Competitive Markets for Personal Data

Simone Galperti
UCSD

Jacopo Perego
Columbia

June 2023

Motivation

Personal data fuels multi-billion dollar industries and is essential input to the digital economy

Yet, consumers are **imperfectly compensated** for their data, and have **limited control** over their use

- Expropriation and barter, common practice in the industry (FTC '15)

Motivation

Personal data fuels multi-billion dollar industries and is essential input to the digital economy

Yet, consumers are **imperfectly compensated** for their data, and have **limited control** over their use

- Expropriation and barter, common practice in the industry (FTC '15)

This status quo may be source of market failures (Seim et al. '22)

- A well-functioning market for data could avoid these problems

Motivation

Personal data fuels multi-billion dollar industries and is essential input to the digital economy

Yet, consumers are **imperfectly compensated** for their data, and have **limited control** over their use

- Expropriation and barter, common practice in the industry (FTC '15)

This status quo may be source of market failures (Seim et al. '22)

- A well-functioning market for data could avoid these problems **?**

This Paper

We study competitive markets for data and their equilibria

We show that "**pooling externalities**" can make these markets inefficient

- Inefficiency so severe that consumers could be better off under expropriation

This Paper

We study competitive markets for data and their equilibria

We show that "**pooling externalities**" can make these markets inefficient

- Inefficiency so severe that consumers could be better off under expropriation
- Inefficiency even if markets is **perfectly competitive**

This Paper

We study competitive markets for data and their equilibria

We show that "**pooling externalities**" can make these markets inefficient

- Inefficiency so severe that consumers could be better off under expropriation
- Inefficiency even if markets is **perfectly competitive**, and if consumers data is **uncorrelated** e.g., Acemoglu et al 22, Bergemann et al 22

This Paper

We study competitive markets for data and their equilibria

We show that "**pooling externalities**" can make these markets inefficient

- Inefficiency so severe that consumers could be better off under expropriation
- Inefficiency even if markets is **perfectly competitive**, and if consumers data is **uncorrelated** e.g., Acemoglu et al 22, Bergemann et al 22

Key ingredient: Information intermediaries, like platforms

- When withholding information ("pooling"), they create scope for externalities

leading example

(many consumers, one platform, one merchant)

A Simple Data Economy

A **merchant** sells widgets to consumers and charges a fee $a \in A$

A Simple Data Economy

A **merchant** sells widgets to consumers and charges a fee $a \in A$

A **consumer** has unit demand for the widget and WTP $\omega \in \{1, 2\}$

A Simple Data Economy

A **merchant** sells widgets to consumers and charges a fee $a \in A$

A **consumer** has unit demand for the widget and WTP $\omega \in \{1, 2\}$

Mass of consumers: $\bar{q}(1) = 1$ and $\bar{q}(2) = 2$

A Simple Data Economy

A **merchant** sells widgets to consumers and charges a fee $a \in A$

A **consumer** has unit demand for the widget and WTP $\omega \in \{1, 2\}$

Mass of consumers: $\bar{q}(1) = 1$ and $\bar{q}(2) = 2$

A Simple Data Economy

A **merchant** sells widgets to consumers and charges a fee $a \in A$

A **consumer** has unit demand for the widget and WTP $\omega \in \{1, 2\}$

Mass of consumers: $\bar{q}(1) = 1$ and $\bar{q}(2) = 2$

Each consumer owns a **data record** that, if sold to the **platform**, verifiably reveals her type ω

A Simple Data Economy

A **merchant** sells widgets to consumers and charges a fee $a \in A$

A **consumer** has unit demand for the widget and WTP $\omega \in \{1, 2\}$

Mass of consumers: $\bar{q}(1) = 1$ and $\bar{q}(2) = 2$

Each consumer owns a **data record** that, if sold to the **platform**, verifiably reveals her type ω

Two periods:

1. Consumers and platform trade data records in a competitive market
2. Given acquired data records, platform intermediates corresponding consumers with merchant

Period 1

Taking as given the **prices** for data records, $p(1)$ and $p(2)$

- Platform chooses how many records to buy, $q(1)$ and $q(2)$
- Consumer ω chooses whether to sell her record to platform: $\zeta(\omega)$

Period 1

Taking as given the **prices** for data records, $p(1)$ and $p(2)$

- Platform chooses how many records to buy, $q(1)$ and $q(2)$
- Consumer ω chooses whether to sell her record to platform: $\zeta(\omega)$
  - If she doesn't sell, she obtains reservation payoff $\underline{u} > 0$

Period 1

Taking as given the **prices** for data records, $p(1)$ and $p(2)$

- Platform chooses how many records to buy, $q(1)$ and $q(2)$
- Consumer ω chooses whether to sell her record to platform: $\zeta(\omega)$
  - If she doesn't sell, she obtains reservation payoff $\underline{u} > 0$
  - If she sells, price $+$ interacts with merchant in period 2...

Period 1

Taking as given the **prices** for data records, $p(1)$ and $p(2)$

- Platform chooses how many records to buy, $q(1)$ and $q(2)$
- Consumer ω chooses whether to sell her record to platform: $\zeta(\omega)$
  - If she doesn't sell, she obtains reservation payoff $\underline{u} > 0$
  - If she sells, price $+$ interacts with merchant in period 2...

Period 2

Given database $q = (q(1), q(2))$, platform acts as **information intermediary**:

- It provides merchant with information about each consumer in database, so to influence fee a

Period 2

Given database $q = (q(1), q(2))$, platform acts as **information intermediary**:

- It provides merchant with information about each consumer in database, so to influence fee a

Gross payoffs:

Merchant: $\pi(a, \omega) = a\mathbb{1}(\omega \geq a)$

Consumer: $u(a, \omega) = \max\{\omega - a, 0\}$

Platform: $v(a, \omega) = \beta\, u(a, \omega) + \gamma\, \pi(a, \omega)$

Period 2

Given database $q = (q(1), q(2))$, platform acts as **information intermediary**:

– It provides merchant with information about each consumer in database, so to influence fee a

Gross payoffs:

Merchant: $\pi(a, \omega) = a\mathbb{1}(\omega \geq a)$

Consumer: $u(a, \omega) = \max\{\omega - a, 0\}$

Platform: $v(a, \omega) = \beta\, u(a, \omega) + \gamma\, \pi(a, \omega)$

Formally, platform solves standard **information-design** problem:

Period 2

Given database $q = (q(1), q(2))$, platform acts as **information intermediary**:

– It provides merchant with information about each consumer in database, so to influence fee a

Gross payoffs:

Merchant: $\pi(a, \omega) = a\mathbb{1}(\omega \geq a)$

Consumer: $u(a, \omega) = \max\{\omega - a, 0\}$

Platform: $v(a, \omega) = \beta\, u(a, \omega) + \gamma\, \pi(a, \omega)$

Formally, platform solves standard **information-design** problem:

$$\overbrace{V(q)}^{\substack{\text{platform's} \\ \text{expected payoff}}}$$

Period 2

Given database $q = (q(1), q(2))$, platform acts as **information intermediary**:

– It provides merchant with information about each consumer in database, so to influence fee a

Gross payoffs:

Merchant: $\pi(a, \omega) = a\mathbb{1}(\omega \geq a)$

Consumer: $u(a, \omega) = \max\{\omega - a, 0\}$

Platform: $v(a, \omega) = \beta\, u(a, \omega) + \gamma\, \pi(a, \omega)$

Formally, platform solves standard **information-design** problem:

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

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

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

Period 2

Given database $q = (q(1), q(2))$, platform acts as **information intermediary**:

– It provides merchant with information about each consumer in database, so to influence fee a

Gross payoffs:

Merchant: $\pi(a, \omega) = a\mathbb{1}(\omega \geq a)$

Consumer: $u(a, \omega) = \max\{\omega - a, 0\}$

TODAY: $\gamma = 0 \Rightarrow$ Platform: $v(a, \omega) = \beta\, u(a, \omega) + \gamma\, \pi(a, \omega)$

Formally, platform solves standard **information-design** problem:

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

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

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

Equilibrium

An **equilibrium** consists of prices p^*, data demand q^*, data supply ζ^* s.t.

Equilibrium

An **equilibrium** consists of prices p^*, data demand q^*, data supply ζ^* s.t.

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

$$\max_q V(q) - \sum_\omega p^*(\omega) q(\omega)$$

Equilibrium

An **equilibrium** consists of prices p^*, data demand q^*, data supply ζ^* s.t.

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

$$\max_q V(q) - \sum_\omega p^*(\omega) q(\omega)$$

2. Given p^* and q^*, ζ^* solves consumers' problem:

$$\max_{z \in [0,1]} (1-z)\underline{u} + z\Big(\mathbb{E}_{q^*}(u(a,\omega)) + p^*(\omega)\Big), \quad \forall \omega$$

Equilibrium

An **equilibrium** consists of prices p^*, data demand q^*, data supply ζ^* s.t.

1. Given p^*, database q^* solves platform's problem:
$$\max_q V(q) - \sum_\omega p^*(\omega) q(\omega)$$
2. Given p^* and q^*, ζ^* solves consumers' problem:
$$\max_{z \in [0,1]} (1-z)\underline{u} + z\Big(\mathbb{E}_{q^*}(u(a,\omega)) + p^*(\omega)\Big), \quad \forall \omega$$
3. Markets clear
$$q^*(\omega) = \zeta^*(\omega)\bar{q}(\omega), \quad \forall \omega$$

Equilibrium

An **equilibrium** consists of prices p^*, data demand q^*, data supply ζ^* s.t.

1. Given p^*, database q^* solves platform's problem:
$$\max_q V(q) - \sum_\omega p^*(\omega) q(\omega)$$
2. Given p^* and q^*, ζ^* solves consumers' problem:
$$\max_{z \in [0,1]} (1-z)\underline{u} + z\Big(\mathbb{E}_{q^*}(u(a,\omega)) + p^*(\omega)\Big), \quad \forall \omega$$
3. Markets clear
$$q^*(\omega) = \zeta^*(\omega)\bar{q}(\omega), \quad \forall \omega$$

Equilibrium

An **equilibrium** consists of prices p^*, data demand q^*, data supply ζ^* s.t.

1. Given p^*, database q^* solves platform's problem:
$$\max_q V(q) - \sum_\omega p^*(\omega) q(\omega)$$
2. Given p^* and q^*, ζ^* solves consumers' problem:
$$\max_{z \in [0,1]} (1-z)\underline{u} + z\Big(\mathbb{E}_{q^*}(u(a,\omega)) + p^*(\omega)\Big), \quad \forall \omega$$
3. Markets clear
$$q^*(\omega) = \zeta^*(\omega)\bar{q}(\omega), \quad \forall \omega$$

Remarks:

- Endogenous "prior" q in an otherwise standard ID problem
- More general model in the paper

analysis

A Useful Step: Expropriation

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

How would the platform use the implied database?

A Useful Step: Expropriation

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

How would the platform use the implied database?

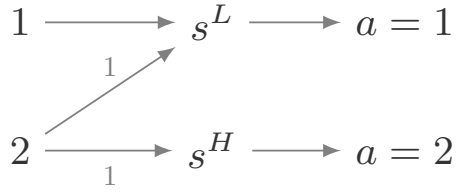
Record Type	Prices	Consumers Keep	Platform Buys
$\omega = 1$	0	0	1
$\omega = 2$	0	0	2

A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use
$\omega = 1$	0	0	$1 \to s^L$	$a = 1$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$

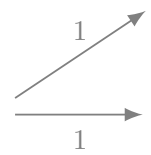


A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		

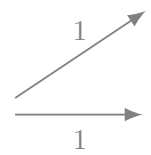


A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		

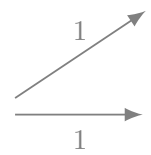


A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		

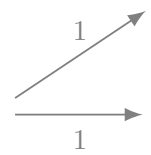


A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		

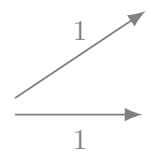


A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		

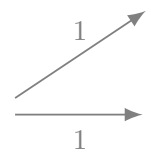


A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		

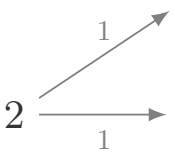


A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		



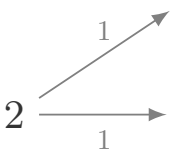
Comments:

A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		



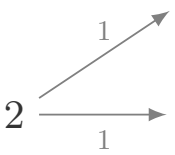
Comments:

A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		



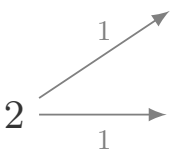
Comments:

A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		



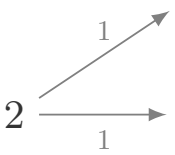
Comments:

A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		



Comments:

A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L$	$\to a = 1$	1
$\omega = 2$	0	0	$2 \to s^H$ (1 / 1)	$\to a = 2$	

Comments:

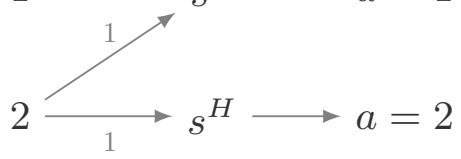
1. Conflict of interest btw platform and merchant: It is optimal to withhold some information

A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		1
$\omega = 2$	0	0	$2 \to s^H \to a = 2$		



Comments:

2. If records are pooled, a consumer's payoff depends on database composition (suppose a low type leaves...)

A Useful Step: Expropriation

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

How would the platform use the implied database?

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	0	0	$1 \to s^L \to a = 1$		
			$2 \to s^H \to a = 2$ (1)		1
$\omega = 2$	0	0			

Comments:

3. This will be the root of the inefficiency in our data economy

.

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	Prices
$\omega = 1$	β
$\omega = 2$	0

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	Prices	Consumers Keep	Platform Buys
$\omega = 1$	β	1	0
$\omega = 2$	0	2	0

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)
$\omega = 1$	β	1	0 $\longrightarrow$ s^L $\longrightarrow$	$a = 1$
$\omega = 2$	0	2	0 $\longrightarrow$ s^H $\longrightarrow$	$a = 2$

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Comments:

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Comments:

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Comments:

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Comments:

1. Low-type consumers have no incentive to sell:
   Price $p^*(1) = \beta$ is too low

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Comments:

2. High-type consumers have no incentive to sell:
   Price $p^*(2) = 0$ is too low

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Comments:

3. Platform has no strict incentive to buy at these prices.
   $p^* =$ marginal values of data — that's where $(\beta, 0)$ comes from

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Comments:

4. When $\underline{u}$ small, eqm welfare is lower than under expropriation
   Eqm is inefficient more broadly (not shown today)

Equilibrium: "Too Little Data"

Let's compute the **competitive equilibrium** of this data economy:

Imagine first $\beta < \underline{u}$ ("high value of privacy")

Record Type	**Prices**	**Consumers Keep**	**Platform Buys**	**Use** (Trivial)	**Consumer Welfare**
$\omega = 1$	β	1	0	$\longrightarrow s^L \longrightarrow a = 1$	$3\underline{u}$
$\omega = 2$	0	2	0	$\longrightarrow s^H \longrightarrow a = 2$	

Comments:

5. High-type consumers would like to subsidize low-type consumers by paying negative prices, but not an eqm

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices
$\omega = 1$	β
$\omega = 2$	0

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys
$\omega = 1$	β	0	1
$\omega = 2$	0	0	2

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

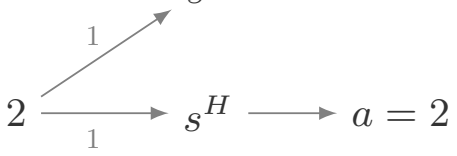
Record Type	Prices	Consumers Keep	Platform Buys	Use
$\omega = 1$	β	0	$1 \xrightarrow{1} s^L$	$a = 1$
$\omega = 2$	0	0	$2 \xrightarrow{1} s^H$	$a = 2$

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

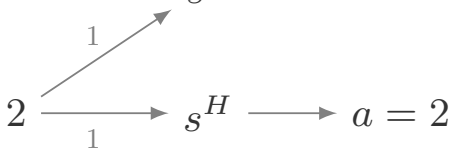


Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

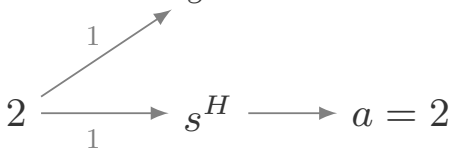


Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

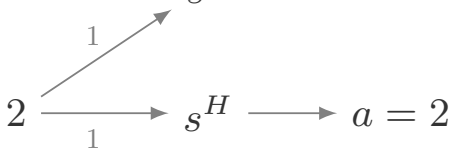


Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

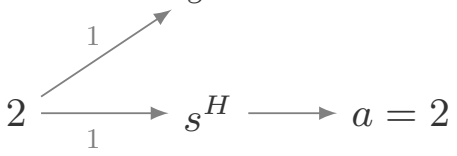


Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

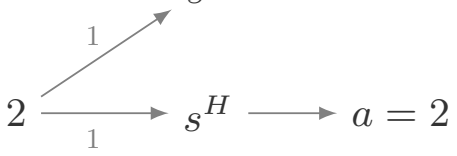


Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

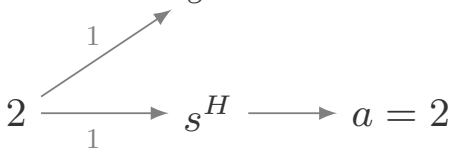


Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	



Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

Comments:

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

Comments:

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$	$a = 2$	

Comments:

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \xrightarrow{1} s^L$	$\rightarrow a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \xrightarrow{1} s^H$	$\rightarrow a = 2$	

Comments:

1. Low-types consumers sell, since $p^*(1) = \beta \geq \underline{u}$

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$ (1) / $2 \to s^L$ (1)	$a = 2$	

Comments:

2. High-types consumers sell too, attracted by the possibility of being charged low fee

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \to s^L$	$a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \to s^H$ (1)	$a = 2$	

Comments:

3. Consumers do better than under expropriation. But eqm still inefficient: Too many high-type consumers sell

Equilibrium: "Too Much Data"

Consider now the opposite case: $\beta \geq \underline{u}$ "low" value of privacy

Let's compute again **competitive equilibrium**:

Record Type	Prices	Consumers Keep	Platform Buys	Use	Consumer Welfare
$\omega = 1$	β	0	$1 \xrightarrow{1} s^L$	$\longrightarrow a = 1$	$1 + \beta$
$\omega = 2$	0	0	$2 \xrightarrow{1} s^H$	$\longrightarrow a = 2$	

Comments:

4. Negative price on high-type consumers? Again, not an equilibrium...

Takeaways from Example

Competition does not always lead to efficient data allocation (for all $\underline{u}, \beta$)

Welfare can be even lower than under expropriation

- Perverse consequence of empowering consumers

Takeaways from Example

Competition does not always lead to efficient data allocation (for all $\underline{u}, \beta$)

Welfare can be even lower than under expropriation

- Perverse consequence of empowering consumers

Market failure originates from information intermediation:

- Conflict of interest $\rightsquigarrow$ info withholding $\rightsquigarrow$ pooling externalities
- In paper: Necessity result (no COI $\rightsquigarrow$ efficiency)

conclusion

Summary

1. A framework to study competitive markets for personal data and their equilibria

2. We show that **pooling externalities** are a source of market failure

3. In paper, we discuss possible remedies
   - Prices conditional on use (Lindhal Economy)
   - Data unions

remedies

An Economy with Lindhal Prices

How can we fix inefficiencies discussed so far?

An Economy with Lindhal Prices

How can we fix inefficiencies discussed so far?

A classic solution: following e.g. Arrow 69, Laffont 78

- Platform has to buy record for a **specific purpose** (i.e. an action recommendation a)
- More complete markets: Prices are $p(\omega, a)$

An Economy with Lindhal Prices

How can we fix inefficiencies discussed so far?

A classic solution: following e.g. Arrow 69, Laffont 78

- Platform has to buy record for a **specific purpose** (i.e. an action recommendation a)
- More complete markets: Prices are $p(\omega, a)$

Note: This presumes data use is contractible

An Economy with Lindhal Prices

How can we fix inefficiencies discussed so far?

A classic solution: following e.g. Arrow 69, Laffont 78

- Platform has to buy record for a **specific purpose** (i.e. an action recommendation a)
- More complete markets: Prices are $p(\omega, a)$

Note: This presumes data use is contractible

Result. Equilibria in this economy exist and are (first-best) efficient.

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

Record Type	Prices	Records Kept	Platform's Database	Use
$\omega = 1$	$p^*(1,1) = \beta + 1 - \varepsilon$ $p^*(1,2) = 0$	0	$1 \longrightarrow s^L$	$\longrightarrow a = 1$
$\omega = 2$	$p^*(2,1) = -(1-\varepsilon)$ $p^*(2,2) = 0$	1	$1 \nearrow s^L$	

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

Record Type	Prices	Records Kept	Platform's Database	Use	Welfare
$\omega = 1$	$p^*(1,1) = \beta + 1 - \varepsilon$ $p^*(1,2) = 0$	0	1	$\rightarrow s^L \rightarrow a = 1$	$1 + \beta + \varepsilon$
$\omega = 2$	$p^*(2,1) = -(1-\varepsilon)$ $p^*(2,2) = 0$	1	1		

Back to Example

To illustrate, return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

Record Type	Prices	Records Kept	Platform's Database	Use	Welfare
$\omega = 1$	$p^*(1,1) = \beta + 1 - \varepsilon$ $p^*(1,2) = 0$	0	1 → s^L → $a = 1$		$1 + \beta + \varepsilon$
$\omega = 2$	$p^*(2,1) = -(1-\varepsilon)$ $p^*(2,2) = 0$	1	1 → s^L → $a = 1$		

Comments

- High-type consumers subsidize the platform to acquire low-type consumers data
- Previously this was not an equilibrium. Why?

How Realistic?

How Realistic?

It captures a qualitative feature of recent privacy-protection policies

- EU's GDPR: "*The **specific purposes** for which personal data are processed should be explicit and determined at the time of the collection of the personal data*"

How Realistic?

It captures a qualitative feature of recent privacy-protection policies

- EU's GDPR: "*The **specific purposes** for which personal data are processed should be explicit and determined at the time of the collection of the personal data*"

However, number of markets required is unrealistically large

How Realistic?

It captures a qualitative feature of recent privacy-protection policies

- EU's GDPR: "*The **specific purposes** for which personal data are processed should be explicit and determined at the time of the collection of the personal data*"

However, number of markets required is unrealistically large

Open Questions:

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

⏩ to conclusions

general model

Model 1/3

There are I competing **platforms**

Model 1/3

There are I competing **platforms**

A **merchant** is active on platform i and sells product for a_i

Model 1/3

There are I competing **platforms**

A **merchant** is active on platform i and sells product for a_i

Consumers has preference type ω

Model 1/3

There are I competing **platforms**

A **merchant** is active on platform i and sells product for a_i

Consumers has preference type ω

Type-ω consumers are of mass $\bar{q}(\omega) \geq 0$

Model 1/3

There are I competing **platforms**

A **merchant** is active on platform i and sells product for a_i

Consumers has preference type ω

Type-ω consumers are of mass $\bar{q}(\omega) \geq 0$

When a ω-consumer transacts with platform i's vendor, payoffs realize

Model 1/3

There are I competing **platforms**

A **merchant** is active on platform i and sells product for a_i

Consumers has preference type ω

Type-ω consumers are of mass $\bar{q}(\omega) \geq 0$

When a ω-consumer transacts with platform i's vendor, payoffs realize

$$\begin{aligned} \text{consumer:} \quad & g_i(a_i, \omega) \\ \text{merchant:} \quad & \pi_i(a_i, \omega) \\ \text{platform:} \quad & u_i(a_i, \omega) \end{aligned}$$

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)
2. Exclusive access to this consumer

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)
2. Exclusive access to this consumer (e.g., email, IP address, etc.)

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)
2. Exclusive access to this consumer (e.g., email, IP address, etc.)

Exclusivity is key: Data record is rival good

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)
2. Exclusive access to this consumer (e.g., email, IP address, etc.)

Exclusivity is key: Data record is rival good

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)
2. Exclusive access to this consumer (e.g., email, IP address, etc.)

Exclusivity is key: Data record is rival good

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)
2. Exclusive access to this consumer (e.g., email, IP address, etc.)

Exclusivity is key: Data record is rival good

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)
2. Exclusive access to this consumer (e.g., email, IP address, etc.)

Exclusivity is key: Data record is rival good

Model 2/3: Data Records

Each consumer is uniquely identified by a **data record**, which provides:

1. Info about this consumer's WTP ω (e.g., age, gender, edu, etc.)
2. Exclusive access to this consumer (e.g., email, IP address, etc.)

Exclusivity is key: Data record is rival good

A collection of data records is called a **database**: denoted $q_i \in \mathbb{R}_+^{\Omega}$

Model 3/3: Platforms

Suppose platform i has acquired database q_i

Model 3/3: Platforms

Suppose platform i has acquired database q_i

- i.e., i has exclusive access to consumers whose records belong to q_i

Model 3/3: Platforms

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

Model 3/3: Platforms

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

$$\max_{x_i : A \times \Omega \to \mathbb{R}_+} \quad \sum_{\omega, a} u_i(a, \omega) x_i(a, \omega)$$

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

$$\sum_{a} x_i(a, \omega) = q_i(\omega) \qquad \forall\, \omega \in \Omega$$

Model 3/3: Platforms

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

$$\max_{x_i : A \times \Omega \to \mathbb{R}_+} \quad \sum_{\omega, a} u_i(a, \omega) x_i(a, \omega)$$

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

$$\sum_{a} x_i(a, \omega) = q_i(\omega) \qquad \forall\, \omega \in \Omega$$

Model 3/3: Platforms

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

$$\overbrace{U_i(q_i)}^{\substack{\text{platform's} \\ \text{exptd payoff}}} = \max_{x_i : A \times \Omega \to \mathbb{R}_+} \sum_{\omega, a} u_i(a, \omega) x_i(a, \omega)$$

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

$$\sum_{a} x_i(a, \omega) = q_i(\omega) \qquad \forall\, \omega \in \Omega$$

Model 3/3: Platforms

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

$$\overbrace{U_i(q_i)}^{\substack{\text{platform's}\\ \text{exptd payoff}}} = \max_{x_i : A\times\Omega\to\mathbb{R}_+} \sum_{\omega,a} u_i(a,\omega)x_i(a,\omega)$$

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

$$\sum_{a} x_i(a,\omega) = q_i(\omega) \qquad \forall\, \omega\in\Omega$$

Denote by $x^*_{q_i}$ a solution (note: it depends on the entire database)

Model 3/3: Platforms

Suppose platform i has acquired database q_i

i.e., i has exclusive access to consumers whose records belong to q_i

Platform i is an **information designer**: It sends a signal about each consumers's ω to its vendor to influence his price a

$$\overbrace{U_i(q_i)}^{\substack{\text{platform's}\\ \text{exptd payoff}}} = \max_{x_i : A\times\Omega\to\mathbb{R}_+} \sum_{\omega,a} u_i(a,\omega)x_i(a,\omega)$$

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

$$\sum_{a} x_i(a,\omega) = q_i(\omega) \qquad \forall\, \omega\in\Omega$$

Denote by $x^*_{q_i}$ a solution (note: it depends on the entire database)

Summing Up

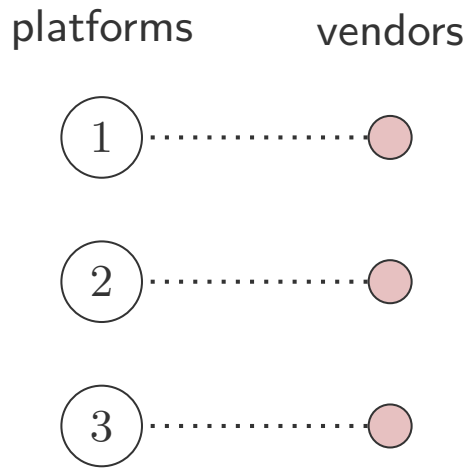
platforms

1

2

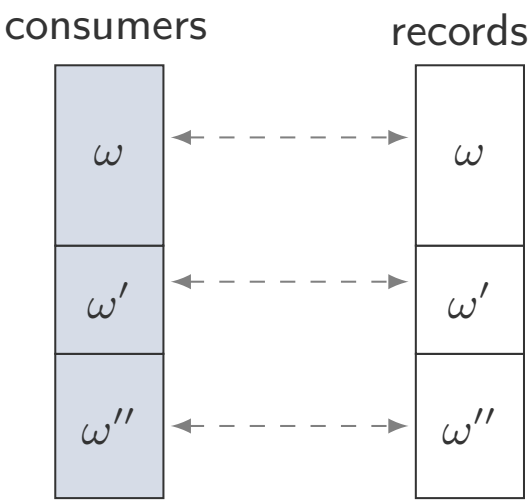
3

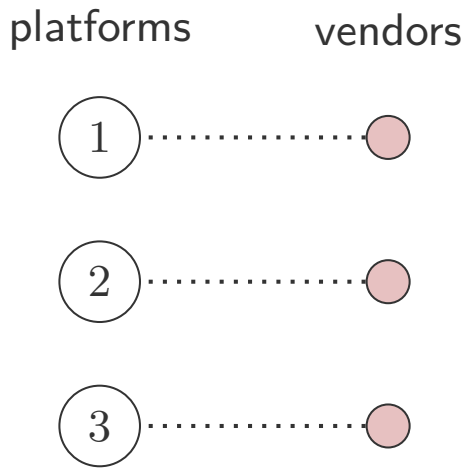
Summing Up



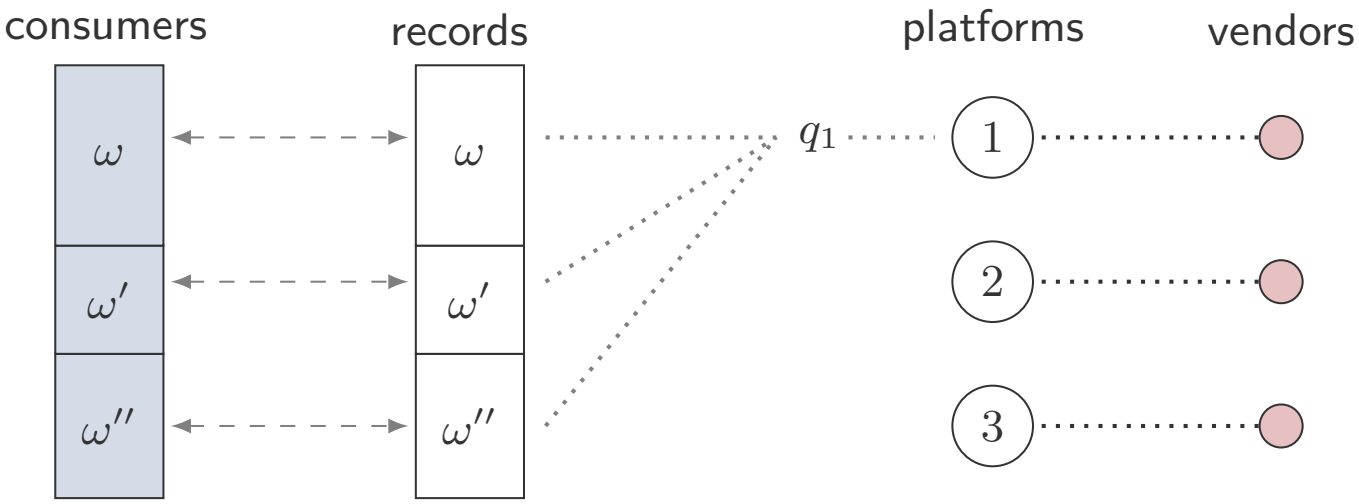
Summing Up

Summing Up


consumers
records
ω
ω'
ω''
ω
ω'
ω''


platforms
vendors
1
2
3

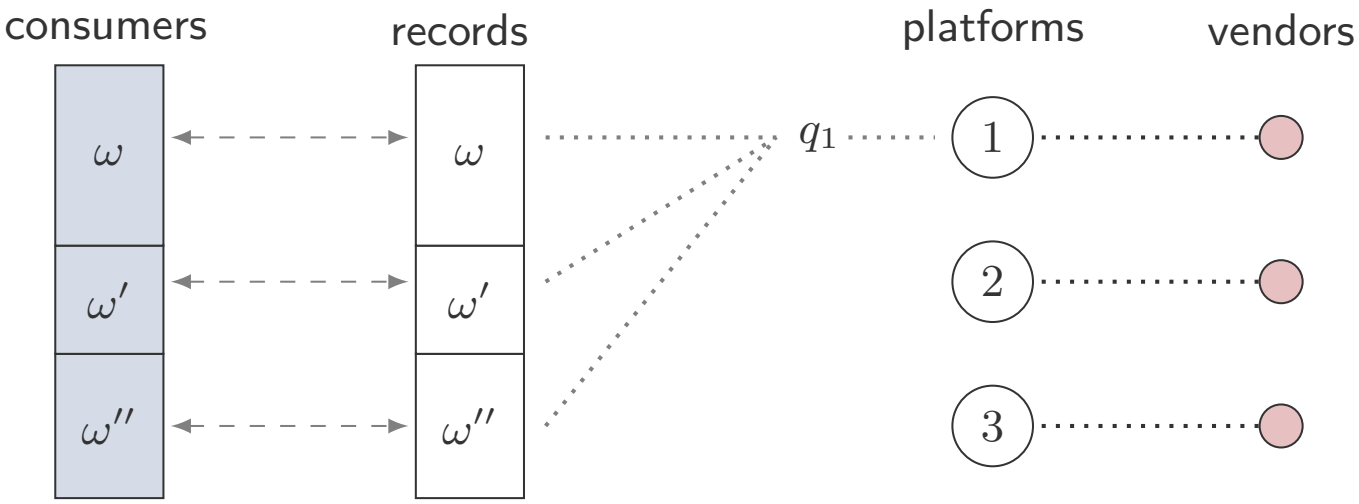
Summing Up



Timing:

1. Platform i acquires q_i in a **competitive market** for data records

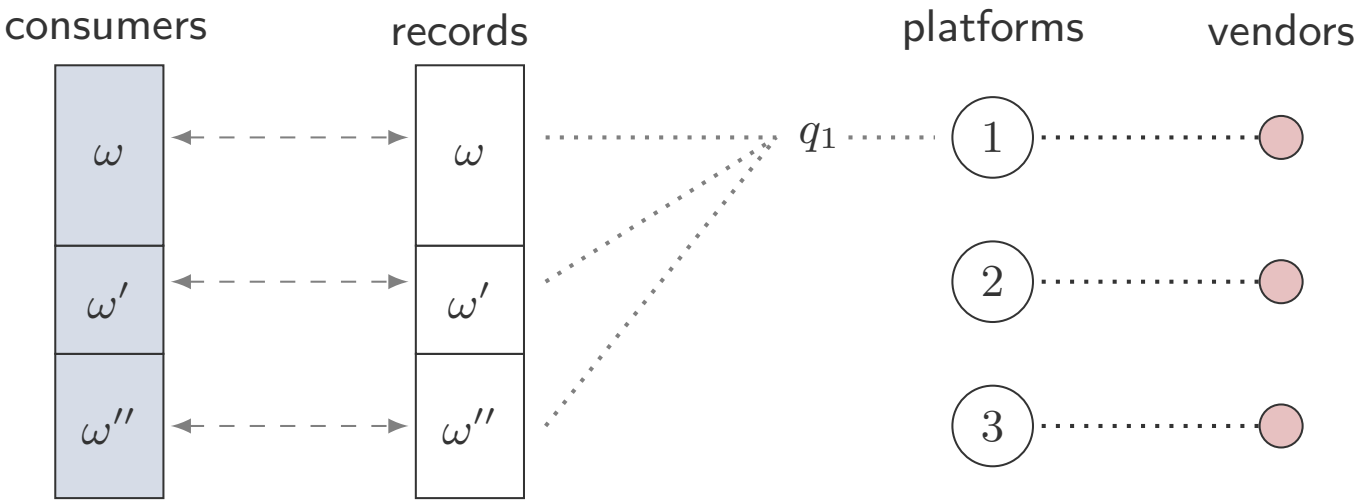
Summing Up



Timing:

1. Platform i acquires q_i in a **competitive market** for data records
2. Given database q_i, platform i chooses $x^*_{q_i}$

Summing Up



Timing:

1. Platform i acquires q_i in a **competitive market** for data records
2. Given database q_i, platform i chooses $x^*_{q_i}$
3. Payoffs realize

results

Three Economies

We study the welfare properties of three different economies:

$\mathcal{E}_1$ An economy with expropriation

Platforms own consumers data and can trade

$\mathcal{E}_2$ An economy with data ownership

Consumers own their data and can trade

$\mathcal{E}_3$ An economy with data ownership and Lindhal prices

Data are priced conditional on how it is used

$\mathcal{E}_1$ – An Economy with Expropriation

In this economy:

- Consumers "expropriated" of their records: no control, imperfect compns
- Platforms trade records among each other, taking prices as given

$\mathcal{E}_1$ – An Economy with Expropriation

In this economy:

- Consumers "expropriated" of their records: no control, imperfect compns
- Platforms trade records among each other, taking prices as given

Definition. Equilibrium in $\mathcal{E}_1$

Prices $p^* \in \mathbb{R}^{\Omega}$ and a feasible data allocation $q^* \in \mathbb{R}_+^{\Omega \times I}$ are an equilibrium of $\mathcal{E}_1$ if:

1. Platforms maximize given prices $\quad q_i^* \in \arg\max_{q_i} U_i(q_i) - \sum_{\omega} p^*(\omega) q_i(\omega)$
2. All markets clear $\quad \text{for all } \omega,\ p^*(\omega)\left(\bar{q}(\omega) - \sum_i q_i^*(\omega)\right) = 0$

$\mathcal{E}_1$ – An Economy with Expropriation

Platform i's payoff depends only on q_i, not on q_j (exclusivity)

$\mathcal{E}_1$ – An Economy with Expropriation

Platform i's payoff depends only on q_i, not on q_j (exclusivity)

Proposition. Equilibrium Characterization in $\mathcal{E}_1$

Equilibria of $\mathcal{E}_1$ exist and maximize the sum of platforms' payoffs

Every platform-optimal allocation can be supported as an equilibrium of $\mathcal{E}_1$

Equilibria are solutions of a grand-information design (LP) problem

$\mathcal{E}_1$ – An Economy with Expropriation

Platform i's payoff depends only on q_i, not on q_j (exclusivity)

Proposition. Equilibrium Characterization in $\mathcal{E}_1$

Equilibria of $\mathcal{E}_1$ exist and maximize the sum of platforms' payoffs

Every platform-optimal allocation can be supported as an equilibrium of $\mathcal{E}_1$

Equilibria are solutions of a grand-information design (LP) problem

$\mathcal{E}_1$ – An Economy with Expropriation

Platform i's payoff depends only on q_i, not on q_j (exclusivity)

Proposition. Equilibrium Characterization in $\mathcal{E}_1$

Equilibria of $\mathcal{E}_1$ exist and maximize the sum of platforms' payoffs

Every platform-optimal allocation can be supported as an equilibrium of $\mathcal{E}_1$

Equilibria are solutions of a grand-information design (LP) problem

$\mathcal{E}_2$ – An Economy with Data Ownership

$\mathcal{E}_2$ is the economy we analyzed earlier

$\mathcal{E}_2$ – An Economy with Data Ownership

$\mathcal{E}_2$ is the economy we analyzed earlier

Equilibrium in $\mathcal{E}_2$:

Prices $p^* \in \mathbb{R}^{\Omega}$, data allocation $q^* \in \mathbb{R}_+^{\Omega \times (I+1)}$, consumers' decisions $\alpha^* \in (\Delta(I))^{\Omega}$ are an equilibrium if:

1. Given p^*, database q_i^* solves platform i's problem

$$q_i^* \in \arg\max_{q_i} U_i(q_i) - \sum_{\omega} p^*(\omega) q_i(\omega)$$

2. Given p^* and q^*, $\alpha^*(\omega)$ solves ω-consumer's problem

$$\alpha^*(\omega) \in \arg\max_{\alpha(\omega) \in \Delta(I)} (1 - \alpha(0|\omega)) r(\omega) + \sum_i \alpha(i|\omega) \Big(p^*(\omega) + \mathbb{E}_{q_i^*}(g_i(a_i, \omega)) \Big)$$

3. Markets clear

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

$\mathcal{E}_2$ – An Economy with Data Ownership

What We Know:

- Equilibrium *can* be inefficient $\rightsquigarrow$ our leading example
- Sufficient conditions for efficiency:

Proposition. No-Intermediation Case

When $u_i = \pi_i$ for all i, equilibria in $\mathcal{E}_2$ exist and are efficient

$\mathcal{E}_2$ – An Economy with Data Ownership

What We Know:

- Equilibrium *can* be inefficient $\rightsquigarrow$ our leading example
- Sufficient conditions for efficiency:

Proposition. No-Intermediation Case

When $u_i = \pi_i$ for all i, equilibria in $\mathcal{E}_2$ exist and are efficient

What We Don't Know (yet):

- Sufficient conditions for inefficiency beyond examples?
- Sufficient conditions for existence in the intermediation case?

$\mathcal{E}_3$ – An Economy with Lindhal prices

How can we fix inefficiencies discussed so far?

$\mathcal{E}_3$ – An Economy with Lindhal prices

How can we fix inefficiencies discussed so far?

We enrich our economy by opening "more complete" markets

following e.g. Arrow 69, Laffont 78

- Consumers can sell record for a **specific purpose** (i.e. an action a_i)
- A richer price system: prices $p_i(\omega, a_i)$ depend on record type, on platform identity, and on intended use a_i

$\mathcal{E}_3$ – An Economy with Lindhal prices

How can we fix inefficiencies discussed so far?

We enrich our economy by opening "more complete" markets

following e.g. Arrow 69, Laffont 78

- Consumers can sell record for a **specific purpose** (i.e. an action a_i)
- A richer price system: prices $p_i(\omega, a_i)$ depend on record type, on platform identity, and on intended use a_i

This presumes data use is contractible

$\mathcal{E}_3$ – An Economy with Lindhal prices

How can we fix inefficiencies discussed so far?

We enrich our economy by opening "more complete" markets

following e.g. Arrow 69, Laffont 78

- Consumers can sell record for a **specific purpose** (i.e. an action a_i)
- A richer price system: prices $p_i(\omega, a_i)$ depend on record type, on platform identity, and on intended use a_i

This presumes data use is contractible

Proposition. Equilibrium Characterization in $\mathcal{E}_3$

Equilibria in $\mathcal{E}_3$ exist and are (first-best) efficient.

Every (first-best) efficient data allocation can be supported in an eqm

Equilibria are solutions of a grand-information design (LP) problem

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Back to Example

Illustrate functioning of economy $\mathcal{E}_3$ with our earlier example

Return to case of market unravelling ($\beta < \epsilon$, $\gamma = 0$):

$$u(a, \omega) = \beta\ g(a, \omega) + \gamma\ \pi(a, \omega)$$

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

Comments

- High-type consumers subsidize the platform to acquire low-type consumers data
- Previously this was not an equilibrium. Why?

How Realistic is $\mathcal{E}_3$?

How Realistic is $\mathcal{E}_3$?

It captures a qualitative feature of recent privacy-protection policies

- EU's GDPR: "*The **specific purposes** for which personal data are processed should be explicit and determined at the time of the collection of the personal data*"

How Realistic is $\mathcal{E}_3$?

It captures a qualitative feature of recent privacy-protection policies

- EU's GDPR: "*The **specific purposes** for which personal data are processed should be explicit and determined at the time of the collection of the personal data*"

However, number of markets required is unrealistically large

How Realistic is $\mathcal{E}_3$?

It captures a qualitative feature of recent privacy-protection policies

- EU's GDPR: "*The **specific purposes** for which personal data are processed should be explicit and determined at the time of the collection of the personal data*"

However, number of markets required is unrealistically large

Open Questions:

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

conclusions

Summary

1. We introduce framework to study competitive markets for personal data and their equilibria

- Rather general setting: many platforms, many merchants, arbitrary objectives, partially informative records, multiple types

2. We identify a novel externality that can make these markets inefficient

- The way platforms withhold information creates externalities that can lead to market failures

3. We discuss possible remedies and their limits

thank you