Competitive Markets for Personal Data

Simone Galperti
UCSD

Tianhao Liu
Columbia

Jacopo Perego
Columbia

May 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 distortions (Seim et al. 2023)

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: ("GE meets ID")

- 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

2. Propose three solutions to this market failure:

- Data unions, Data taxes, Lindahl prices for data

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** $q = (q(\omega))_{\omega \in \Omega}$ and 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** $q = (q(\omega))_{\omega \in \Omega}$ and 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

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** $q = (q(\omega))_{\omega \in \Omega}$ and pays $\sum_\omega q(\omega)p(\omega)$

The supply side:

- If a type-ω consumer sells her record to the platform, she is paid $p(\omega)$ and is later intermediated with merchant
- If type-ω consumer doesn't sell her record, she gets 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.

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

(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.,

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

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. $\quad q^*(\omega) = \zeta^*(\omega)\bar{q}(\omega) \qquad \forall \omega$

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

(in)efficiency of the data economy

Efficiency Benchmark

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

Efficiency Benchmark

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

Call (q, x) an allocation and denote by $\mathcal{W}(q, x)$ the **welfare** of consumers and platform

Efficiency Benchmark

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

Call (q, x) an allocation and denote by $\mathcal{W}(q, x)$ the **welfare** of consumers and platform

Definition

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

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

Efficiency Benchmark

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

Call (q, x) an allocation and denote by $\mathcal{W}(q, x)$ the **welfare** of consumers and platform

Definition

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

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

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

Private and Social Benefit of Selling a Record

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

The **private benefit** for a type-ω consumer from selling her record is:

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

Private and Social Benefit of Selling a Record

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

The **private benefit** for a type-ω consumer from selling her record is:

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

Private and Social Benefit of Selling a Record

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

The **private benefit** for a type-ω consumer from selling her record is:

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

The **social benefit** when a type-ω consumer sells her record is:

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

Private and Social Benefit of Selling a Record

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

The **private benefit** for a type-ω consumer from selling her record is:

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

The **social benefit** when a type-ω consumer sells her record is:

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

Result (informal): If U^* and ψ_q^* sufficiently aligned, eqm is constrained efficient

Main Characterization

When are U^* and ψ_q^* aligned? It depends on how platform uses the data and, thus, on its objective

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

Main Characterization

When are U^* and ψ_q^* aligned? It depends on how platform uses the data and, thus, 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 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

When are U^* and ψ_q^* aligned? It depends on how platform uses the data and, thus, on its objective

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

Proposition

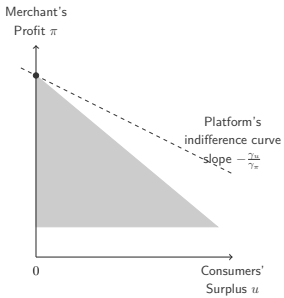
- If $\gamma_u < \gamma_\pi$, equilibria are constrained efficient and consumers' welfare is maximal
- If $\gamma_u \geq \gamma_\pi$, equilibria can be inefficient (and consumers' welfare can be minimal, i.e., $\sum_\omega r(\omega)\bar{q}(\omega)$)

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$

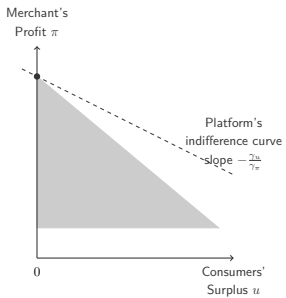
Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



Intuition for Efficiency

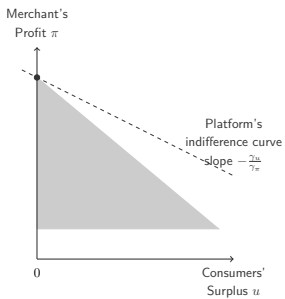
If $\gamma_u < \gamma_\pi$



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

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



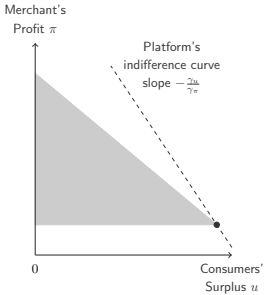
- At all q, **full disclosure** is optimal
- Merchant extracts surplus from all consumers
- Therefore, $\xi^*(\omega) = \sum_a x^*(a,\omega)u(a,\omega) = 0$
- Therefore, $\psi_q^* = U^*$, perfect alignement
- **Result:** All equilibria are constrained efficient

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$

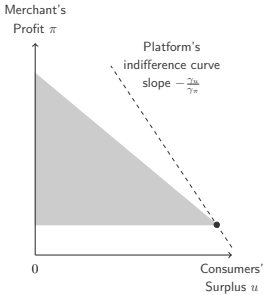
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



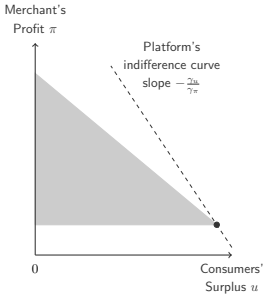
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



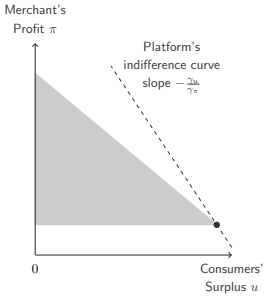
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



Intuition for Inefficiency

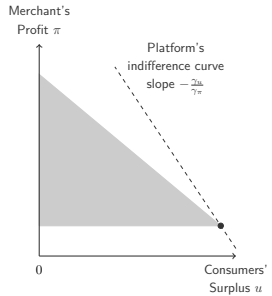
If $\gamma_u > \gamma_\pi$



– 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
- Externalities $\rightsquigarrow$ Eqm inefficiency

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 and when this practice can lead to market failure

Inefficiency we emphasize is more general than our price-discrimination application with a monopolist merchant

remedies

Remedies

We study three solutions to this market failure:

1. Introducing **data unions**

   A new intermediary manages consumer data on their behalf and coordinates the data sale

Remedies

We study three solutions to this market failure:

1. Introducing **data unions**

   A new intermediary manages consumer data on their behalf and coordinates the data sale

2. Implementing **data taxes**

   A tax $\tau(\omega) \in \mathbb{R}$ is levied by govt when a type-ω record is sold

Remedies

We study three solutions to this market failure:

1. Introducing **data unions**

   A new intermediary manages consumer data on their behalf and coordinates the data sale

2. Implementing **data taxes**

   A tax $\tau(\omega) \in \mathbb{R}$ is levied by govt when a type-ω record is sold

3. Lindahl Pricing

   Richer data prices $p(\omega, a)$ that also depend on "intended use"

Remedies

We study three solutions to this market failure:

1. Introducing **data unions**

   A new intermediary manages consumer data on their behalf and coordinates the data sale

2. Implementing **data taxes**

   A tax $\tau(\omega) \in \mathbb{R}$ is levied by govt when a type-ω record is sold

3. Lindahl Pricing

   Richer data prices $p(\omega, a)$ that also depend on "intended use"

   "*The **specific purpose** for which personal data is used should be determined at the time of the collection*" (GDPR)

related work

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 — 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); 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 — 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

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

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

back