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 Contributions

1. Highlight novel **market failure**, despite competition and property rights

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 Contributions

1. Highlight novel **market failure**, despite competition and property rights

- Stems from **pooling externalities**, which are enabled by how platform endogenously uses the 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 Contributions

1. Highlight novel **market failure**, despite competition and property rights

- Stems from **pooling externalities**, which are enabled by how platform endogenously uses the data — Galperti, Levkun, Perego (2023)
- Unrelated to "learning" externalities — Choi et al ('19), Ichihashi ('21), Bergemann et al. ('22), Acemoglu et al. ('22)

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 Contributions

1. Highlight novel **market failure**, despite competition and property rights

- Stems from **pooling externalities**, which are enabled by how platform endogenously uses the data — Galperti, Levkun, Perego (2023)
- Unrelated to "learning" externalities — Choi et al ('19), Ichihashi ('21) Bergemann et al. ('22), Acemoglu et al. ('22)

2. Propose three alternative **solutions**

- Data taxes; a data union; an making data markets more complete

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

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 Stylized 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 allows platform to intermediate her with merchant

A Stylized 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 allows platform to intermediate her with merchant

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

Period 1: Data Trade

Period 1: Data Trade

Platform and consumers trade the data records in a **competitive market**

Period 1: Data Trade

Platform and consumers trade the data records in a **competitive market**

Demand Side:

Platform demands database $q = (q(\omega_1), \ldots, q(\omega_N))$ and pays unit price $p(\omega)$ for each ω-record

Supply Side:

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

If consumer does not sell, she forgoes price and platform's "service" and obtains 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:	$u(a, \omega) =$ surplus
Merchant's:	$\pi(a, \omega) =$ profits
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

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

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

analysis

Efficiency Benchmark

Do equilibria "efficiently" allocate data records between consumers and platform?

Efficiency Benchmark

Do equilibria "efficiently" allocate data records between consumers and platform?

Call (q, x) an **outcome** and let $W(q, x)$ be welfare of consumers and platform:

Efficiency Benchmark

Do equilibria "efficiently" allocate data records between consumers and platform?

Call (q, x) an **outcome** and let $W(q, x)$ be welfare of consumers and platform:

Definition

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

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

Main Result

Question: Are equilibrium outcomes constrained efficient?

Eqm efficiency depends on how platform uses data, thus on its objective

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

Main Result

Question: Are equilibrium outcomes constrained efficient?

Eqm efficiency depends on how platform uses data, 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 thus consumers' welfare is maximal
- If $\gamma_u \geq \gamma_\pi$, equilibria can be inefficient (and consumers' welfare can be as low as $R = \sum_\omega r(\omega)\bar{q}(\omega)$)

Main Result

Question: Are equilibrium outcomes constrained efficient?

Eqm efficiency depends on how platform uses data, 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 thus consumers' welfare is maximal
- 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?**

Intuition for Inefficiency

- When $\gamma_u \geq \gamma_\pi$, platform withholds some information from merchant
- How? Platform pools consumers of different types
- Composition of a pool determines merchant's belief, thus his fee
- If a consumer does not sell her data, she affects pool composition and, thus, other consumers' payoff
- Consumers exert **externality** on each other (Relation to GLP '23)

Intuition for Inefficiency

- When $\gamma_u \geq \gamma_\pi$, platform withholds some information from merchant
- How? Platform pools consumers of different types
- Composition of a pool determines merchant's belief, thus his fee
- If a consumer does not sell her data, she affects pool composition and, thus, other consumers' payoff
- Consumers exert **externality** on each other (Relation to GLP '23)

In the Paper

See paper for:

- Existence
- General characterization of eqm efficiency
- Stronger sufficiency conditions for inefficiency
- Social welfare
- An illustrative example

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 outcome can be supported by 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 outcome can be supported by an equilibrium of the competitive economy with taxation

Data Taxes

Proposition

Fix any constrained-efficient outcome (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 outcome. 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, outcomes 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