q^*, x^*) and taxes

$$t^*(\omega) := \sum_a u(a, \omega) x^*(a|\omega) + p^*(\omega) - \psi_{q^*}(\omega) \qquad \forall \omega$$

supporting the constrained-efficient data allocation.

Data Taxes

Proposition

Fix any constrained-efficient data allocation (q^*, x^*).

There exists an equilibrium of the competitive economy (p^*, ζ^*, q^*, x^*) and taxes

$$t^*(\omega) := \sum_a u(a, \omega) x^*(a|\omega) + p^*(\omega) - \psi_{q^*}(\omega) \qquad \forall \omega$$

supporting the constrained-efficient data allocation. Thus, consumer welfare is maximized, regardless of platform's preferences.

Data Taxes

Proposition

Fix any constrained-efficient data allocation (q^*, x^*).

There exists an equilibrium of the competitive economy (p^*, ζ^*, q^*, x^*) and taxes

$$t^*(\omega) := \sum_a u(a, \omega) x^*(a|\omega) + p^*(\omega) - \psi_{q^*}(\omega) \qquad \forall \omega$$

supporting the constrained-efficient data allocation. Thus, consumer welfare is maximized, regardless of platform's preferences.

Moreover, the budget is balanced.

Data Taxes

Proposition

Fix any constrained-efficient data allocation (q^*, x^*).

There exists an equilibrium of the competitive economy (p^*, ζ^*, q^*, x^*) and taxes

$$t^*(\omega) := \sum_a u(a,\omega)x^*(a|\omega) + p^*(\omega) - \psi_{q^*}(\omega) \qquad \forall\omega$$

supporting the constrained-efficient data allocation. Thus, consumer welfare is maximized, regardless of platform's preferences.

Moreover, the budget is balanced.

Idea: with data tax, consumers internalize social benefit of selling their data

data union

Data Union

We design a **data union** that operates as follows: (Posner, Weyl, 18; Seim et al 22)

Data Union

We design a **data union** that operates as follows: (Posner, Weyl, 18; Seim et al 22)

– Union manages data on behalf of all consumers

i.e., all consumers voluntarily participate in the union, $\zeta(\omega) = 1,\ \forall\omega$.

Data Union

We design a **data union** that operates as follows: (Posner, Weyl, 18; Seim et al 22)

- Union manages data on behalf of all consumers

  i.e., all consumers voluntarily participate in the union, $\zeta(\omega) = 1, \ \forall \omega$.

- Union sells part of its database $\bar{q}$ to platform (price maker)

  i.e., it sells $q \leq \bar{q}$ for $V(q)$ (extracting platform's payoff)

Data Union

We design a **data union** that operates as follows: (Posner, Weyl, 18; Seim et al 22)

- Union manages data on behalf of all consumers

  i.e., all consumers voluntarily participate in the union, $\zeta(\omega) = 1, \; \forall \omega$.

- Union sells part of its database $\bar{q}$ to platform (price maker)

  i.e., it sells $q \leq \bar{q}$ for $V(q)$ (extracting platform's payoff)

- Union distributes proceeds $V(q)$ of the sale back to consumers

  i.e. it chooses p s.t. $\sum_\omega p(\omega)\bar{q}(\omega) = V(q)$ to guarantee participation

Data Union

We design a **data union** that operates as follows: (Posner, Weyl, 18; Seim et al 22)

- Union manages data on behalf of all consumers

  i.e., all consumers voluntarily participate in the union, $\zeta(\omega) = 1, \ \forall \omega$.

- Union sells part of its database $\bar{q}$ to platform (price maker)

  i.e., it sells $q \leq \bar{q}$ for $V(q)$ (extracting platform's payoff)

- Union distributes proceeds $V(q)$ of the sale back to consumers

  i.e. it chooses p s.t. $\sum_{\omega} p(\omega)\bar{q}(\omega) = V(q)$ to guarantee participation

Proposition

Equilibria of the data-union economy are constrained efficient and maximize consumers' welfare, regardless of platform's incentives

more-complete markets

Lindahl Economy

We allow consumers to trade **the way** their records are used by platform

More-complete markets:

- There is a market where type-ω records can be sold for "intended use a"
- The price of ω-records, $p(a, \omega)$, can now depend on how it is used

Lindahl Economy

We allow consumers to trade **the way** their records are used by platform

More-complete markets:

- There is a market where type-ω records can be sold for "intended use a"
- The price of ω-records, $p(a, \omega)$, can now depend on how it is used

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

Lindahl Economy

We allow consumers to trade **the way** their records are used by platform

More-complete markets:

- There is a market where type-ω records can be sold for "intended use a"
- The price of ω-records, $p(a, \omega)$, can now depend on how it is used

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

Reminiscent of EU's GDPR: "*The **specific purposes** for which personal data are used should be determined at the time of the collection*"

Lindahl Economy

We allow consumers to trade **the way** their records are used by platform

More-complete markets:

- There is a market where type-ω records can be sold for "intended use a"
- The price of ω-records, $p(a, \omega)$, can now depend on how it is used

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

Reminiscent of EU's GDPR: "*The* ***specific purposes*** *for which personal data are used should be determined at the time of the collection*"

Proposition

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

monopsonist platform

Monopsonist Economy

Drop competitive market assumption and suppose platform is **price maker**

- Platform sets data prices p by making take-it-or-leave-it offers to consumers

Monopsonist Economy

Drop competitive market assumption and suppose platform is **price maker**

– Platform sets data prices p by making take-it-or-leave-it offers to consumers

Definition

(p^*, ζ^*, q^*, x^*) is an equilibrium of the **monopsonist economy** if it solves

$$\begin{aligned} \max_{(p,\zeta,q,x)} \quad & V(q) - \sum_{\omega} p(\omega) q(\omega) \\ \text{s.t.} \quad & \text{conditions } (b), (c), (d) \text{ satisfied} \end{aligned}$$

Monopsonist Economy

Drop competitive market assumption and suppose platform is **price maker**

- Platform sets data prices p by making take-it-or-leave-it offers to consumers

Definition

(p^*, ζ^*, q^*, x^*) is an equilibrium of the **monopsonist economy** if it solves

$$\begin{aligned} \max_{(p,\zeta,q,x)} \quad & V(q) - \sum_{\omega} p(\omega) q(\omega) \\ \text{s.t.} \quad & \text{conditions } (b), (c), (d) \text{ satisfied} \end{aligned}$$

Remark

In a monopsony equilibrium, allocations are constrained-efficient (and vice versa). Moreover, platform's payoff is maximal, while consumers' welfare is minimal

conclusion

Summary

1. A stylized framework to study competitive markets for personal data

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

2. Emphasize a novel market failure

   Platform's role as an information intermediary enables an externality that leads to inefficiencies

3. We propose three alternative market designs that fix inefficiency: data taxes, data unions, more-complete data markets