Competitive Markets for Personal Data

Simone Galperti Tianhao Liu Jacopo Perego

December 2024

Motivation

Consumers supply a crucial input for modern economy: their **personal data**

Yet, they often have **limited control** over how and by whom their data is used:

- This may lead to inefficiencies and inequality

New legislation gives consumers more control over their data

- Lays foundations upon which **data markets** could emerge

What properties would these markets have, and how should they be designed to promote desirable outcomes?

This Paper

Model. A stylized **competitive data market** where

- Consumers own their data and can sell it to intermediaries
- Intermediaries use the acquired consumer data to provide information to merchants and interact them with the consumers

This Paper

Model. A stylized **competitive data market** where

- Consumers own their data and can sell it to intermediaries
- Intermediaries use the acquired consumer data to provide information to merchants and interact them with the consumers

Main Results

1. We identify **novel externality** in this market and show it depends on how intermediaries use the data:

- If full disclosure $\Rightarrow$ No Externalities $\Rightarrow$ Efficiency
- If some pooling $\Rightarrow$ Externalities $\Rightarrow$ Inefficiencies

This Paper

Model. A stylized **competitive data market** where

- Consumers own their data and can sell it to intermediaries
- Intermediaries use the acquired consumer data to provide information to merchants and interact them with the consumers

Main Results

1. We identify **novel externality** in this market and show it depends on how intermediaries use the data:

- If full disclosure $\Rightarrow$ No Externalities $\Rightarrow$ Efficiency
- If some pooling $\Rightarrow$ Externalities $\Rightarrow$ Inefficiencies

2. Propose three solutions to this market failure:

- Data unions; Data taxes; "Lindahl" pricing for the data

model

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω in the population

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω in the population

Each consumer owns a **data record** that fully reveals her corresponding ω

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω in the population

Each consumer owns a **data record** that fully reveals her corresponding ω

Each consumer has an outside option $r \sim F_\omega$ (fully supported on $\mathbb{R}$)

Basic Ingredients

A unit mass of consumers, one (representative) platform, one merchant

Each **consumer** has unit demand for merchant's product

Her **preference type** is denoted by $\omega \in \Omega$ (finite) (e.g., WTP, location, ...)

Let $\bar{q} \in \Delta(\Omega)$ be the distribution of ω in the population

Each consumer owns a **data record** that fully reveals her corresponding ω

Each consumer has an outside option $r \sim F_\omega$ (fully supported on $\mathbb{R}$)

Two periods: 1. Data markets are open 2. Product market is open

Period 1: Competitive Data Markets

Period 1: Competitive Data Markets

The consumers and the platform trade data records at prices $p = (p(\omega))_{\omega \in \Omega}$, which they take as given

Period 1: Competitive Data Markets

The consumers and the platform trade data records at prices $p = (p(\omega))_{\omega \in \Omega}$, which they take as given

The demand side:

- Platform demands **database** $q = (q(\omega))_{\omega \in \Omega}$, for which it pays $\sum_\omega q(\omega)p(\omega)$

Period 1: Competitive Data Markets

The consumers and the platform trade data records at prices $p = (p(\omega))_{\omega\in\Omega}$, which they take as given

The demand side:

- Platform demands **database** $q = (q(\omega))_{\omega\in\Omega}$, for which it pays $\sum_\omega q(\omega)p(\omega)$

The supply side:

- If a type-ω consumer sells her record to the platform, she is paid $p(\omega)$ and is later intermediated with merchant
- If she doesn't, she enjoys her outside option $r \sim F_\omega$

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It sends signal to merchant about each consumer in its database
- Given signal, the merchant chooses an action $a \in A$ (finite)
- Together, a and ω determine period-2 payoffs:

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It sends signal to merchant about each consumer in its database
- Given signal, the merchant chooses an action $a \in A$ (finite)
- Together, a and ω determine period-2 payoffs:

Consumer's:	$u(a, \omega)$	e.g., trading surplus
Merchant's:	$\pi(a, \omega)$	e.g., profits
Platform's:	$v(a, \omega)$	

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It sends signal to merchant about each consumer in its database
- Given signal, the merchant chooses an action $a \in A$ (finite)
- Together, a and ω determine period-2 payoffs:

Consumer's:	$u(a, \omega)$	e.g., trading surplus
Merchant's:	$\pi(a, \omega)$	e.g., profits
Platform's:	$v(a, \omega)$	

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It sends signal to merchant about each consumer in its database
- Given signal, the merchant chooses an action $a \in A$ (finite)

Remark: Info design problem equivalent to a linear program: (BM '16)

$$V(q) = \max_{x:\Omega \to \Delta(A)} \sum_{\omega, a} v(a, \omega) x(a|\omega) q(\omega)$$

$$\text{s.t. } \forall a, a': \; \sum_{\omega} \Big(\pi(a, \omega) - \pi(a', \omega) \Big) x(a|\omega) q(\omega) \geq 0$$

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It sends signal to merchant about each consumer in its database
- Given signal, the merchant chooses an action $a \in A$ (finite)

Remark: Info design problem equivalent to a linear program: (BM '16)

$$V(q) = \max_{x:\Omega \to \Delta(A)} \sum_{\omega,a} v(a,\omega)x(a|\omega)q(\omega)$$

$$\text{s.t. } \forall a, a' \colon \sum_{\omega} \Big(\pi(a,\omega) - \pi(a',\omega)\Big) x(a|\omega)q(\omega) \geq 0$$

Denote its solutions by $\mathcal{X}(q)$ (standard ID problem, but with endogenous prior)

Period 2: Product Market

Given acquired database q, platform acts as **information designer**

- It sends signal to merchant about each consumer in its database
- Given signal, the merchant chooses an action $a \in A$ (finite)

Remark: Info design problem equivalent to a linear program: (BM '16)

$$V(q) = \max_{x:\Omega\to\Delta(A)} \sum_{\omega,a} v(a,\omega)x(a|\omega)q(\omega)$$

$$\text{s.t. } \forall a, a' \colon \ \sum_{\omega} \Big(\pi(a,\omega) - \pi(a',\omega)\Big) x(a|\omega)q(\omega) \geq 0$$

Denote its solutions by $\mathcal{X}(q)$ (standard ID problem, but with endogenous prior)

Equilibrium

Definition

A profile (p^*, q^*, x^*)

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

(a). Given p^*, q^* solves the platform's problem in the first period, i.e.,

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

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

(a). Given p^*, q^* solves the platform's problem in the first period, i.e.,

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

(b). Given q^*, x^* solves platform's info-design problem, i.e., $x^* \in \mathcal{X}(q^*)$

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

(a). Given p^*, q^* solves the platform's problem in the first period, i.e.,

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

(b). Given q^*, x^* solves platform's info-design problem, i.e., $x^* \in \mathcal{X}(q^*)$

(c). Given p^* and x^*, a type-ω consumer with outside option r sells record if

$$r \leq p^*(\omega) + \sum_a x^*(a|\omega) u(a, \omega)$$

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

(a). Given p^*, q^* solves the platform's problem in the first period, i.e.,

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

(b). Given q^*, x^* solves platform's info-design problem, i.e., $x^* \in \mathcal{X}(q^*)$

(c). Given p^* and x^*, a type-ω consumer with outside option r sells record if

$$r \leq p^*(\omega) + \sum_a x^*(a|\omega) u(a, \omega) \triangleq r^*(\omega)$$

Equilibrium

Definition

A profile (p^*, q^*, x^*) is an equilibrium of the economy if:

(a). Given p^*, q^* solves the platform's problem in the first period, i.e.,

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

(b). Given q^*, x^* solves platform's info-design problem, i.e., $x^* \in \mathcal{X}(q^*)$

(c). Given p^* and x^*, a type-ω consumer with outside option r sells record if

$$r \leq p^*(\omega) + \sum_a x^*(a|\omega) u(a, \omega) \triangleq r^*(\omega)$$

(d). Data markets clear, i.e. $\quad q^*(\omega) = \bar{q}(\omega) F_{\omega}(r^*(\omega)) \qquad \forall \omega$

discussion

Discussion

Results extend to large class of **information-intermediation problems**:

- Multiple agents (e.g., competing merchants)
- Arbitrary downstream (finite) games (e.g., a second-price auctions, hotelling)
- More than information design (e.g., platform takes a enforceable action)

Leading applications: Online marketplaces and advertisement auctions

Discussion

Results extend to large class of **information-intermediation problems**:

- Multiple agents (e.g., competing merchants)
- Arbitrary downstream (finite) games (e.g., a second-price auctions, hotelling)
- More than information design (e.g., platform takes a enforceable action)

Leading applications: Online marketplaces and advertisement auctions

Substantive assumptions we made:

- The data market is competitive
- Platform is a "gate keeper" — alt see BB '23
- A data record combines "access" and information — alt see ALV '22

Two Useful Technical Lemmas

Lemma

For generic v, $\mathcal{X}(q)$ is a singleton **for all** q, which we denote by x_q

Two Useful Technical Lemmas

Lemma

For generic v, $\mathcal{X}(q)$ is a singleton **for all** q, which we denote by x_q

The paper studies these generic economies

Two Useful Technical Lemmas

Lemma

For generic v, $\mathcal{X}(q)$ is a singleton **for all** q, which we denote by x_q

The paper studies these generic economies

Lemma

For all $q \in \mathbb{R}^{\Omega}_{++}$, $V(q)$ and x_q are directionally differentiable in q.

Moreover, they are differentiable at almost all $q \in \mathbb{R}^{\Omega}_{++}$

Two Useful Technical Lemmas

Lemma

For generic v, $\mathcal{X}(q)$ is a singleton **for all** q, which we denote by x_q

The paper studies these generic economies

Lemma

For all $q \in \mathbb{R}^{\Omega}_{++}$, $V(q)$ and x_q are directionally differentiable in q.

Moreover, they are differentiable at almost all $q \in \mathbb{R}^{\Omega}_{++}$

For this talk only: I'll focus on differentiable $V(q)$ and x_q

efficiency

Efficiency of the Data Market

Our Main Question: Is the **competitive** data market in which platform and consumers participate **efficient**?

- Namely, does it maximize platform and consumers welfare?

Efficiency of the Data Market

Our Main Question: Is the **competitive** data market in which platform and consumers participate **efficient**?

- Namely, does it maximize platform and consumers welfare?

Definition

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

$$\max_{q,x} \quad \sum_{a,\omega} v(a,\omega)x(a|\omega)q(\omega) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega} \bar{q}(\omega) \int_{F_\omega^{-1}\left(\frac{q(\omega)}{\bar{q}(\omega)}\right)}^{\infty} r dF_\omega(r)$$

$$\text{s.t.} \quad q \leq \bar{q} \quad \text{and} \quad x \in \mathcal{X}(q)$$

Efficiency of the Data Market

Our Main Question: Is the **competitive** data market in which platform and consumers participate **efficient**?

- Namely, does it maximize platform and consumers welfare?

Definition

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

$$\max_{q,x} \quad \sum_{a,\omega} v(a,\omega)x(a|\omega)q(\omega) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega} \bar{q}(\omega)\int_{F_\omega^{-1}\left(\frac{q(\omega)}{\bar{q}(\omega)}\right)}^{\infty} r dF_\omega(r)$$

$$\text{s.t.} \quad q \leq \bar{q} \ \text{ and } \ x \in \mathcal{X}(q)$$

Efficiency of the Data Market

Our Main Question: Is the **competitive** data market in which platform and consumers participate **efficient**?

- Namely, does it maximize platform and consumers welfare?

Definition

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

$$\begin{aligned} \max_{q,x} \quad & \sum_{a,\omega} v(a,\omega)x(a|\omega)q(\omega) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega} \bar{q}(\omega) \int_{F_\omega^{-1}\left(\frac{q(\omega)}{\bar{q}(\omega)}\right)}^{\infty} r dF_\omega(r) \\ \text{s.t.} \quad & q \le \bar{q} \;\; \text{and} \;\; x \in \mathcal{X}(q) \end{aligned}$$

No apriori reason why competitive market should maximize merchant's welfare:
To illustrate inefficiency, less-demanding benchmark more desirable

Efficiency of the Data Market

Our Main Question: Is the **competitive** data market in which platform and consumers participate **efficient**?

- Namely, does it maximize platform and consumers welfare?

Definition

An allocation (q, x_q) is **efficient** if it maximizes

$$\begin{aligned} \max_{q} \quad & \sum_{a,\omega} v(a,\omega) x_q(a|\omega) q(\omega) + \sum_{a,\omega} u(a,\omega) x_q(a|\omega) q(\omega) + \sum_{\omega} \bar{q}(\omega) \int_{F_\omega^{-1}\left(\frac{q(\omega)}{\bar{q}(\omega)}\right)}^{\infty} r dF_\omega(\\ \text{s.t.} \quad & q \leq \bar{q} \end{aligned}$$

No apriori reason why competitive market should maximize merchant's welfare:

To illustrate inefficiency, less-demanding benchmark more desirable

Efficiency of the Data Market

Our Main Question: Is the **competitive** data market in which platform and consumers participate **efficient**?

- Namely, does it maximize platform and consumers welfare?

Definition

An allocation (q, x_q) is **efficient** if it maximizes

$$\max_{q} \quad V(q) \quad + \sum_{a,\omega} u(a,\omega) x_q(a|\omega) q(\omega) + \sum_{\omega} \bar{q}(\omega) \int_{F_\omega^{-1}\left(\frac{q(\omega)}{\bar{q}(\omega)}\right)}^{\infty} r dF_\omega($$

$$\text{s.t.} \quad q \leq \bar{q}$$

No apriori reason why competitive market should maximize merchant's welfare:

To illustrate inefficiency, less-demanding benchmark more desirable

data externality

Data Externality

Why could there be an externality in this data market?

A high-level intuition:

- If type-ω consumer sells her record, she affects $q(\omega)$, which may change x_q, which affects other consumers' payoffs

Data Externality

Why could there be an externality in this data market?

A high-level intuition:

– If type-ω consumer sells her record, she affects $q(\omega)$, which may change x_q, which affects other consumers' payoffs

To formalize this intuition, let's compare:

– How records are allocated in equilibrium

– How records are allocated by social planner

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$p^*(\omega) + \sum_a u(a, \omega) x_{q^*}(a|\omega)$$

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$p^*(\omega) + \sum_a u(a, \omega) x_{q^*}(a|\omega)$$

In equilibrium, it must hold that $p^*(\omega) = \dfrac{\partial V(q^*)}{\partial q^*(\omega)}$ platform makes no profit

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega)$$

In equilibrium, it must hold that $p^*(\omega) = \dfrac{\partial V(q^*)}{\partial q^*(\omega)}$ platform makes no profit

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega)$$

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega)$$

The **gross social gain** when an additional type-ω consumer sells her record is

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a, \omega) x_{q^*}(a|\omega)$$

The **gross social gain** when an additional type-ω consumer sells her record is

$$\frac{\partial}{\partial q^*(\omega)} \Big(V(q^*) + \sum_{a,\omega} u(a, \omega) x_{q^*}(a|\omega) q^*(\omega) \Big)$$

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega)$$

The **gross social gain** when an additional type-ω consumer sells her record is

$$\frac{\partial}{\partial q^*(\omega)} \Big(V(q^*) + \sum_{a,\omega} u(a,\omega) x_{q^*}(a|\omega) q^*(\omega) \Big)$$
$$= \frac{\partial V(q^*)}{\partial q^*(\omega)}$$

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a, \omega) x_{q^*}(a|\omega)$$

The **gross social gain** when an additional type-ω consumer sells her record is

$$\begin{aligned} &\frac{\partial}{\partial q^*(\omega)} \Big(V(q^*) + \sum_{a,\omega} u(a,\omega) x_{q^*}(a|\omega) q^*(\omega) \Big) \\ = &\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a, \omega) x_{q^*}(a|\omega) \end{aligned}$$

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega)$$

The **gross social gain** when an additional type-ω consumer sells her record is

$$\frac{\partial}{\partial q^*(\omega)} \Big(V(q^*) + \sum_{a,\omega} u(a,\omega) x_{q^*}(a|\omega) q^*(\omega) \Big)$$
$$= \frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega) + \sum_{a,\omega'} q^*(\omega') u(a,\omega') \tfrac{\partial x_{q^*}(a|\omega')}{\partial q^*(\omega)}$$

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega)$$

The **gross social gain** when an additional type-ω consumer sells her record is

$$\begin{aligned} &\frac{\partial}{\partial q^*(\omega)} \Big(V(q^*) + \sum_{a,\omega} u(a,\omega) x_{q^*}(a|\omega) q^*(\omega) \Big) \\ &= \frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega) + \underbrace{\sum_{a,\omega'} q^*(\omega') u(a,\omega') \frac{\partial x_{q^*}(a|\omega')}{\partial q^*(\omega)}}_{\textbf{data externality}} \end{aligned}$$

Data Externality

Fix an equilibrium (p^*, q^*, x_{q^*})

The **gross private gain** when an additional type-ω consumer sells her record is

$$\frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega)$$

The **gross social gain** when an additional type-ω consumer sells her record is

$$\begin{aligned} &\frac{\partial}{\partial q^*(\omega)} \Big(V(q^*) + \sum_{a,\omega} u(a,\omega) x_{q^*}(a|\omega) q^*(\omega) \Big) \\ &= \frac{\partial V(q^*)}{\partial q^*(\omega)} + \sum_a u(a,\omega) x_{q^*}(a|\omega) + \underbrace{\sum_{a,\omega'} q^*(\omega') u(a,\omega') \frac{\partial x_{q^*}(a|\omega')}{\partial q^*(\omega)}}_{\textbf{data externality}} \end{aligned}$$

Efficiency Characterization

The data externality arises because the platform's use of consumer data may depend on the specific data it acquires

Efficiency Characterization

The data externality arises because the platform's use of consumer data may depend on the specific data it acquires

Proposition

An equilibrium (p^*, q^*, x_{q^*}) is efficient only if

$$\sum_{a,\omega'} q^*(\omega') u(a,\omega') \frac{\partial x_{q^*}(a|\omega')}{\partial q^*(\omega)} = 0 \qquad \forall \omega$$

If planner's payoff is concave, such a condition is also sufficient for efficiency

Efficiency Characterization

The data externality arises because the platform's use of consumer data may depend on the specific data it acquires

Proposition

An equilibrium (p^*, q^*, x_{q^*}) is efficient only if

$$\sum_{a,\omega'} q^*(\omega') u(a,\omega') \frac{\partial x_{q^*}(a|\omega')}{\partial q^*(\omega)} = 0 \qquad \forall \omega$$

If planner's payoff is concave, such a condition is also sufficient for efficiency

So what?

Efficiency Characterization

The data externality arises because the platform's use of consumer data may depend on the specific data it acquires

Proposition

An equilibrium (p^*, q^*, x_{q^*}) is efficient only if

$$\sum_{a,\omega'} q^*(\omega')u(a,\omega')\frac{\partial x_{q^*}(a|\omega')}{\partial q^*(\omega)} = 0 \qquad \forall\omega$$

If planner's payoff is concave, such a condition is also sufficient for efficiency

So what? E.g., what ways of using consumer data leads to inefficiency?

A Typology of Recommendation Mechanisms

A Typology of Recommendation Mechanisms

Definition

A recommendation mechanism $x : \Omega \to \Delta(A)$ is a:

1. **Full-Disclosure Mechanism** if, for each ω, it recommends only actions the merchant would choose under complete information

A Typology of Recommendation Mechanisms

Definition

A recommendation mechanism $x : \Omega \to \Delta(A)$ is a:

1. **Full-Disclosure Mechanism** if, for each ω, it recommends only actions the merchant would choose under complete information

2. **Mixing(?) Mechanism** if there is ω at which it recommends more than one action to the merchant

A Typology of Recommendation Mechanisms

Definition

A recommendation mechanism $x : \Omega \to \Delta(A)$ is a:

1. **Full-Disclosure Mechanism** if, for each ω, it recommends only actions the merchant would choose under complete information
2. **Mixing(?) Mechanism** if there is ω at which it recommends more than one action to the merchant

Note: If $x_q \in (1)$ then $x_q \notin (2)$.

A Typology of Recommendation Mechanisms

Definition

A recommendation mechanism $x : \Omega \to \Delta(A)$ is a:

1. **Full-Disclosure Mechanism** if, for each ω, it recommends only actions the merchant would choose under complete information
2. **Mixing(?) Mechanism** if there is ω at which it recommends more than one action to the merchant

Note: If $x_q \in (1)$ then $x_q \notin (2)$. Converse is false (e.g., no-disclosure)

Result 1: Full Disclosure Leads to Efficiency

Result 1: Full Disclosure Leads to Efficiency

Proposition

If x_q is a full-disclosure mechanism at some $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Result 1: Full Disclosure Leads to Efficiency

Proposition

If x_q is a full-disclosure mechanism at some $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Proof Sketch:

- If full disclosure is optimal at interior q, it is optimal at all q

Result 1: Full Disclosure Leads to Efficiency

Proposition

If x_q is a full-disclosure mechanism at some $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Proof Sketch:

- If full disclosure is optimal at interior q, it is optimal at all q
- Thus, x_q constant in $q \Rightarrow$ (1) No data externality; (2) Planner's payoff is concave $\Rightarrow$ Efficiency

Result 1: Full Disclosure Leads to Efficiency

Proposition

If x_q is a full-disclosure mechanism at some $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Proof Sketch:

- If full disclosure is optimal at interior q, it is optimal at all q
- Thus, x_q constant in $q \Rightarrow$ (1) No data externality; (2) Planner's payoff is concave $\Rightarrow$ Efficiency

Corollary

Fix an equilibrium (p^*, q^*, x^*). If x^* is a full-disclosure mechanism, the equilibrium is efficient.

Result 1: Full Disclosure Leads to Efficiency

Proposition

If x_q is a full-disclosure mechanism at some $q \in \mathbb{R}^{\Omega}_{++}$, all equilibria are efficient.

Proof Sketch:

- If full disclosure is optimal at interior q, it is optimal at all q
- Thus, x_q constant in $q \Rightarrow$ (1) No data externality; (2) Planner's payoff is concave $\Rightarrow$ Efficiency

Corollary

Fix an equilibrium (p^*, q^*, x^*). If x^* is a full-disclosure mechanism, the equilibrium is efficient.

Result 2: Mixing Mechanisms Lead to Inefficiency

Previous result implies that, in any inefficient equilibrium, the platform **must withhold** some information from the merchant

Result 2: Mixing Mechanisms Lead to Inefficiency

Previous result implies that, in any inefficient equilibrium, the platform **must withhold** some information from the merchant

Proposition

Fix an equilibrium (p^*, q^*, x^*). If x^* is a mixing mechanism, the equilibrium is inefficient

Result 2: Mixing Mechanisms Lead to Inefficiency

Previous result implies that, in any inefficient equilibrium, the platform **must withhold** some information from the merchant

Proposition

Fix an equilibrium (p^*, q^*, x^*). If x^* is a mixing mechanism, the equilibrium is inefficient

(Poor) Intuition. Exact mixture depends on q

Result 2: Mixing Mechanisms Lead to Inefficiency

Previous result implies that, in any inefficient equilibrium, the platform **must withhold** some information from the merchant

Proposition

Fix an equilibrium (p^*, q^*, x^*). If x^* is a mixing mechanism, the equilibrium is inefficient

(Poor) Intuition. Exact mixture depends on q

Work in progress: "Only if" direction also holds under some additional conditions

an application

Mediated Price Discrimination

So far, minimal assumptions on the intermediation problem

Let's specialize setting to canonical application: Price discrimination *à la* BBM

Mediated Price Discrimination

So far, minimal assumptions on the intermediation problem

Let's specialize setting to canonical application: Price discrimination *à la* BBM

- Let $\omega \in \mathbb{R}_{++}$ be the consumer's WTP for merchant's product
- Let a denote price the merchant sets for his product
- Players payoffs are

$$\begin{aligned} &\text{Consumer's:} && u(a,\omega) = \max\{\omega - a, 0\} \\ &\text{Merchant's:} && \pi(a,\omega) = a\, \mathbb{1}(\omega \geq a) \\ &\text{Platform's:} && v(a,\omega) = \gamma_u\, u(a,\omega) + \gamma_\pi\, \pi(a,\omega) \end{aligned}$$

Equilibrium Characterization

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

Equilibrium Characterization

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

Proposition

Fix any equilibrium.

- If $\gamma_u < \gamma_\pi$, the equilibrium is efficient. Consumers' welfare is maximized
- If $\gamma_u > \gamma_\pi$, the equilibrium is efficient if and only the platform provides no information to the merchant.

Equilibrium Characterization

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

Proposition

Fix any equilibrium.

- If $\gamma_u < \gamma_\pi$, the equilibrium is efficient. Consumers' welfare is maximized
- If $\gamma_u > \gamma_\pi$, the equilibrium is efficient if and only the platform provides no information to the merchant.

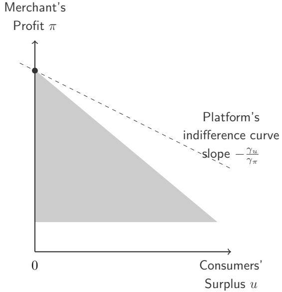
That is, any "non-trivial" use of information by the platform will lead to inefficiencies

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$

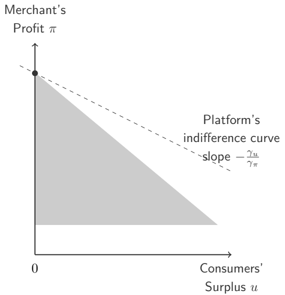
Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



Intuition for Efficiency

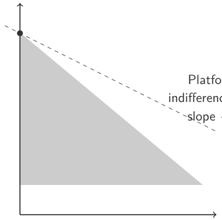
If $\gamma_u < \gamma_\pi$



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

Intuition for Efficiency

If $\gamma_u < \gamma_\pi$



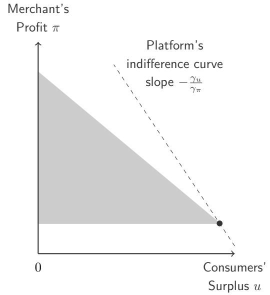
- At all q, **full disclosure** is optimal
- Merchant extracts surplus from all consumers
- Therefore, $x^*(a, \omega)$ does not depend on q
- Therefore, no externality! All equilibria are constrained efficient

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$

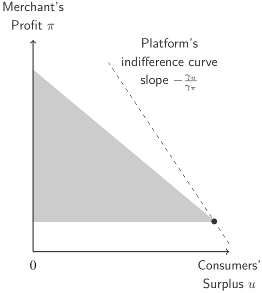
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



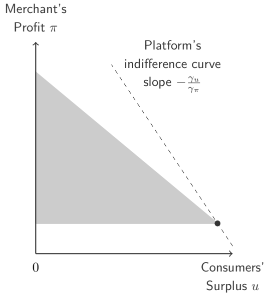
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



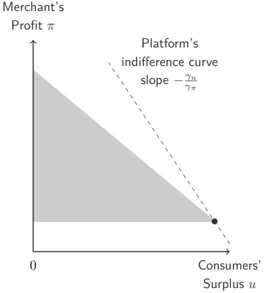
Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



Intuition for Inefficiency

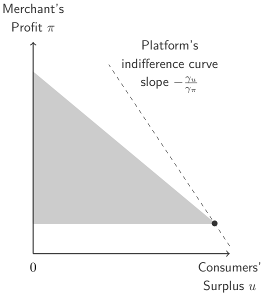
If $\gamma_u > \gamma_\pi$



– Platform **withholds information** from merchant

Intuition for Inefficiency

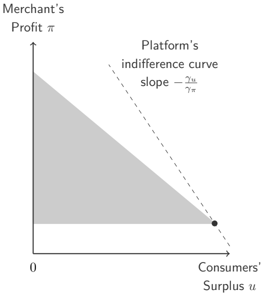
If $\gamma_u > \gamma_\pi$



- Platform **withholds information** from merchant
- Pooling types required to keep merchant indifferent at some as

Intuition for Inefficiency

If $\gamma_u > \gamma_\pi$



- Platform **withholds information** from merchant
- Pooling types required to keep merchant indifferent at some as
- Thus, x_q depends on q
- Thus, $\sum_{\omega'} q(\omega') \frac{\partial U(\omega', x^*)}{\partial q^*(\omega)}$ can be non-zero: data externality

Bigger Picture

Application highlights when there is **no conflict of interest** btw platform and merchant $\Rightarrow$ full disclosure is optimal $\Rightarrow$ data market is efficient

Special case: if platform is the merchant $\quad\Rightarrow\quad$ no intermediation

Bigger Picture

Application highlights when there is **no conflict of interest** btw platform and merchant $\Rightarrow$ full disclosure is optimal $\Rightarrow$ data market is efficient

Special case: if platform is the merchant $\quad\Rightarrow\quad$ no intermediation

Thus the source of the inefficiency is the role platforms play as **information intermediaries**

- Platforms typically balance conflicting interests, which they rarely resolve with full disclosure — otw, no info-design literature! :)
- Instead, they often garble the data they have collected

Bigger Picture

Application highlights when there is **no conflict of interest** btw platform and merchant $\Rightarrow$ full disclosure is optimal $\Rightarrow$ data market is efficient

Special case: if platform is the merchant $\Rightarrow$ no intermediation

Thus the source of the inefficiency is the role platforms play as **information intermediaries**

- Platforms typically balance conflicting interests, which they rarely resolve with full disclosure — otw, no info-design literature! :)
- Instead, they often garble the data they have collected

This paper shows how this practice can lead to a failure of the first-welfare theorem in a competitive data market

example

A Simple Example to Illustrate

Suppose:

- $\gamma_u > \gamma_\pi = 0$, i.e. platform only cares about consumers' surplus
- Only two types of consumers: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Suppose $\bar{r} < \frac{1+\gamma_u}{2}$ so that some trade is efficient

A Simple Example to Illustrate

Suppose:

- $\gamma_u > \gamma_\pi = 0$, i.e. platform only cares about consumers' surplus
- Only two types of consumers: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Suppose $\bar{r} < \frac{1+\gamma_u}{2}$ so that some trade is efficient

There is a **unique** constrained-efficient allocation (q°, x°):

- All low-type consumers sell: $q^\circ(1) = \bar{q}(1)$
- Only some high-type consumers sell: $q^\circ(2) = \bar{q}(1) < \bar{q}(2)$
- Platform provides no info to merchant, who charges lowest fee to all consumers in database: $x^\circ(a = 1|\omega) = 1, \ \forall\omega$

A Simple Example to Illustrate

Suppose:

- $\gamma_u > \gamma_\pi = 0$, i.e. platform only cares about consumers' surplus
- Only two types of consumers: $\Omega = \{1, 2\}$ with $\bar{q}(1) < \bar{q}(2)$
- Suppose $\bar{r} < \frac{1+\gamma_u}{2}$ so that some trade is efficient

There is a **unique** constrained-efficient allocation (q°, x°):

- All low-type consumers sell: $q^\circ(1) = \bar{q}(1)$
- Only some high-type consumers sell: $q^\circ(2) = \bar{q}(1) < \bar{q}(2)$
- Platform provides no info to merchant, who charges lowest fee to all consumers in database: $x^\circ(a = 1|\omega) = 1, \ \forall\omega$

A Simple Example to Illustrate

Claim: If $\gamma_u < \bar{r}$, all equilibria are inefficient $\rightsquigarrow$ no trade (Corollary 1)

A Simple Example to Illustrate

Claim: If $\gamma_u < \bar{r}$, all equilibria are inefficient $\rightsquigarrow$ no trade (Corollary 1)

It can be shown that $p^*(\omega) \leq \gamma_u$. This implies that:

- Low-type consumers do not want to sell their records, $q^*(1) = 0$

  Why? $U^*(1) = p^*(1) \leq \gamma_u < \bar{r}$

  Do not internalize positive externality that selling their record generate for high-type consumers

A Simple Example to Illustrate

Claim: If $\gamma_u < \bar{r}$, all equilibria are inefficient $\rightsquigarrow$ no trade (Corollary 1)

It can be shown that $p^*(\omega) \leq \gamma_u$. This implies that:

- Low-type consumers do not want to sell their records, $q^*(1) = 0$

  Why? $U^*(1) = p^*(1) \leq \gamma_u < \bar{r}$

  Do not internalize positive externality that selling their record generate for high-type consumers

- Hence, high-type consumer do not want to sell either

  Why? $U^*(2) = p^*(2) \leq \gamma_u < \bar{r}$

A Simple Example to Illustrate

Claim: If $\gamma_u < \bar{r}$, all equilibria are inefficient $\rightsquigarrow$ no trade (Corollary 1)

It can be shown that $p^*(\omega) \leq \gamma_u$. This implies that:

- Low-type consumers do not want to sell their records, $q^*(1) = 0$

  Why? $U^*(1) = p^*(1) \leq \gamma_u < \bar{r}$

  Do not internalize positive externality that selling their record generate for high-type consumers

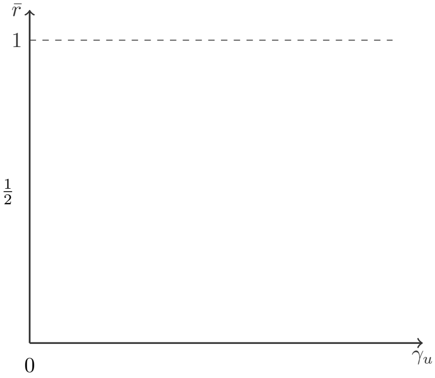
- Hence, high-type consumer do not want to sell either

  Why? $U^*(2) = p^*(2) \leq \gamma_u < \bar{r}$

- Market unravels $\rightsquigarrow$ No trade $\rightsquigarrow$ Inefficiency

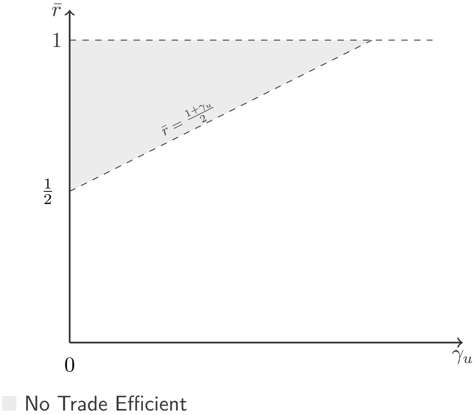
Example

Complete equilibrium characterization for this example:



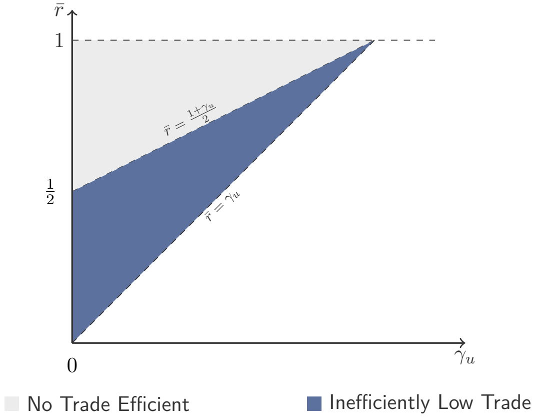
Example

Complete equilibrium characterization for this example:



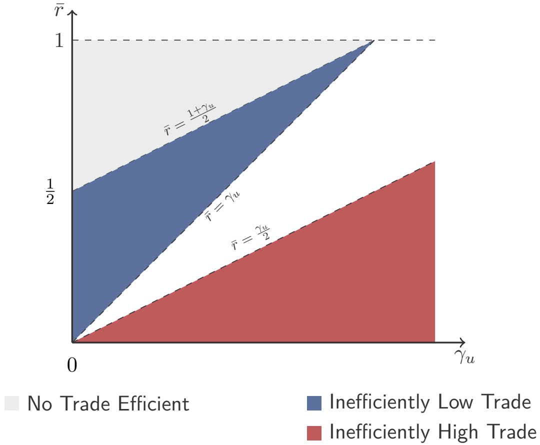
Example

Complete equilibrium characterization for this example:



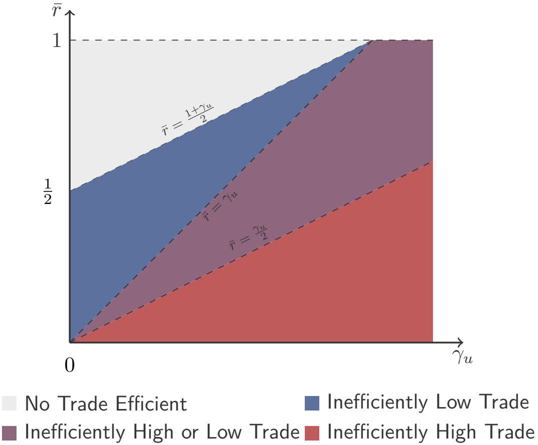
Example

Complete equilibrium characterization for this example:



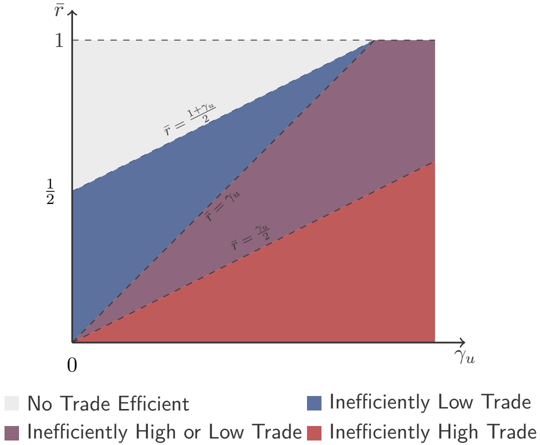
Example

Complete equilibrium characterization for this example:



Example

Complete equilibrium characterization for this example:



remedies

Remedies

How to fix this market failure?

We explore three alternative market designs:

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

data union

Data Unions

Recent policy proposals for the data economy (Posner-Weyl 18; Bergemann et al 23)

A data union would represent consumers by managing data on their behalf

We offer some theoretical support to these policy proposals

Data Union

How does a data union work?

- Consumers can participate in the union
- If they do, they relinquish their data to the union
- Union sells some of this data to the platform

  Consumers retain reservation utility unless record is sold to platform

- With the proceeds of sale, union compensates all participating consumers (to incentivize their participation)
- Union maximizes welfare of participating consumers

Data Union

Formally, the data union problem is:

$$
\begin{aligned}
\max_{(p,q,x)} \quad & \sum_{\omega} p(\omega)\bar{q}(\omega) + \sum_{a,\omega} u(a,\omega)x(a|\omega)q(\omega) + \sum_{\omega}(\bar{q}(\omega) - q(\omega))\bar{r} \\
\text{such that} \quad & q \leq \bar{q}, \\
\text{and} \quad & \textstyle\sum_{\omega} p(\omega)\bar{q}(\omega) = V(q), \\
\text{and} \quad & x \text{ solves } \mathcal{P}_q, \\
\text{and} \quad & p(\omega) + \tfrac{q(\omega)}{\bar{q}(\omega)} \textstyle\sum_{a} u(a,\omega)x(a|\omega) + \left(1 - \tfrac{q(\omega)}{\bar{q}(\omega)}\right)\bar{r} \geq \bar{r}.
\end{aligned}
$$

Data Union

Proposition

Equilibria of the data-union economy are constrained efficient and maximize consumers' welfare (and vice versa), regardless of the platform's objective

Some intuition:

Data union coordinates consumers by deciding which records to sell and how to compensate them

By doing so, data union acts as a substitute for the competitive market and avoids market failure

data taxes

Data Taxes

Enrich competitive economy by introducing a simple **data tax**:

- When selling her record, consumer pays tax $\tau(\omega) \in \mathbb{R}$ to the govt

When properly designed, data taxes force consumers to internalize effects that selling their records create on economy

Data Taxes

Enrich competitive economy by introducing a simple **data tax**:

- When selling her record, consumer pays tax $\tau(\omega) \in \mathbb{R}$ to the govt

When properly designed, data taxes force consumers to internalize effects that selling their records create on economy

Proposition

Let (q°, x°) be a constrained-efficient allocation. There exists a profile of taxes τ^*, of prices p^*, and of consumer choices ζ^*, such that $(p^*, \zeta^*, q^\circ, x^\circ)$ is an equilibrium of the economy with taxation τ^* and the government does not run a deficit.

Sketch of the Proof

Let allocation (q°, x°) be constrained efficient

Let p^* be a supergradient of $V(q^\circ)$

Define $\boxed{\tau^*(\omega) \triangleq p^*(\omega) + \sum_a x^\circ(a|\omega) u(a,\omega) - \bar{r}}$

Notice that $U^*(\omega) - \tau^*(\omega) \equiv \bar{r}$

Therefore, all consumers indifferent $\rightsquigarrow$ choose ζ^* to implement q° $\square$

more-complete markets

Lindahl Economy

We let price of data depend not only on its type (i.e., ω) but also on its "intended use" (i.e., a)

> Platform and the consumer trade on **how** record will be used—i.e., which fee a platform will recommend to the merchant

Lindahl Economy

We let price of data depend not only on its type (i.e., ω) but also on its "intended use" (i.e., a)

> Platform and the consumer trade on **how** record will be used—i.e., which fee a platform will recommend to the merchant

This adapts to our setting the standard approach for modeling economies with externalities (Arrow 1969, Laffont 1976)

Lindahl Economy

We let price of data depend not only on its type (i.e., ω) but also on its "intended use" (i.e., a)

> Platform and the consumer trade on **how** record will be used—i.e., which fee a platform will recommend to the merchant

This adapts to our setting the standard approach for modeling economies with externalities (Arrow 1969, Laffont 1976)

This is reminiscent of GDPR: "*The **specific purposes** for which personal data are used should be determined at the time of the collection*"

Lindahl Economy

A market for each (a, ω), where ω-records can be traded for use a at price $p(a, \omega)$

Our equilibrium definition extends naturally to this richer economy

In particular, timing is the same

Lindahl Economy

A market for each (a, ω), where ω-records can be traded for use a at price $p(a, \omega)$

Our equilibrium definition extends naturally to this richer economy

In particular, timing is the same

Proposition

Equilibria of this economy are (unconstrained) efficient and maximize consumers' welfare, regardless of platform's incentives

conclusion

Summary

1. A framework to study competitive markets for personal data

2. Identify novel inefficiency leading this otherwise perfectly competitive market to fail

   Show how inefficiency critically depends on platform's role as an information intermediary

3. Propose three alternative market designs that fix inefficiency: data unions, data taxes, richer data prices

Thank You!

Thank You!

The Merchant of Venice, Gilbert (1873)

Efficiency Benchmark

To illustrate market failure, a less demanding efficiency benchmark is desirable:

Efficiency Benchmark

To illustrate market failure, a less demanding efficiency benchmark is desirable:

1. We require x° to be optimal given q° **for the platform**

   If not, detect inefficiency driven by platform lack of commitment in period 1

   (main results extend to "unconstrained" efficiency)

Efficiency Benchmark

To illustrate market failure, a less demanding efficiency benchmark is desirable:

1. We require x° to be optimal given q° **for the platform**

   If not, detect inefficiency driven by platform lack of commitment in period 1

   (main results extend to "unconstrained" efficiency)

2. We exclude merchant's payoff from $W(q, x)$

   If not, detect inefficiency driven by platform not fully internalizing merchant's payoff (main results extend to "aggregate" welfare)

Efficiency Benchmark

To illustrate market failure, a less demanding efficiency benchmark is desirable:

1. We require x° to be optimal given q° **for the platform**

   If not, detect inefficiency driven by platform lack of commitment in period 1

   (main results extend to "unconstrained" efficiency)

2. We exclude merchant's payoff from $W(q, x)$

   If not, detect inefficiency driven by platform not fully internalizing merchant's payoff (main results extend to "aggregate" welfare)

Bonus: In eqm, platform makes not profits. Thus, $W(q^*, x^*)$ equals consumer welfare. Thus, any constrained-efficient eqm maximizes consumer welfare

back