Competitive Markets for Personal Data

Simone Galperti UCSD

Jacopo Perego Columbia

February 2023

Motivation

Personal data fuels multi-billion dollar industries — essential input in digital economy

Status quo: consumers are imperfectly compensated for their data, and have limited control over their use $\rightsquigarrow$ possible market failures (Seim et al. 22)

Establishing competitive markets for personal data could improve efficiency and equity

Growing interest in studying theoretical properties of these markets and their problems e.g, Choi et al. 19, Bergemann et al. 22, Acemoglu et al. 22

This Paper

We identify a **novel externality** that makes these markets inefficient

Inefficiency can be so severe that consumers would be better off in an economy with expropriation (akin to status quo)

This Paper

We identify a **novel externality** that makes these markets inefficient

Inefficiency can be so severe that consumers would be better off in an economy with expropriation (akin to status quo)

This inefficiency:

- Originates from the specific role that **information intermediaries** play
- Present even if markets are otherwise **perfectly competitive**
- Present even if consumers data is **uncorrelated**

My Goal Today: Illustrate inefficiency through a simple example

example

(many consumers, one platform, one vendor)

A Simple Data Economy

The **vendor** sells a product to consumers and charges fee $a \in \{1, 2\}$

Consumers have unit demand and WTP of $\omega \in \{1, 2\} = \Omega$

Assume mass of consumers: $\bar{q}(\omega = 1) = 1$ and $\bar{q}(\omega = 2) = 2$

A Simple Data Economy

The **vendor** sells a product to consumers and charges fee $a \in \{1, 2\}$

Consumers have unit demand and WTP of $\omega \in \{1, 2\} = \Omega$

Assume mass of consumers: $\bar{q}(\omega = 1) = 1$ and $\bar{q}(\omega = 2) = 2$

Each consumer is uniquely identified by a **data record**, which reveals her ω

A collection of records is called a **database**, denoted by $q \in \mathbb{R}_+^{\Omega}$

A Simple Data Economy

Given a database q, the **platform** can intermediate these consumers with the vendor

Specifically, platform acts as **information designer** like in BBM15

- It sends a signal about a consumer's ω to the vendor to influence the fee he charges, and to maximize its own objective

A Simple Data Economy

Given a database q, the **platform** can intermediate these consumers with the vendor

Specifically, platform acts as **information designer** like in BBM15

- It sends a signal about a consumer's ω to the vendor to influence the fee he charges, and to maximize its own objective

Denote:

platform's expected payoff from database q

$$\overbrace{U(q)}$$

A Simple Data Economy

Given a database q, the **platform** can intermediate these consumers with the vendor

Specifically, platform acts as **information designer** like in BBM15

- It sends a signal about a consumer's ω to the vendor to influence the fee he charges, and to maximize its own objective

Denote:

platform's expected payoff from database q

$$\overbrace{U(q)}^{\text{platform's expected payoff from database } q} = \max_{x: A\times\Omega\to\mathbb{R}_+} \sum_{\omega,a} u(a,\omega)x(a,\omega)$$

$$\text{such that:} \quad \sum_{\omega} \big(\pi(a,\omega) - \pi(\hat{a},\omega)\big)x(a,\omega) \geq 0 \qquad \forall\, a,\hat{a}\in A$$

$$\sum_{a} x(a,\omega) = q(\omega) \qquad \forall\, \omega\in\Omega$$

a standard information design problem

A Simple Data Economy

Market for data records is assumed to be competitive

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation
   - **Platform** acquires data records from consumers

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation
   - **Platform** acquires data records from consumers
2. Platform intermediates consumers in its database with the vendor

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation
   - **Platform** acquires data records from consumers
2. Platform intermediates consumers in its database with the vendor

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation
   - **Platform** acquires data records from consumers
2. Platform intermediates consumers in its database with the vendor

Payoffs:

	if consumer: keeps record	sells record
Consumer	ε	
Vendor	0	
Platform	0	

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation
   - **Platform** acquires data records from consumers
2. Platform intermediates consumers in its database with the vendor

Payoffs:

	if consumer: keeps record	sells record	
Consumer	ε	$\max\{\omega - a, 0\}$	consumer's surplus
Vendor	0		
Platform	0		

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation
   - **Platform** acquires data records from consumers
2. Platform intermediates consumers in its database with the vendor

Payoffs:

	if consumer: keeps record	sells record	
Consumer	ε	$\max\{\omega - a, 0\}$	consumer's surplus
Vendor	0	$a\mathbb{1}(\omega \geq a)$	vendor's profit
Platform	0		

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation
   - **Platform** acquires data records from consumers
2. Platform intermediates consumers in its database with the vendor

Payoffs:

	if consumer: keeps record	sells record	
Consumer	ε	$\max\{\omega - a, 0\}$	consumer's surplus
Vendor	0	$a\mathbb{1}(\omega \geq a)$	vendor's profit
Platform	0	$\beta\ \max\{\omega - a, 0\}$	a fraction of CS

A Simple Data Economy

Market for data records is assumed to be competitive

1. Taking the **market price** $p(\omega)$ of each ω records as given:
   - Each **consumer** either:
     - Sells her record to the platform
     - Keeps her record and avoids platform's intermediation
   - **Platform** acquires data records from consumers
2. Platform intermediates consumers in its database with the vendor

Payoffs:

	if consumer: keeps record	sells record	
Consumer	ε	$\max\{\omega - a, 0\}$	consumer's surplus
Vendor	0	$a\mathbb{1}(\omega \geq a)$	vendor's profit
Platform	0	$\beta \ \max\{\omega - a, 0\}$	a fraction of CS

assume: $\beta < \varepsilon < 1/3$

Competitive Equilibrium

A **competitive equilibrium** is: prices $p^* \in \mathbb{R}_+^{\Omega}$; database $q^* \in \mathbb{R}_+^{\Omega}$; consumers' decisions $\alpha^* \in [0,1]^{\Omega}$ such that

Competitive Equilibrium

A **competitive equilibrium** is: prices $p^* \in \mathbb{R}_+^\Omega$; database $q^* \in \mathbb{R}_+^\Omega$; consumers' decisions $\alpha^* \in [0,1]^\Omega$ such that

1. Given p^*, platform optimizes.

$$q^* \in \arg\max_q U(q) - \sum_\omega p^*(\omega) q(\omega)$$

2. Given p^* and q^*, each consumer optimizes.

$$\alpha^*(\omega) \in \arg\max_\alpha (1-\alpha)\varepsilon + \alpha\Big(p^*(\omega) + \mathbb{E}_{q^*}(g(a,\omega))\Big)$$

3. And markets clear.

$$q^*(\omega) = \alpha^*(\omega)\bar{q}(\omega), \quad \forall \omega$$

Competitive Equilibrium

A **competitive equilibrium** is: prices $p^* \in \mathbb{R}_+^\Omega$; database $q^* \in \mathbb{R}_+^\Omega$; consumers' decisions $\alpha^* \in [0,1]^\Omega$ such that

1. Given p^*, platform optimizes.

$$q^* \in \arg\max_q U(q) - \sum_\omega p^*(\omega) q(\omega)$$

2. Given p^* and q^*, each consumer optimizes.

$$\alpha^*(\omega) \in \arg\max_\alpha (1-\alpha)\varepsilon + \alpha\Big(p^*(\omega) + \mathbb{E}_{q^*}(g(a,\omega))\Big)$$

3. And markets clear.

$$q^*(\omega) = \alpha^*(\omega)\bar{q}(\omega), \quad \forall \omega$$

In the paper: many platforms, many vendors, arbitrary objectives, partially informative records, multiple types

analysis

A Useful Step

Imagine platform **expropriated** consumers of their records ($\approx$ status quo)

How would the platform use database $\bar{q}$?

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database
$\omega = 1$	0	1
$\omega = 2$	0	2

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

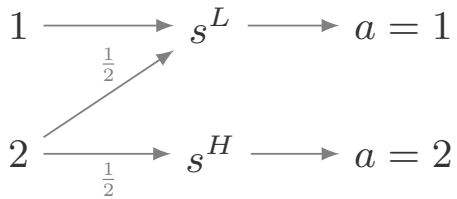
Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L$	$a = 1$
$\omega = 2$	0	$2 \to \frac{1}{2}: s^L$, $\frac{1}{2}: s^H$	$a = 2$

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	



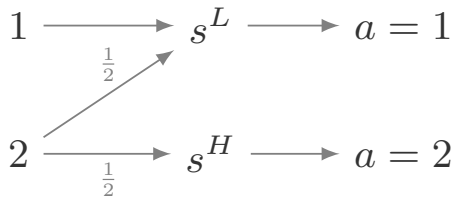
Payoffs		
C	**P**	**V**
1	β	4

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	



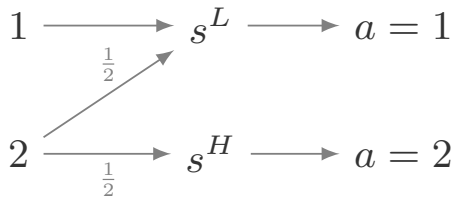
Payoffs		
C	**P**	**V**
1	β	4

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	



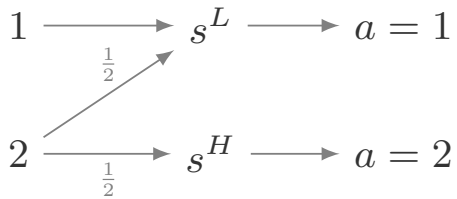
Payoffs		
C	**P**	**V**
1	β	4

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	



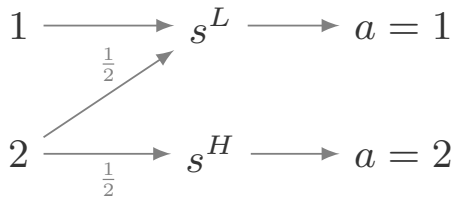
Payoffs		
C	**P**	**V**
1	β	4

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	



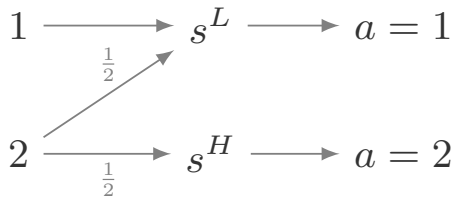
Payoffs		
C	**P**	**V**
1	β	4

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	



Payoffs		
C	**P**	**V**
1	β	4

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	

$\frac{1}{2}$ $\frac{1}{2}$

Payoffs		
C	**P**	**V**
1	β	4

Comments:

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	

$\frac{1}{2}$ $\frac{1}{2}$

Payoffs		
C	**P**	**V**
1	β	4

Comments:

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	

$\frac{1}{2}$ $\frac{1}{2}$

Payoffs		
C	**P**	**V**
1	β	4

Comments:

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	

$\frac{1}{2}$ $\frac{1}{2}$

Payoffs		
C	**P**	**V**
1	β	4

Comments:

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	

$\frac{1}{2}$ $\frac{1}{2}$

Payoffs		
C	**P**	**V**
1	β	4

Comments:

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$ (prob $\frac{1}{2}$)	
$\omega = 2$	0	$2 \to s^H \to a = 2$ (prob $\frac{1}{2}$)	

C	P	V
1	β	4

Payoffs

Comments:

1. Since platform's and vendor's objectives are not aligned, it is optimal to withhold some information (key feature of platform mkts: examples)

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	

(each arrow out of 1 and 2 has probability $\frac{1}{2}$)

Payoffs		
C	**P**	**V**
1	β	4

Comments:

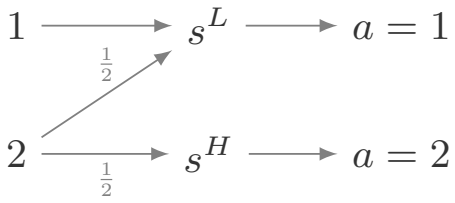
2. When data records are pooled, one consumer's payoff depends on what other records platform has

A Useful Step

Imagine platform **expropriated** consumers of their records (≈ status quo)

How would the platform use database $\bar{q}$?

Record Type	Records Kept	Platform's Database	Use
$\omega = 1$	0	$1 \to s^L \to a = 1$	
$\omega = 2$	0	$2 \to s^H \to a = 2$	



Payoffs		
C	**P**	**V**
1	β	4

Comments:

3. This "pooling externality" (GLP '22) is the root of inefficiency in our data economy

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	Prices	Records Kept	Platform's Database
$\omega = 1$	β	1	0
$\omega = 2$	0	2	0

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow$	$a = 1$
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow$	$a = 2$

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)		**Payoffs**	
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)		**Payoffs**	
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)		**Payoffs**	
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)		**Payoffs**	
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)		**Payoffs**	
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)		**Payoffs**	
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)		**Payoffs**	
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)		**Payoffs**	
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)	**Payoffs**		
					C	**P**	**V**
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$		3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$				

Comments:

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)	**Payoffs**		
					C	**P**	**V**
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$		3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$				

Comments:

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)	**Payoffs**		
					C	**P**	**V**
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$		3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$				

Comments:

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)	**Payoffs**		
					C	**P**	**V**
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$		3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$				

Comments:

1. Consumers have no incentive to sell: prices too low

$(\beta < \varepsilon)$

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)	**Payoffs**		
					C	**P**	**V**
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$		3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$				

Comments:

2. Platform has no (strict) incentive to buy: prices too high $(\beta < \varepsilon)$
   The value of a pair of records is β

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)	**Payoffs**		
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$	**C**	**P**	**V**
				3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$			

Comments:

3. Equilibrium is inefficient ("too little data"): Consumer welfare is lower than under expropriation (remember $\varepsilon < \frac{1}{3}$)

Competitive Equilibrium

Let's compute the (unique) **competitive equilibrium** of the data economy:

Record Type	**Prices**	**Records Kept**	**Platform's Database**	**Use** (Trivial)	**Payoffs**		
					C	**P**	**V**
$\omega = 1$	β	1	$0 \longrightarrow s^L \longrightarrow a = 1$		3ε	0	0
$\omega = 2$	0	2	$0 \longrightarrow s^H \longrightarrow a = 2$				

Comments:

4. High-type consumers would like to subsidize low-type consumers to sell their data, but this market is too incomplete

More Generally

Beyond this example?

More Generally

Beyond this example?

- Our data economy is generally inefficient, i.e., equilibrium data allocation does not maximize consumers' welfare

More Generally

Beyond this example?

- Our data economy is generally inefficient, i.e., equilibrium data allocation does not maximize consumers' welfare
- The inefficiency can be so severe that consumers' welfare would be higher in an economy with expropriation
  - Unitended consequence of giving consumers control

More Generally

Beyond this example?

- Our data economy is generally inefficient, i.e., equilibrium data allocation does not maximize consumers' welfare
- The inefficiency can be so severe that consumers' welfare would be higher in an economy with expropriation
  - Unitended consequence of giving consumers control
- "Pooling externalities" are the root of inefficiency:

  Economy is efficient when platforms don't have incentives to withhold information (back to example)

More Generally

Beyond this example?

- Our data economy is generally inefficient, i.e., equilibrium data allocation does not maximize consumers' welfare
- The inefficiency can be so severe that consumers' welfare would be higher in an economy with expropriation
  - Unitended consequence of giving consumers control
- "Pooling externalities" are the root of inefficiency:

  Economy is efficient when platforms don't have incentives to withhold information (back to example)
- This problem distinct and complementary to "social data"

More Generally

Beyond this example?

- Our data economy is generally inefficient, i.e., equilibrium data allocation does not maximize consumers' welfare
- The inefficiency can be so severe that consumers' welfare would be higher in an economy with expropriation
  - Unitended consequence of giving consumers control
- "Pooling externalities" are the root of inefficiency:

  Economy is efficient when platforms don't have incentives to withhold information (back to example)
- This problem distinct and complementary to "social data"

solutions?

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type
$\omega = 1$
$\omega = 2$

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type	**Prices**
$\omega = 1$	$p^*(1)$
$\omega = 2$	$p^*(2)$

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type	**Prices**
$\omega = 1$	$p^*(1, a = 1)$
	$p^*(1, a = 2)$
$\omega = 2$	$p^*(2, a = 1)$
	$p^*(2, a = 2)$

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type	**Prices**
$\omega = 1$	$p^*(1, a = 1)$
	$p^*(1, a = 2)$
$\omega = 2$	$p^*(2, a = 1)$
	$p^*(2, a = 2)$

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type	**Prices**
$\omega = 1$	$p^*(1,1) = \beta + 1 - \varepsilon$
	$p^*(1,2) = 0$
$\omega = 2$	$p^*(2,1) = -(1-\varepsilon)$
	$p^*(2,2) = 0$

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type	**Prices**	**Records Kept**
$\omega = 1$	$p^*(1,1) = \beta + 1 - \varepsilon$ $p^*(1,2) = 0$	0
$\omega = 2$	$p^*(2,1) = -(1-\varepsilon)$ $p^*(2,2) = 0$	1

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type	Prices	Records Kept	Records Used	Payoffs		
				C	**P**	**V**
$\omega = 1$	$p^*(1,1) = \beta + 1 - \varepsilon$ $p^*(1,2) = 0$	0	$x(1,1) = 1$ $x(1,2) = 0$	$1 + \varepsilon + \beta$	0	2
$\omega = 2$	$p^*(2,1) = -(1-\varepsilon)$ $p^*(2,2) = 0$	1	$x(2,1) = 1$ $x(2,2) = 0$			

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type	Prices	Records Kept	Records Used	Payoffs		
				C	**P**	**V**
$\omega = 1$	$p^*(1,1) = \beta + 1 - \varepsilon$ $p^*(1,2) = 0$	0	$x(1,1) = 1$ $x(1,2) = 0$	$1 + \varepsilon + \beta$	0	2
$\omega = 2$	$p^*(2,1) = -(1-\varepsilon)$ $p^*(2,2) = 0$	1	$x(2,1) = 1$ $x(2,2) = 0$			

Contracting on "Data Use"

A classic solution: Arrow (1969), Laffont (1978)

- Open markets where parties can trade the way data is **used**.
  This presumes data use can be contractible

Back to example:

Record Type	Prices	Records Kept	Records Used	Payoffs		
				C	**P**	**V**
$\omega = 1$	$p^*(1,1) = \beta + 1 - \varepsilon$ $p^*(1,2) = 0$	0	$x(1,1) = 1$ $x(1,2) = 0$	$1 + \varepsilon + \beta$	0	2
$\omega = 2$	$p^*(2,1) = -(1-\varepsilon)$ $p^*(2,2) = 0$	1	$x(2,1) = 1$ $x(2,2) = 0$			

Result. In this economy, every equilibrium data allocation maximizes consumer surplus (and vice versa)

Discussion

How realistic?

Discussion

How realistic?

Open Questions:

1. Intermediate solutions, partial decentralization?

Discussion

How realistic?

Open Questions:

1. Intermediate solutions, partial decentralization?
2. "Non-market" solutions: Data Unions? Seim et al. 2022

conclusions

Summary

We study competitive markets for personal data:

1. We identify a novel inefficiency that a data market can suffer from, even when it is otherwise perfectly competitive

   It shows how giving consumers control over their data can backfire

2. The inefficiency originates from how data is endogenously used by the intermediaries "pooling externality"

3. We discuss possible solutions and their limits

thank you