Competitive Markets for Personal Data

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

March 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:

- "Expropriation" and barter, common practice in the industry (FTC '15)
- This may lead to inefficiencies and inequality (Seim 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?

A Competitive Data Market

A stylized model of a **competitive data market**:

- Each **consumer** owns her data and can sell it to a platform
- The **platform** pays her a market price for the data and offers her a "service"
- It gives consumer access to a **merchant**, from whom she can buy a product, and
- Platform uses acquired database to inform merchant about consumer's type

A Competitive Data Market

A stylized model of a **competitive data market**:

- Each **consumer** owns her data and can sell it to a platform
- The **platform** pays her a market price for the data and offers her a "service"
- It gives consumer access to a **merchant**, from whom she can buy a product, and
- Platform uses acquired database to inform merchant about consumer's type

A canonical info-design problem but platform's database is **endogenous**, pinned down in the equilibrium of the data market

Main Results

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

    Consumers exert an externality on each other, enabled by how the platform endogenously uses the data

    If platform's objective is suff aligned with merchant's, data market is efficient

    If platform's objective is suff aligned with consumers', data market can be inefficient (and, in some case, unravels entirely)

Main Results

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

    Consumers exert an externality on each other, enabled by how the platform endogenously uses the data

    If platform's objective is suff aligned with merchant's, data market is efficient

    If platform's objective is suff aligned with consumers', data market can be inefficient (and, in some case, unravels entirely)

2. We then propose three solutions to this market failure:

    Introducing a **data union**; Implementing **data taxes**; and making data markets "more complete"

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); How data is priced (this paper)

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); How data is priced (this paper)

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
- Idea builds on GLP '23, who characterized how much platform values the data of a single consumer in a larger database

model

A Stylized Data Economy

One merchant, one platform, a unit mass of consumers

A Stylized Data Economy

One merchant, one 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 Stylized Data Economy

One merchant, one 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 Stylized Data Economy

One merchant, one 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

Denote by $\zeta(\omega) \in [0, 1]$ the 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

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

– If a type-ω consumer sells her record, she obtains reservation utility $r(\omega)$

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 $\omega \geq a$

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.

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

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

(canonical 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 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.,

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

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

(c). Given p^* and x^*, 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). Data markets clear, i.e. $\quad q^*(\omega) = \zeta^*(\omega)\bar{q}(\omega) \qquad \forall \omega$

Discussion of Main Assumptions

Discussion of Main Assumptions

Single platform takes data prices as given:

- Substantive: price-taking behavior, i.e. competitiveness of the market
- Expositional: single platform richer economy studied in GP '22

Discussion of Main Assumptions

Single platform takes data prices as given:

- Substantive: price-taking behavior, i.e. competitiveness of the market
- Expositional: single platform — richer economy studied in GP '22

Platform's payoff $v(a, \omega)$ is linear combination of $\pi(a, \omega)$ and $u(a, \omega)$:

- Tractability — a dynamic microfoundation in XY '23

Discussion of Main Assumptions

Single platform takes data prices as given:

- Substantive: price-taking behavior, i.e. competitiveness of the market
- Expositional: single platform — richer economy studied in GP '22

Platform's payoff $v(a, \omega)$ is linear combination of $\pi(a, \omega)$ and $u(a, \omega)$:

- Tractability — a dynamic microfoundation in XY '23

Three aspects of the consumer problem have been simplified:

- Record fully reveals underlying type — alt see GLP '23
- Record bundles access and information — alt see ALV '22
- Reservation utility $r(\omega)$ is exogenous — alt see BB '23

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?

Call (q, x) an **allocation**

Efficiency Benchmark

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

Call (q, x) an **allocation**, and denote the **welfare** of consumers and platform by

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

Efficiency Benchmark

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

Call (q, x) an **allocation**, and denote the **welfare** of consumers and platform by

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

Definition

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

$$\begin{aligned} W^\circ = \quad & \max_{q,x} \quad \mathcal{W}(q,x) \\ & \text{s.t.} \quad q \le \bar{q} \ \text{ and } \ x \text{ solves platform' problem } \mathcal{P}_q \end{aligned}$$

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

inefficiency of the data economy

Roadmap for the Analysis

Goal: Identify necessary and sufficient conditions for eqm efficiency

1. Characterize constrained efficiency of equilibrium allocations
2. Identify an externality that can lead to market failure
3. State our main result and its intuition

“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 $r(\omega) \geq 0$

"Social" Cost and Benefit of Data Records

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

- It's $r(\omega) \geq 0$

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 $r(\omega) \geq 0$

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

- Fix any q. Consider the following maximization problem:

$$\begin{aligned} W(q) \triangleq \quad & \max_{x:\Omega\to\Delta(A)} \quad \sum_{a,\omega}(v(a,\omega)+u(a,\omega))x(a|\omega)q(\omega) \\ & \text{s.t.} \qquad x \text{ solves } \mathcal{P}_q \end{aligned}$$

“Social” Cost and Benefit of Data Records

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

– It's $r(\omega) \geq 0$

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

– Fix any q. Consider the following maximization problem:

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

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

"Social" Cost and Benefit of Data Records

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

– It's $r(\omega) \geq 0$

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

– Fix any q. Consider the following maximization problem:

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

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

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

Simple cost-benefit analysis: Necessity is obvious, sufficiency less so, but crucial for what comes next

“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

$$U^*(\omega) \triangleq p^*(\omega) + \sum_a x^*(a, \omega) u(a, \omega)$$

"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

$$U^*(\omega) \triangleq p^*(\omega) + \sum_a x^*(a,\omega)u(a,\omega)$$

In equilibrium, the optimality of consumer behavior requires that:

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

Alignment of Social vs Private Benefits

Thus, an equilibrium is constrained-efficient if and only if the social (ψ_{q^*}) and private (U^*) benefit of data records are sufficiently **aligned**

– i.e., $U^*(\omega) \geq r(\omega) \Rightarrow \psi_{q^*}(\omega) \geq r(\omega)$ and $U^*(\omega) \leq r(\omega) \Rightarrow \psi_{q^*}(\omega) \leq r(\omega)$

Alignment of Social vs Private Benefits

Thus, an equilibrium is constrained-efficient if and only if the social (ψ_{q^*}) and private (U^*) benefit of data records are sufficiently **aligned**

- i.e., $U^*(\omega) \geq r(\omega) \Rightarrow \psi_{q^*}(\omega) \geq r(\omega)$ and $U^*(\omega) \leq r(\omega) \Rightarrow \psi_{q^*}(\omega) \leq r(\omega)$

Thus, key question is: When are ψ_{q^*} and U^* aligned?

A Data Externality

Recall definition of **private benefit** of selling ω-record:

$$U^*(\omega) \triangleq p^*(\omega) + \sum_a x^*(a,\omega)u(a,\omega)$$

A Data Externality

Recall definition of **private benefit** of selling ω-record:

$$U^*(\omega) \triangleq p^*(\omega) + \underbrace{\textstyle\sum_a x^*(a,\omega)u(a,\omega)}_{\substack{\text{change in trading surplus} \\ \text{for } \textbf{this consumer}}}$$

A Data Externality

Recall definition of **private benefit** of selling ω-record:

$$U^*(\omega) \triangleq p^*(\omega) + \underbrace{\textstyle\sum_a x^*(a,\omega)u(a,\omega)}_{\substack{\text{change in trading surplus} \\ \text{for \textbf{this consumer}}}}$$

In eqm, the **social benefit** of selling ω-record can be decomposed as:

$$\psi_{q^*}(\omega) = \underbrace{p^*(\omega)}_{\substack{\text{change in platform's} \\ \text{gross payoff}}} + \underbrace{\xi^*(\omega)}_{\substack{\text{change in trading} \\ \text{surplus of \textbf{all consumers}}}}$$

A Data Externality

Recall definition of **private benefit** of selling ω-record:

$$U^*(\omega) \triangleq p^*(\omega) + \underbrace{\textstyle\sum_a x^*(a,\omega)u(a,\omega)}_{\substack{\text{change in trading surplus} \\ \text{for } \textbf{this consumer}}}$$

In eqm, the **social benefit** of selling ω-record can be decomposed as:

$$\psi_{q^*}(\omega) = \underbrace{p^*(\omega)}_{\substack{\text{change in platform's} \\ \text{gross payoff}}} + \underbrace{\xi^*(\omega)}_{\substack{\text{change in trading} \\ \text{surplus of } \textbf{all consumers}}}$$

Discrepancy between $\sum_a x^*(a,\omega)u(a,\omega)$ and $\xi^*(\omega)$ captures an **externality** that type-ω consumer exerts on others when selling her record to the platform

Main Characterization

This externality, and thus the eqm efficiency, depends on how platform uses the data, i.e., on its objective

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

Main Characterization

This externality, and thus the eqm efficiency, depends on how platform uses the data, i.e., 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 thus consumers' welfare is maximal
- If $\gamma_u \geq \gamma_\pi$, equilibria can be inefficient (and consumers' welfare can be minimal, i.e., $\sum_\omega r(\omega)\bar{q}(\omega)$)

Main Characterization

This externality, and thus the eqm efficiency, depends on how platform uses the data, i.e., on its objective

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

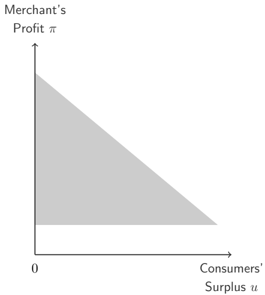
Proposition

- If $\gamma_u < \gamma_\pi$, equilibria are constrained efficient and thus consumers' welfare is maximal
- If $\gamma_u \geq \gamma_\pi$, equilibria can be inefficient (and consumers' welfare can be minimal, i.e., $\sum_\omega r(\omega)\bar{q}(\omega)$)

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

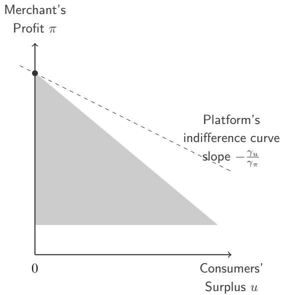
Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



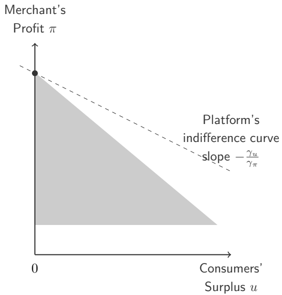
Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



Intuition for Efficiency

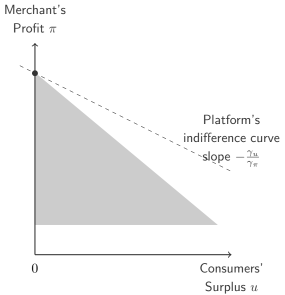
If $\gamma_u < \gamma_\pi$



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

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



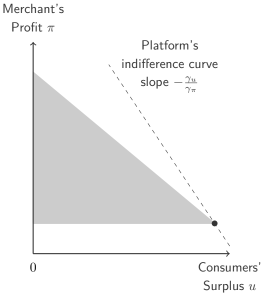
- At all q, **full disclosure** is optimal
- Merchant extracts surplus from all consumers
- Therefore, $\xi^*(\omega) = \sum_a x^*(a,\omega)u(a,\omega) = 0$
- Therefore, $\psi_q^* = U^*$, perfect alignement
- Therefore, all equilibria are constrained efficient

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$

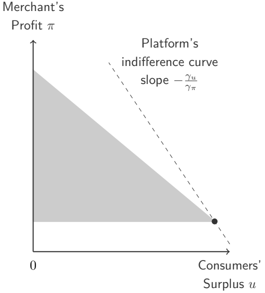
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



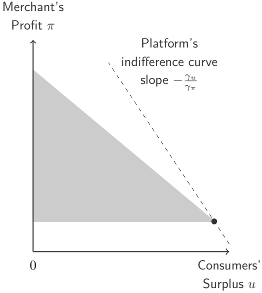
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



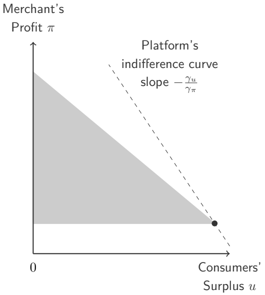
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



Intuition for Inefficiency

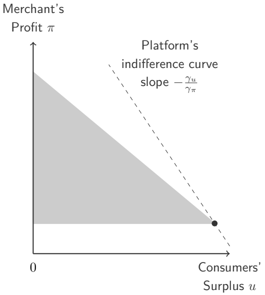
If $\gamma_u > \gamma_\pi$



– Platform **withholds information** from merchant

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



- Platform **withholds information** from merchant
- Pooling different consumers together makes their payoff inter-dependent
- Thus, $\xi^*(\omega) \neq \sum_a x^*(a,\omega)u(a,\omega)$
- Example: think of lowest-type consumer

A Stronger Negative Result

We can sharpen the negative part of the previous result:

To avoid trivial cases, focus on economies where the constrained efficient allocation requires some trade, i.e., $W^\circ > \sum_\omega \bar{q}(\omega) r(\omega)$

Corollary

Let $\gamma_\pi \leq \gamma_u$. Additionally, suppose $\gamma_u \underline{\omega} < r(\underline{\omega}) < (1 + \gamma_u)\underline{\omega}$.
Then, **all equilibria** are inefficient.

Stepping Back

In many digital industries, information intermediaries play a ubiquitous role
Acquisti et al. '16

A defining feature of these intermediaries is that they balance interests of conflicting agents (sellers-buyers; drivers-riders, etc.)

Typically, in theory and in practice, intermediaries manage these conflicts by optimally withholding some information from the agents

This paper illustrates how this practice can create market failures

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 $\bar{q}(1) < \bar{q}(2)$
- Same reservation utility: $r(\omega) = \bar{r} \in (0, \frac{1+\gamma_u}{2})$, for all ω

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 $\bar{q}(1) < \bar{q}(2)$
- Same reservation utility: $r(\omega) = \bar{r} \in (0, \frac{1+\gamma_u}{2})$, for all ω

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 $\bar{q}(1) < \bar{q}(2)$
- Same reservation utility: $r(\omega) = \bar{r} \in (0, \frac{1+\gamma_u}{2})$, for all ω

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)

We can show 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)

We can show 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)

We can show 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

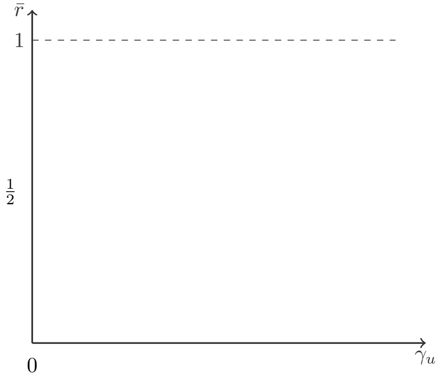
- Hence, high-type consumer do not want to sell either

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

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

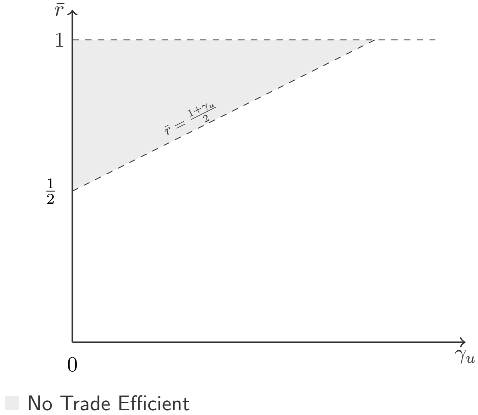
Example

Complete equilibrium characterization for this example:



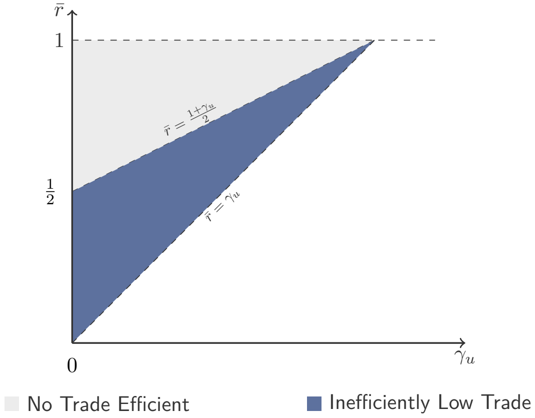
Example

Complete equilibrium characterization for this example:



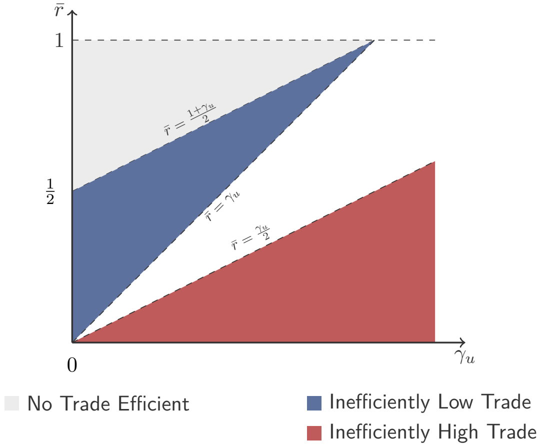
Example

Complete equilibrium characterization for this example:



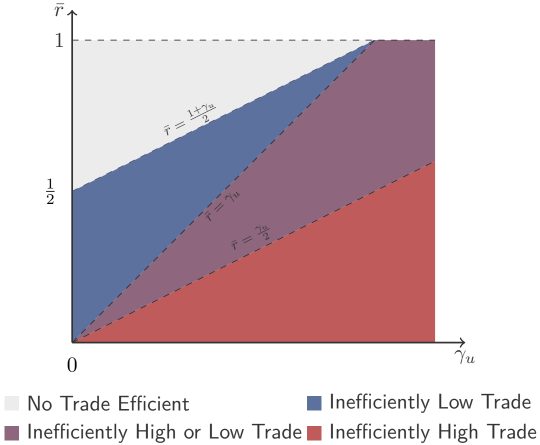
Example

Complete equilibrium characterization for this example:



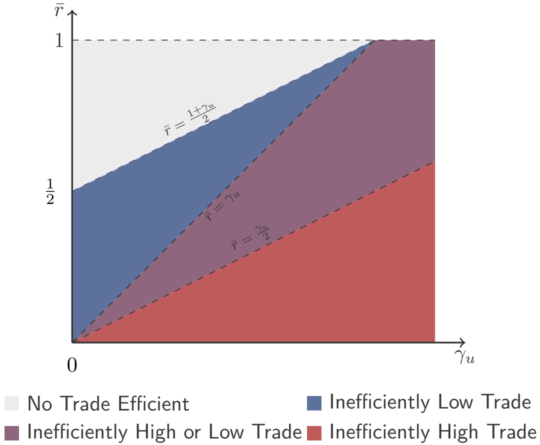
Example

Complete equilibrium characterization for this example:



Example

Complete equilibrium characterization for this example:



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 Union

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

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

First model of a data union, we offer theoretical support to these policy proposals

Data Union

How does a data union work?

- Consumers choose whether to become members of 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:

$$\begin{array}{rl}
\max\limits_{(p,q,x)} & \displaystyle\sum_{\omega} p(\omega)\bar{q}(\omega) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega}(\bar{q}(\omega) - q(\omega))r(\omega) \\
\text{such that} & q \leq \bar{q}, \\
\text{and} & \sum_{\omega} p(\omega)\bar{q}(\omega) = V(q), \\
\text{and} & x \text{ solves } \mathcal{P}_q, \\
\text{and} & p(\omega) + \frac{q(\omega)}{\bar{q}(\omega)} \sum_a u(a,\omega)x(a|\omega) + \big(1 - \frac{q(\omega)}{\bar{q}(\omega)}\big)r(\omega) \geq r(\omega).
\end{array}$$

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

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

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)

Thus, the platform and the consumer must 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)

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

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

It 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)

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

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

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

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, same timing and assumptions on commitment power

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, same timing and assumptions on commitment power

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

March 2024