Competitive Markets for Personal Data

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

August 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

Main Result.

- This economy can be **inefficient**, despite its competitive nature 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. Two alternative market designs to avoid inefficiency:

- Introducing a **data union**
- Making data markets **more complete**

Literature

This paper contributes to a growing literature on data markets (Bonatti and Bergemann, 2019)

Our model is rooted in GE tradition, but leverages recent progress in information-design literature (Bergemann and Morris (2019)

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)

Focus competitive markets and emphasize a novel source of inefficiency Galperti et al. '23

- It hinges on unique role modern platforms have as info intermediaries

model

A Simple Data Economy

The players: one merchant, one platform, a unit mass of consumers

A Simple Data Economy

The players: 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

The players: 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 (record bundles access + info)

A Simple Data Economy

The players: 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 (record bundles access + info)

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

Period 1: Data Trade

Platform and consumers trade the data records

Period 1: Data Trade

Platform and consumers trade the data records

Suppose unit **prices** of data records are $p = (p(\omega))_\omega \in \mathbb{R}^\Omega$

(supply) ω-consumer sells her record with prob $\zeta(\omega)$ and is paid $p(\omega)$

(demand) Platform buys quantity $q(\omega)$ of ω-records at unit price $p(\omega)$

Period 1: Data Trade

Platform and consumers trade the data records

Suppose unit **prices** of data records are $p = (p(\omega))_\omega \in \mathbb{R}^\Omega$

(supply) ω-consumer sells her record with prob $\zeta(\omega)$ and is paid $p(\omega)$

(demand) Platform buys quantity $q(\omega)$ of ω-records at unit price $p(\omega)$

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

As if platform chose recommendation mechanism $x : \Omega \to \Delta(A)$ to solve

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

(ID problem with endogenous q)

Summary

Four endogenous variables: $\Big(p, \zeta, q, x\Big)$

Summary

Four endogenous variables: $\Big(p, \zeta, q, x\Big)$

Definition

A profile (p^*, ζ^*, q^*, x^*) is **consistent** if

- Given p^* and x^*, consumers' behave optimally, i.e.

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

- Given q^*, mechanism x^* solves the platform's problem $\mathcal{P}_{q^*}$

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

Summary

Four endogenous variables: $\Big(p, \zeta, q, x\Big)$

Definition

A profile (p^*, ζ^*, q^*, x^*) is **consistent** if

- Given p^* and x^*, consumers' behave optimally, i.e.

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

- Given q^*, mechanism x^* solves the platform's problem $\mathcal{P}_{q^*}$
- Markets clear, i.e. $q^*(\omega) = \zeta^*(\omega)\bar{q}(\omega) \qquad \forall\omega$

How are data prices p determined?

Summary

Four endogenous variables: $\Big(p, \zeta, q, x\Big)$

Definition

A profile (p^*, ζ^*, q^*, x^*) is **consistent** if

- Given p^* and x^*, consumers' behave optimally, i.e.

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

- Given q^*, mechanism x^* solves the platform's problem $\mathcal{P}_{q^*}$
- Markets clear, i.e. $q^*(\omega) = \zeta^*(\omega)\bar{q}(\omega) \qquad \forall\omega$

How are data prices p determined?

monopsonist economy

Monopsonist Economy

Let's start with a simple benchmark: the platform is a **price maker**

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

Monopsonist Economy

Let's start with a simple benchmark: the platform is a **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

$$\begin{aligned} \max_{(p,\zeta,q,x)} \quad & V(q) - \sum_{\omega} p(\omega) q(\omega) \\ \text{s.t.} \quad & (p, \zeta, q, x) \text{ is consistent} \end{aligned}$$

Monopsonist Economy

Let's start with a simple benchmark: the platform is a **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

$$\begin{aligned} \max_{(p,\zeta,q,x)} \quad & V(q) - \sum_{\omega} p(\omega) q(\omega) \\ \text{s.t.} \quad & (p, \zeta, q, x) \text{ is consistent} \end{aligned}$$

Result: Equilibria maximize the **sum of platform's and consumers' payoffs**

Constrained Efficiency

Definition

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

$$\max_{q,x} \quad V(q) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega} \Big(\bar{q}(\omega) - q(\omega)\Big) r(\omega)$$

$$\text{s.t.} \quad q \leq \bar{q} \text{ and } x \text{ solves } \mathcal{P}_q.$$

"Constrained" as x° needs to be platform-optimal given q°

Constrained Efficiency

Definition

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

$$\max_{q,x} \quad V(q) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega} \Big(\bar{q}(\omega) - q(\omega)\Big) r(\omega)$$

$$\text{s.t.} \quad q \leq \bar{q} \text{ and } x \text{ solves } \mathcal{P}_q.$$

"Constrained" as x° needs to be platform-optimal given q°

Remark

In a monopsony equilibrium:

- Allocations are constrained-efficient (and vice versa)
- Platform's payoff is maximal, while consumers' welfare is minimal

Constrained Efficiency

Definition

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

$$\begin{aligned} \max_{q,x} \quad & V(q) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega} \Big(\bar{q}(\omega) - q(\omega)\Big) r(\omega) \\ \text{s.t.} \quad & q \leq \bar{q} \text{ and } x \text{ solves } \mathcal{P}_q. \end{aligned}$$

"Constrained" as x° needs to be platform-optimal given q°

Remark

In a monopsony equilibrium:

- Allocations are constrained-efficient (and vice versa)
- Platform's payoff is maximal, while consumers' welfare is minimal

Efficient equilibria yet consumers don't reap the benefits of their data
$\rightsquigarrow$ Can a competitive economy do better?

competitive economy

Competitive Economy

Take bargaining power away from platform: assume platform is **price taker**

Prices determined by market clearing, as in a competitive market with multiple identical platforms

Competitive Economy

Take bargaining power away from platform: assume platform is **price taker**

Prices determined by market clearing, as in a competitive market with multiple identical platforms

Definition

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

- Given p^*, database q^* solves platform's problem:

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

- (p^*, ζ^*, q^*, x^*) is consistent

Competitive Economy

Take bargaining power away from platform: assume platform is **price taker**

Prices determined by market clearing, as in a competitive market with multiple identical platforms

Definition

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

- Given p^*, database q^* solves platform's problem:

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

- (p^*, ζ^*, q^*, x^*) is consistent

Are equilibria still constrained efficient? Are consumers better off?

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 in the monopsonist economy

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

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

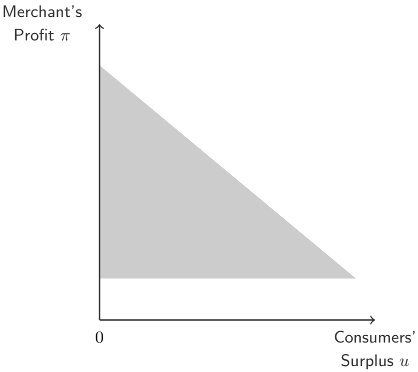
Proposition

- 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 in the monopsonist economy

When platform "cares too much" about consumers, competitive equilibria can be inefficient $\rightsquigarrow$ **Why?**

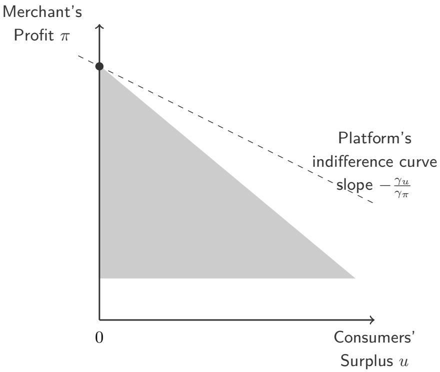
Intuition for Inefficiency

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



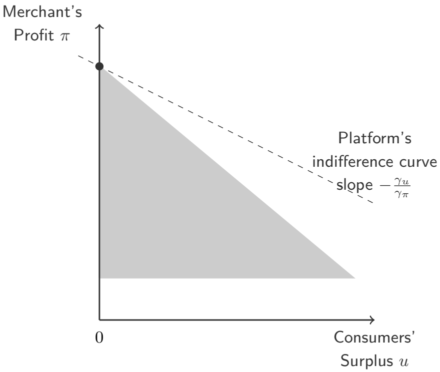
Intuition for Inefficiency

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



Intuition for Inefficiency

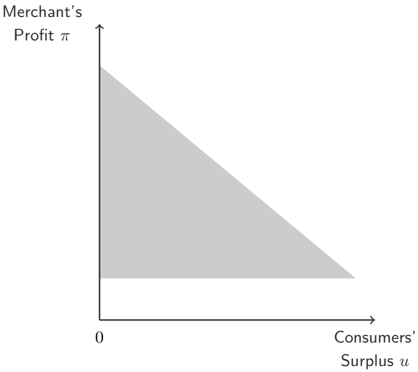
Recall "surplus triangle" from Bergemann et al. (2015):



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

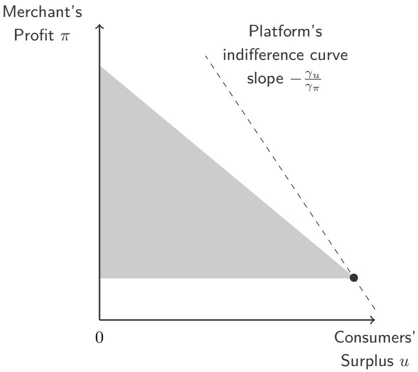
Intuition for Inefficiency

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



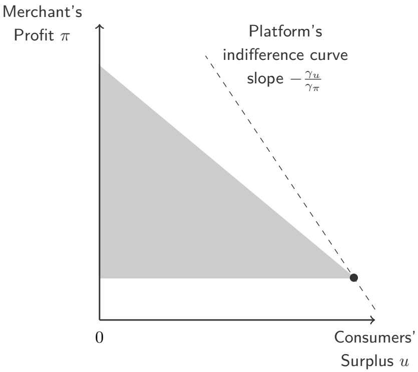
Intuition for Inefficiency

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

example

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 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

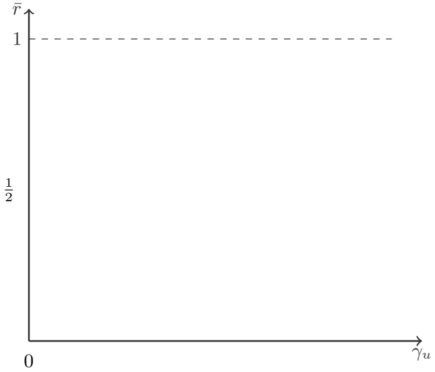
The constrained-efficient allocation involves:

- 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: $x^\circ(1|\omega) = 1, \ \forall\omega$

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

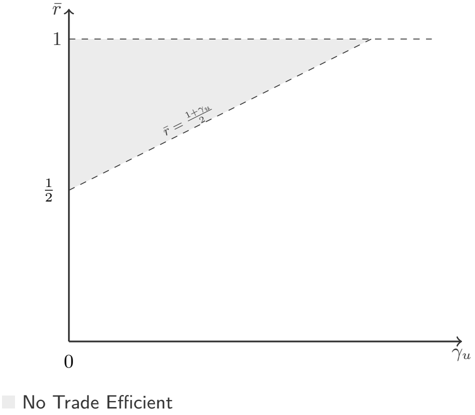
- Low-type consumers do not sell their records $\rightsquigarrow$ neg externality
- Hence, high-type consumer do not want to sell
- Market unravels $\rightsquigarrow$ No trade $\rightsquigarrow$ Inefficiency

Complete equilibrium characterization for this example:

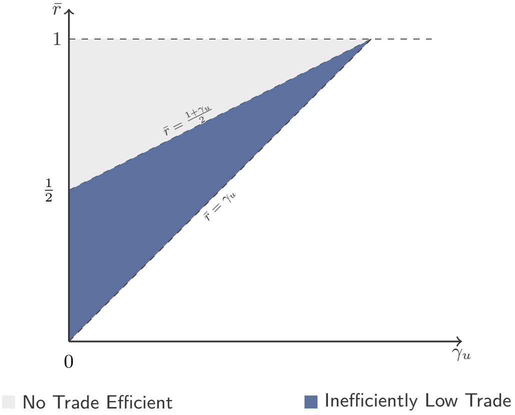


Example

Complete equilibrium characterization for this example:

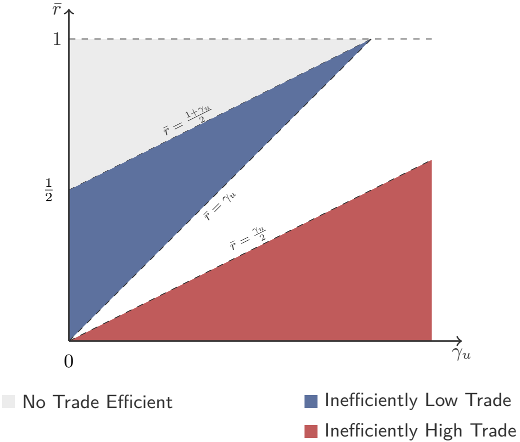


Complete equilibrium characterization for this example:



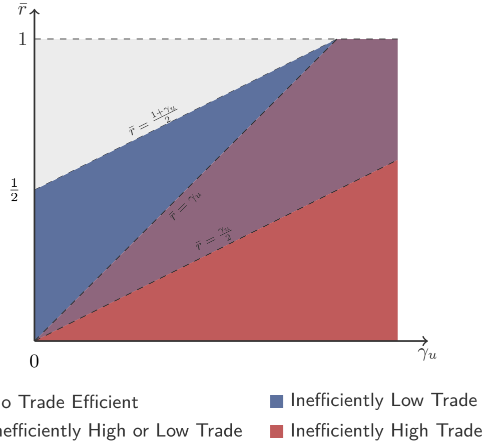
Example

Complete equilibrium characterization for this example:



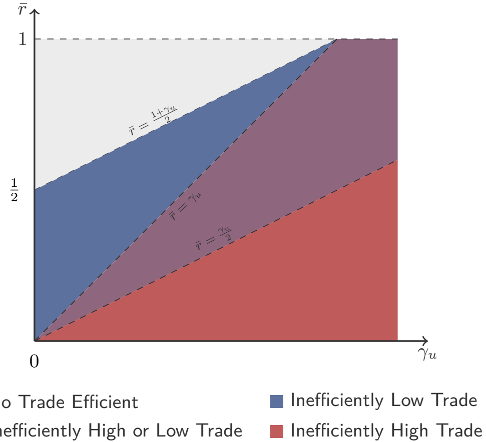
Example

Complete equilibrium characterization for this example:



Example

Complete equilibrium characterization for this example:



Takeaway: Whenever some trade is required for efficiency, the equilibrium allocation is inefficient

remedies

Remedies

We propose two solutions to avoid this market inefficiency:

1. Introducing a **data union**
2. Making data markets **more complete**

Data Union

Our **data union** operates as follows: (blueprint from Posner Weyl, 2018)

Data Union

Our **data union** operates as follows: (blueprint from Posner Weyl, 2018)

- Union manages data on behalf of all consumers

  i.e., it has database $\bar{q}$; all consumers participate $\zeta(\omega) = 1,\ \forall \omega$

Data Union

Our **data union** operates as follows: (blueprint from Posner Weyl, 2018)

- Union manages data on behalf of all consumers

  i.e., it has database $\bar{q}$; all consumers participate $\zeta(\omega) = 1, \ \forall\omega$

- Union sells part of it to platform and extract its payoff

  i.e., it chooses $q \leq \bar{q}$ and sells it for to $V(q)$

Data Union

Our **data union** operates as follows: (blueprint from Posner Weyl, 2018)

- Union manages data on behalf of all consumers

  i.e., it has database $\bar{q}$; all consumers participate $\zeta(\omega) = 1, \ \forall\omega$

- Union sells part of it to platform and extract its payoff

  i.e., it chooses $q \leq \bar{q}$ and sells it for to $V(q)$

- Union distributes proceeds to make consumer participation IC

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

Data Union

Our **data union** operates as follows: (blueprint from Posner Weyl, 2018)

- Union manages data on behalf of all consumers

  i.e., it has database $\bar{q}$; all consumers participate $\zeta(\omega) = 1, \ \forall \omega$

- Union sells part of it to platform and extract its payoff

  i.e., it chooses $q \leq \bar{q}$ and sells it for to $V(q)$

- Union distributes proceeds to make consumer participation IC

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

Proposition

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

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

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 two alternative market designs that fix inefficiency