The Value of Data Records

Simone Galperti
UC San Diego

Aleksandr Levkun
UC San Diego

Jacopo Perego
Columbia University

December 2021

Motivation

Personal data is an essential input in many modern industries

Markets for data are rapidly developing and raise new questions for policy and regulation (Federal Trade Commission, 2014, Stigler Report, 2019)

Our Goal: better understand **demand side** of data markets, how they work, and how to fairly compensate data sources (Lanier, 2013; Acquisti et al., 2016; Posner and Weyl, 2018)

Basic Question: What is the value of an individual piece of data?

> E.g.: for an e-commerce platform, is one consumer's data more valuable than another's? Why? How much should each be worth/paid?

example

An Illustrative Example

(builds on BBM ’15)

A monopolistic product **firm** sells its product through an e-commerce **platform**

Platform is used by a population of **buyers**, each with independent valuation for the product

An Illustrative Example

(builds on BBM '15)

A monopolistic product **firm** sells its product through an e-commerce **platform**

Platform is used by a population of **buyers**, each with independent valuation for the product

For each buyer, platform observes a **record** of her personal characteristics, which is informative about her valuation

An Illustrative Example

(builds on BBM '15)

A monopolistic product **firm** sells its product through an e-commerce **platform**

Platform is used by a population of **buyers**, each with independent valuation for the product

For each buyer, platform observes a **record** of her personal characteristics, which is informative about her valuation

Suppose there are only two *types* of such records and that

- ω_1 reveals buyer has valuation 1
- ω_2 reveals buyer has valuation 2

Platform's **database** of records composed of $q(\omega_1) = 3M$ and $q(\omega_2) = 6M$

An Illustrative Example

Only platform observes ω; seller only knows database composition

Platform intermediates buyer-seller interaction by giving seller information about ω so as to influence her price p

Seller maximizes profits (MC = 0)

Suppose platform maximizes buyer's surplus

Question. What is the value of a type-ω record for the platform?

An Illustrative Example

An optimal way to use the database the following:

$$\text{Record of type } \omega_1 \longrightarrow s^L$$

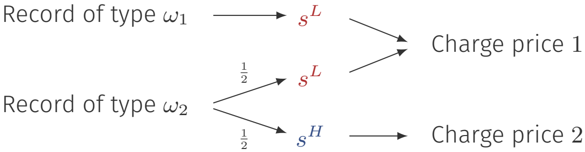
$$\begin{tikzcd}
 & s^L \\
\text{Record of type } \omega_2 \arrow[ur, "\frac{1}{2}"] \arrow[dr, "\frac{1}{2}"'] & \\
 & s^H
\end{tikzcd}$$

An Illustrative Example

An optimal way to use the database the following:

$$\begin{tikzcd}
\text{Record of type } \omega_1 \arrow[r] & s^L \arrow[dr] & \\
 & & \text{Charge price } 1 \\
 & s^L \arrow[ur] & \\
\text{Record of type } \omega_2 \arrow[ur, "\frac{1}{2}"] \arrow[dr, "\frac{1}{2}"'] & & \\
 & s^H \arrow[r] & \text{Charge price } 2
\end{tikzcd}$$

An optimal way to use the database the following:



Thus, platform's **direct payoff** is $u_q^*(\omega) = \begin{cases} 0 & \text{for } \omega_1 \\ \frac{1}{2} & \text{for } \omega_2 \end{cases}$

An Illustrative Example

An optimal way to use the database the following:

Record of type $\omega_1 \longrightarrow s^L \longrightarrow$ Charge price 1

Record of type $\omega_2 \xrightarrow{\frac{1}{2}} s^L \longrightarrow$ Charge price 1

Record of type $\omega_2 \xrightarrow{\frac{1}{2}} s^H \longrightarrow$ Charge price 2

Thus, platform's **direct payoff** is $u_q^*(\omega) = \begin{cases} 0 & \text{for } \omega_1 \\ \frac{1}{2} & \text{for } \omega_2 \end{cases}$

Are type-ω_1 records really worthless?

An Illustrative Example

An optimal way to use the database the following:

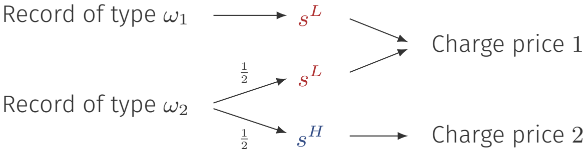
$$
\begin{tikzcd}
\text{Record of type } \omega_1 \arrow[r] & s^L \arrow[dr] & \\
& & \text{Charge price } 1 \\
& s^L \arrow[ur] & \\
\text{Record of type } \omega_2 \arrow[ur, "\frac{1}{2}"] \arrow[dr, "\frac{1}{2}"'] & & \\
& s^H \arrow[r] & \text{Charge price } 2
\end{tikzcd}
$$

Thus, platform's **direct payoff** is $u_q^*(\omega) = \begin{cases} 0 & \text{for } \omega_1 \\ \frac{1}{2} & \text{for } \omega_2 \end{cases}$

Are type-ω_1 records really worthless? No, in fact $v_q^*(\omega) = \begin{cases} 1 & \text{for } \omega_1 \\ 0 & \text{for } \omega_2 \end{cases}$

An Illustrative Example

An optimal way to use the database the following:



Thus, platform's **direct payoff** is $u_q^*(\omega) = \begin{cases} 0 & \text{for } \omega_1 \\ \frac{1}{2} & \text{for } \omega_2 \end{cases}$

Are type-ω_1 records really worthless? No, in fact $v_q^*(\omega) = \begin{cases} 1 & \text{for } \omega_1 \\ 0 & \text{for } \omega_2 \end{cases}$

1. Most valuable records are those yielding lowest payoff
2. ω_1 generates no direct payoff but "helps" ω_2 earn positive surplus
3. Payoff u_q^* gives biased account of the value created by a record

this paper

Questions

We address two main questions

1. What is the value of each record and what are its properties?

   $\rightsquigarrow$ a *benchmark* for compensating buyer with “data dividend”

Questions

We address two main questions

1. What is the value of each record and what are its properties?

   $\rightsquigarrow$ a *benchmark* for compensating buyer with “data dividend”

2. What is the platform’s willingness to pay for “more data”?

Questions

We address two main questions

1. What is the value of each record and what are its properties?

   $\rightsquigarrow$ a *benchmark* for compensating buyer with “data dividend”

2. What is the platform’s willingness to pay for “more data”?

– For *more* records

  E.g., Platform obtains access to *new* buyers

Questions

We address two main questions

1. What is the value of each record and what are its properties?

   $\rightsquigarrow$ a *benchmark* for compensating buyer with “data dividend”

2. What is the platform’s willingness to pay for “more data”?

– For *more* records

  E.g., Platform obtains access to *new* buyers

– For *better* records

  E.g., Platform obtains more information about *existing* buyers

Questions

We address two main questions

1. What is the value of each record and what are its properties?

   $\rightsquigarrow$ a *benchmark* for compensating buyer with "data dividend"

2. What is the platform's willingness to pay for "more data"?

   – For *more* records — Marketing Lists

     E.g., Platform obtains access to *new* buyers

   – For *better* records — Data Appends

     E.g., Platform obtains more information about *existing* buyers

Related Literature

Two strands of the literature on **data markets** Bergemann and Ottaviani (21)

1. Use of a database — how to design and sell information e.g., Admati and Pfleiderer (86, 90), Bergemann and Bonatti (15), Bergemann et al. (18), Yang (20)
   - Focus is not on use; but on inputs to the database (upstream market)
   - Features of *data records*: access + information; valued ex post

Related Literature

Two strands of the literature on **data markets** Bergemann and Ottaviani (21)

1. Use of a database — how to design and sell information e.g., Admati and Pfleiderer (86, 90), Bergemann and Bonatti (15), Bergemann et al. (18), Yang (20)

   - Focus is not on use; but on inputs to the database (upstream market)
   - Features of *data records*: access + information; valued ex post

2. Consumers' incentives to disclose — social dimension of data, learning externalities Choi et al. (19), Bergemann et al. (20), Acemoglu et al. (21), and Ichihashi (21)

   - Ann's record uninformative about Bob's $\rightsquigarrow$ new data externality
   - No disclosure, platform already owns the database

See paper for connections to mechanism/information design literature and duality model

Model

For this talk, the model simply generalizes the example

Model

For this talk, the model simply generalizes the example

Denote **platform** by $i = 0$; denote **seller** by $i = 1$ and his action $a \in A$

Buyer's valuation θ for seller's product is i.i.d

Model

For this talk, the model simply generalizes the example

Denote **platform** by $i = 0$; denote **seller** by $i = 1$ and his action $a \in A$

Buyer's valuation θ for seller's product is i.i.d

A buyer's record is of type $\omega \in \Omega$ and is (partially) informative about her θ

Database composition $q \in \mathbb{R}_{+}^{\Omega}$ is common knowledge

For $i \in \{0, 1\}$, $u_i : A \times \Omega \to \mathbb{R}$ denotes i's expected payoff function

Model

Platform acts as an information designer, wlog focus on $x : \Omega \to \Delta(A)$

Model

Platform acts as an information designer, wlog focus on $x : \Omega \to \Delta(A)$

Information-Design Problem $\mathcal{U}_q$: Bergemann and Morris '16

$$\max_x \sum_{\omega, a} u_0(a,\omega) x(a|\omega) q(\omega)$$

s.t. for all a, a',

$$\sum_{\omega} \Big(u_1(a,\omega) - u_1(a',\omega) \Big) x(a|\omega) q(\omega) \geq 0$$

Model

Platform acts as an information designer, wlog focus on $x : \Omega \to \Delta(A)$

Information-Design Problem $\mathcal{U}_q$: Bergemann and Morris '16

$$\max_x \sum_{\omega,a} u_0(a,\omega)x(a|\omega)q(\omega)$$

s.t. for all a, a',

$$\sum_\omega \Big(u_1(a,\omega) - u_1(a',\omega)\Big)x(a|\omega)q(\omega) \geq 0$$

Letting x_q^* be a solution of $\mathcal{U}_q$, denote

- $u_q^*(\omega) \triangleq \sum_a x_q^*(a|\omega)u_0(a,\omega)$ the **direct payoff** of a record
- $U^*(q) \triangleq \sum_\omega u_q^*(\omega)q(\omega)$ the **total payoff** of the database

the value of a single data record

The Value Problem

In $\mathcal{U}_q$, platform uses records as **inputs** to produce **outputs**

Use LP duality to determine **unit value** of each input Dorfman et al. '87, Gale '89

In $\mathcal{U}_q$, platform uses records as **inputs** to produce **outputs**

Use LP duality to determine **unit value** of each input Dorfman et al. '87, Gale '89

Let $v : \Omega \to \mathbb{R}$ and $\lambda : A \times A \to \mathbb{R}_+$

For every (a, ω), let $t(a, \omega) \triangleq \sum_{a' \in A} \Big(u_1(a, \omega) - u_1(a', \omega) \Big) \lambda(a' | a)$

The Value Problem

In $\mathcal{U}_q$, platform uses records as **inputs** to produce **outputs**

Use LP duality to determine **unit value** of each input Dorfman et al. '87, Gale '89

Let $v : \Omega \to \mathbb{R}$ and $\lambda : A \times A \to \mathbb{R}_+$

For every (a, ω), let $t(a, \omega) \triangleq \sum_{a' \in A} \Big(u_1(a, \omega) - u_1(a', \omega) \Big) \lambda(a'|a)$

Value Problem:

$$\mathcal{V}_q : \qquad \min_{v, \lambda} \sum_{\omega} v(\omega) q(\omega)$$

s.t. for all $\omega \in \Omega$,

$$v(\omega) = \max_{a \in A} \Big\{ u_0(a, \omega) + t(a, \omega) \Big\}$$

Lemma (Duality)

$\mathcal{V}_q$ is the dual of $\mathcal{U}_q$.

For every optimal solution v_q^* and x_q^*,

$$\sum_{\omega\in\Omega} v_q^*(\omega)q(\omega) = U^*(q) \triangleq \sum_{\omega,a} u_q^*(\omega)q(\omega)$$

Lemma (Duality)

$\mathcal{V}_q$ is the dual of $\mathcal{U}_q$.

For every optimal solution v_q^* and x_q^*,

$$\sum_{\omega \in \Omega} v_q^*(\omega) q(\omega) = U^*(q) \triangleq \sum_{\omega, a} u_q^*(\omega) q(\omega)$$

Basic Step

Lemma (Duality)

$\mathcal{V}_q$ is the dual of $\mathcal{U}_q$.

For every optimal solution v_q^* and x_q^*,

$$\sum_{\omega \in \Omega} v_q^*(\omega) q(\omega) = U^*(q) \triangleq \sum_{\omega, a} u_q^*(\omega) q(\omega)$$

- Formally, $v_q^*(\omega)$ is multiplier of feasibility constraint, thus $v_q^*(\omega)$ captures effect of marginal change in $q(\omega)$ on $U^*(q)$
- $v_q^*(\omega)$ is the **unit value** of a record of type ω (Gale '89)
- The goal of the paper is to characterize properties of $v_q^*(\omega)$

Decomposing the Value of a Record

What determines the value of a record?

What determines the value of a record?

Proposition (Decomposition)

The value of a record ω can be decomposed as

$$v_q^*(\omega) = u_q^*(\omega) + t_q^*(\omega) \qquad \text{where:}$$

What determines the value of a record?

Proposition (Decomposition)

The value of a record ω can be decomposed as

$$v_q^*(\omega) = u_q^*(\omega) + t_q^*(\omega) \qquad \text{where:}$$

$$u_q^*(\omega) \triangleq \sum_a u_0(a,\omega) x_q^*(a|\omega) \qquad \text{direct payoff}$$

- $u_q^*(\omega)$ captures the payoff that platform earns directly from record

Decomposing the Value of a Record

What determines the value of a record?

Proposition (Decomposition)

The value of a record ω can be decomposed as

$$v_q^*(\omega) = u_q^*(\omega) + t_q^*(\omega) \qquad \text{where:}$$

$$u_q^*(\omega) \triangleq \sum_a u_0(a,\omega) x_q^*(a|\omega) \qquad \textbf{direct payoff}$$

$$t_q^*(\omega) \triangleq \sum_a t^*(a,\omega) x^*(a|\omega) \overset{\text{a.e.}}{=} \sum_{\omega'} q(\omega') \tfrac{\partial}{\partial q(\omega)} u_q^*(\omega') \qquad \textbf{externality}$$

- $t_q^*(\omega)$ **externality** that ω exterts on payoffs generated by other records

Decomposing the Value of a Record

What determines the value of a record?

Proposition (Decomposition)

The value of a record ω can be decomposed as

$$v_q^*(\omega) = u_q^*(\omega) + t_q^*(\omega) \qquad \text{where:}$$

$$u_q^*(\omega) \triangleq \sum_a u_0(a,\omega) x_q^*(a|\omega) \qquad \textbf{direct payoff}$$

$$t_q^*(\omega) \triangleq \sum_a t^*(a,\omega) x^*(a|\omega) \overset{\text{a.e.}}{=} \sum_{\omega'} q(\omega') \frac{\partial}{\partial q(\omega)} u_q^*(\omega') \qquad \textbf{externality}$$

- $t_q^*(\omega)$ **externality** that ω exterts on payoffs generated by other records
- Clarfies why $u_q^*(\omega)$ is biased measure of value

Externality

Externalities arise because platform may withhold information by pooling data records together

- Specific to **intermediation problems**, as opposed to decision problems

  Ubiquitous due to rise of "info-mediaries" Acquisti et al. (16)

- They arise even if records are statistically independent between buyers

  Unrelated to "learning" externalities in literature (e.g. Acemoglu et al. 2021, Bergemann et al. 2020, Ichihashi (2021))

Example

$\Omega = \{\omega_1, \dots, \omega_K\}$ and record of type ω_k fully reveals that $\theta = \omega_k$

Platform maximizes

$$u_0(a, \omega) = \pi\Big(\underbrace{a\mathbb{1}\{\omega \geq a\}}_{\text{seller's profit}}\Big) + (1-\pi)\Big(\underbrace{\max\{\omega - a, 0\}}_{\text{buyer's surplus}}\Big) \qquad \text{for } \pi \in [0, 1]$$

Notation: a_q = uniform monopoly price

Proposition

Suppose $\pi < 1/2$.

- $t_q^*(\omega) > 0$ if $\omega < a_q$ and $t_q^*(\omega) \leq 0$ if $\omega \geq a_q$

Example

Proposition

Suppose $\pi < 1/2$.

– $t_q^*(\omega) > 0$ if $\omega < a_q$ and $t_q^*(\omega) \leq 0$ if $\omega \geq a_q$

If $\pi < 1/2$, u_q^* provides biased account of the value of each record

Ignoring externality may lead to:

- Underpay low-valuation buyers for their data $u_q^*(\omega) < v_q^*(\omega)$
- Overpay high-valuation buyers for their data $u_q^*(\omega) > v_q^*(\omega)$

Example

Proposition

Suppose $\pi < 1/2$.

- $t_q^*(\omega) > 0$ if $\omega < a_q$ and $t_q^*(\omega) \leq 0$ if $\omega \geq a_q$

Suppose $\pi \geq 1/2$.

- $t_q^*(\omega) = 0$ for all ω

If $\pi < 1/2$, u_q^* provides biased account of the value of each record

Ignoring externality may lead to:

- Underpay low-valuation buyers for their data $u_q^*(\omega) < v_q^*(\omega)$
- Overpay high-valuation buyers for their data $u_q^*(\omega) > v_q^*(\omega)$

changing the database

Having More Data?

What is the platform's **willingness to pay** for more data?

We study two cases ($\approx$ kinds of information products):

1. The platform obtains *more* records
2. The platform obtains *better* records

We can study both problems by exploring how v_q^* depends on q

Having More Data?

What is the platform's **willingness to pay** for more data?

We study two cases ($\approx$ kinds of information products):

1. **The platform obtains *more* records**
2. The platform obtains *better* records

We can study both problems by exploring how v_q^* depends on q

Platform as a consumer, WTP revealed by marginal “utility,” $v_q^*(\omega)$

$v_q^*(\omega)$ is step-wise diminishing (a textbook consumer)

The Demand for Data Records

Platform as a consumer, WTP revealed by marginal "utility," $v_q^*(\omega)$

$v_q^*(\omega)$ is step-wise diminishing (a textbook consumer)

Define frequency of a record type as $\mu_q(\omega) \triangleq \dfrac{q(\omega)}{\sum_{\omega'} q(\omega')}$

Platform as a consumer, WTP revealed by marginal "utility," $v_q^*(\omega)$

$v_q^*(\omega)$ is step-wise diminishing (a textbook consumer)

Define frequency of a record type as $\mu_q(\omega) \triangleq \dfrac{q(\omega)}{\sum_{\omega'} q(\omega')}$

1. Proposition (Scarcity Principle)

Fix q' and q. If $\mu_{q'}(\omega) > \mu_q(\omega)$, then $v_{q'}^*(\omega) \leq v_q^*(\omega)$.

The Demand for Data Records

Platform as a consumer, WTP revealed by marginal "utility," $v_q^*(\omega)$

$v_q^*(\omega)$ is step-wise diminishing (a textbook consumer)

Define frequency of a record type as $\mu_q(\omega) \triangleq \dfrac{q(\omega)}{\sum_{\omega'} q(\omega')}$

1. Proposition (Scarcity Principle)

Fix q' and q. If $\mu_{q'}(\omega) > \mu_q(\omega)$, then $v_{q'}^*(\omega) \leq v_q^*(\omega)$.

When $\mu_q(\omega)$ sufficiently large, $v_q^*(\omega) = \bar{u}(\omega) \rightsquigarrow$ payoff under full-disclosure

Platform as a consumer, WTP revealed by marginal “utility,” $v_q^*(\omega)$

$v_q^*(\omega)$ is step-wise diminishing (a textbook consumer)

Define frequency of a record type as $\mu_q(\omega) \triangleq \dfrac{q(\omega)}{\sum_{\omega'} q(\omega')}$

1. Proposition (Scarcity Principle)

Fix q' and q. If $\mu_{q'}(\omega) > \mu_q(\omega)$, then $v_{q'}^*(\omega) \leq v_q^*(\omega)$.

When $\mu_q(\omega)$ sufficiently large, $v_q^*(\omega) = \bar{u}(\omega) \rightsquigarrow$ payoff under full-disclosure

2. Proposition (Locally Constant)

There is finite collection $\{Q_1, \ldots, Q_N\} \subseteq \mathbb{R}_+^\Omega$ of open, convex, disjoint sets s.t. $\bigcup Q_n$ has full measure and v_q^* is **constant** in $q \in Q_n$ for each n

The Demand for Data Records

Platform as a consumer, WTP revealed by marginal "utility," $v_q^*(\omega)$

$v_q^*(\omega)$ is step-wise diminishing (a textbook consumer)

Define frequency of a record type as $\mu_q(\omega) \triangleq \dfrac{q(\omega)}{\sum_{\omega'} q(\omega')}$

1. Proposition (Scarcity Principle)

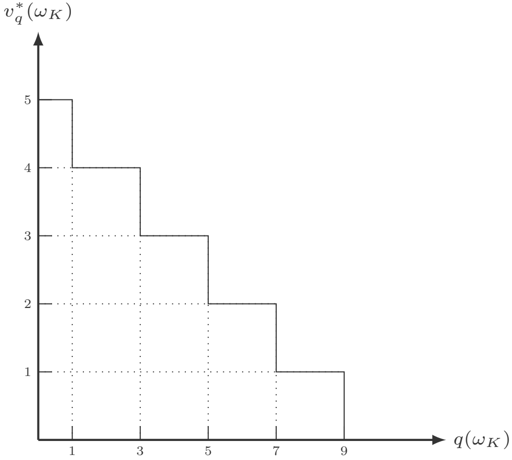
Fix q' and q. If $\mu_{q'}(\omega) > \mu_q(\omega)$, then $v_{q'}^*(\omega) \leq v_q^*(\omega)$.

When $\mu_q(\omega)$ sufficiently large, $v_q^*(\omega) = \bar{u}(\omega) \rightsquigarrow$ payoff under full-disclosure

2. Proposition (Locally Constant)

There is finite collection $\{Q_1, \ldots, Q_N\} \subseteq \mathbb{R}_+^\Omega$ of open, convex, disjoint sets s.t. $\bigcup Q_n$ has full measure and v_q^* is **constant** in $q \in Q_n$ for each n

An example of a **demand curve** for records of type ω_K



Fix $\pi = 0$, $K = 10$, $q(\omega_k) = 1$, $\forall k < K$

Data records exhibits complementarities iff platform withholds information

Proposition

Records are **perfect** substitutes iff for **some** (interior) q it is optimal to **fully disclose** every record. In this case, full disclosure is optimal for every q.

WTP for More Records

Data records exhibits complementarities iff platform withholds information

Proposition

Records are **perfect** substitutes iff for **some** (interior) q it is optimal to **fully disclose** every record. In this case, full disclosure is optimal for every q.

Data records exhibits complementarities iff platform withholds information

Proposition

Records are **perfect** substitutes iff for **some** (interior) q it is optimal to **fully disclose** every record. In this case, full disclosure is optimal for every q.

Two implications:

1. Optimal database: Well-behaved problem, interior solution $\rightsquigarrow$ Standard demand analysis

Data records exhibits complementarities iff platform withholds information

Proposition

Records are **perfect** substitutes iff for **some** (interior) q it is optimal to **fully disclose** every record. In this case, full disclosure is optimal for every q.

Two implications:

1. Optimal database: Well-behaved problem, interior solution $\rightsquigarrow$ Standard demand analysis
2. Revealed preferences:
   - By detecting imperfect substitutability at q, we can infer platform withholds information from the seller

WTP for More Records

Data records exhibits complementarities iff platform withholds information

Proposition

Records are **perfect** substitutes iff for **some** (interior) q it is optimal to **fully disclose** every record. In this case, full disclosure is optimal for every q.

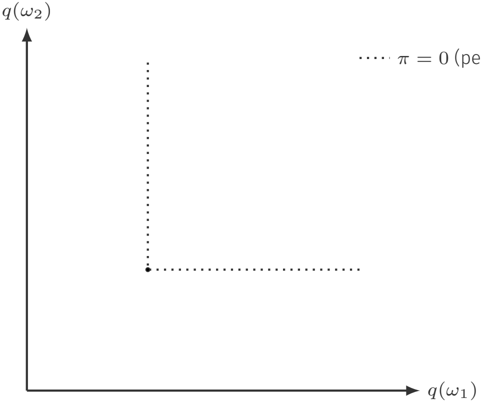
Two implications:

1. Optimal database: Well-behaved problem, interior solution $\rightsquigarrow$ Standard demand analysis
2. Revealed preferences:
   - By detecting imperfect substitutability at q, we can infer platform withholds information from the seller

An example of platform's **indifference curves**

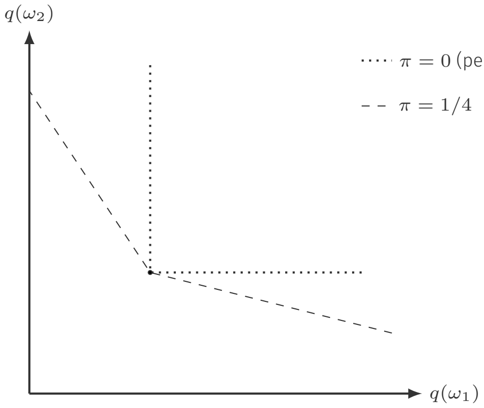
Set $K = 2$

An example of platform's **indifference curves**



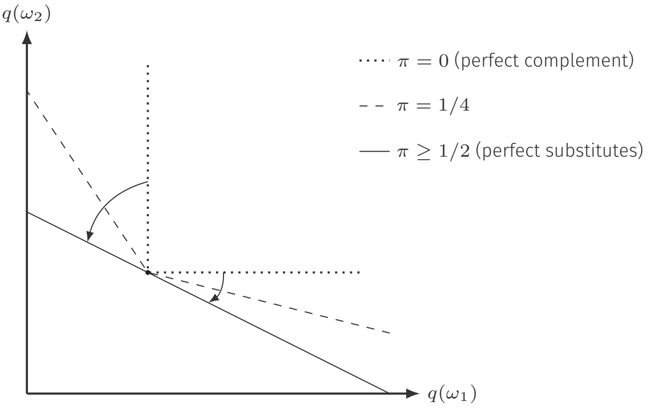
Set $K = 2$

An example of platform's **indifference curves**



Set $K = 2$

An example of platform's **indifference curves**



Set $K = 2$

Having More Data?

What is the platform's **willingness to pay** for more data?

We study two cases ($\approx$ kinds of information products):

1. The platform acquires *more* records
2. **The platform obtains *better* records**

We can study both problems by exploring how v_q^* depends on q

Refining Existing Records

Records are often only partially informative about θ and platform can **learn** more about them $\rightsquigarrow$ **refine records**

Example

$$\begin{tikzcd} & \omega_1 \\ \omega \arrow[ur, "\text{w.p. } \sigma_\omega(\omega_1)"] \arrow[dr, "\text{w.p. } \sigma_\omega(\omega_2)"'] & \\ & \omega_2 \end{tikzcd}$$

(gender, $\varnothing$) (gender, education)

Refining Existing Records

Records are often only partially informative about θ and platform can **learn** more about them $\rightsquigarrow$ **refine records**

Example

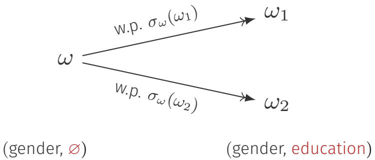
$$\begin{tikzcd} & \omega_1 \\ \omega \arrow[ur, "\text{w.p. } \sigma_\omega(\omega_1)"] \arrow[dr, "\text{w.p. } \sigma_\omega(\omega_2)"'] & \\ & \omega_2 \end{tikzcd}$$

(gender, $\varnothing$) (gender, education)

Refining Existing Records

Records are often only partially informative about θ and platform can **learn** more about them $\rightsquigarrow$ **refine records**

Example



Definition. A **refinement** independently refines a share α of type-ω records wrt σ_ω

It transforms the original database $q \rightsquigarrow q_\alpha$ such that:

$$q_\alpha(\omega) < q(\omega) \qquad \text{and} \qquad q_\alpha(\omega') > q(\omega') \qquad \forall \omega' \in \operatorname{supp} \sigma_\omega$$

How do refinements change the value derived from *each* record?

How do refinements change the value derived from *each* record?

Corollary:

Direct Effects. The value of each *refined* record increases:

$$\sum_{\omega' \in \Omega} v^*_{q_\alpha}(\omega')\sigma_\omega(\omega') \geq v^*_q(\omega)$$

Indirect Effects. The value of *unrefined* records changes as well

$$v^*_{q_\alpha}(\omega) \geq v^*_q(\omega) \qquad \text{and} \qquad v^*_{q_\alpha}(\omega') \leq v^*_q(\omega') \qquad \forall \omega' \in \operatorname{supp} \sigma_\omega$$

How do refinements change the value derived from *each* record?

Corollary:

Direct Effects. The value of each *refined* record increases:

$$\sum_{\omega' \in \Omega} v^*_{q_\alpha}(\omega')\sigma_\omega(\omega') \geq v^*_q(\omega)$$

Indirect Effects. The value of *unrefined* records changes as well

$$v^*_{q_\alpha}(\omega) \geq v^*_q(\omega) \qquad \text{and} \qquad v^*_{q_\alpha}(\omega') \leq v^*_q(\omega') \qquad \forall \omega' \in \operatorname{supp} \sigma_\omega$$

Given these mixed effects, do refinements benefit platform overall?

Given these mixed effects, do refinements benefit platform overall?

Proposition

The platform's benefit from the refinement is:

- Weakly **positive**, $U^*(q_\alpha) \geq U^*(q)$

Yes. WTP for a refinement is positive

Given these mixed effects, do refinements benefit platform overall?

Proposition

The platform's benefit from the refinement is:

– Weakly **positive**, $U^*(q_\alpha) \geq U^*(q)$

Yes. WTP for a refinement is positive

However, WTP can be 0 even if platform acts on new information (x_q^* changes)

- In sharp contrast with decision problems

Given these mixed effects, do refinements benefit platform overall?

Proposition

The platform's benefit from the refinement is:

- Weakly **positive**, $U^*(q_\alpha) \geq U^*(q)$
- **Zero** for all α iff there is $a \in \operatorname{supp} x_q^*(\cdot|\omega'')$ for $\omega'' = \omega$ & $\omega'' \in \operatorname{supp} \sigma_\omega$

Yes. WTP for a refinement is positive

However, WTP can be 0 even if platform acts on new information (x_q^* changes)

- In sharp contrast with decision problems

WTP for Better Records

Given these mixed effects, do refinements benefit platform overall?

Proposition

The platform's benefit from the refinement is:

- Weakly **positive**, $U^*(q_\alpha) \geq U^*(q)$
- **Zero** for all α iff there is $a \in \operatorname{supp} x_q^*(\cdot|\omega'')$ for $\omega'' = \omega$ & $\omega'' \in \operatorname{supp} \sigma_\omega$
- Marginally **decreasing** in α

Yes. WTP for a refinement is positive

However, WTP can be 0 even if platform acts on new information (x_q^* changes)

- In sharp contrast with decision problems

summary

Summary

We characterize **unit value** of a buyer's specific data record using duality

- Direct payoff from a record gives biased account of its value
- Novel data externalities, platform withholds information by pooling records

We study the platform's WTP for more data

- *"More"* records: demand for records, complements vs substitutes
- *"Better"* records: Refinements, effects on values, WTP $= 0$

Overall, a study of the **demand side** of data markets

$\rightsquigarrow$ Welfare effects of policy intervention; demand estimation

Future Directions: Privacy

With the same framework, we can study effects of **privacy** on the value of data

In a richer version of today's model:

- Buyer is an agent $(i = 2)$ and ω is her **private** information
- Platform must elicit such information in IC ways in order to use it

Still a LP problem, same approach as today applies.

Preliminary findings show that

- Privacy decreases the overall value of the database (of course!)
- But it can **increase** the value of some of types of record

How does v_q^* depend on q?

Stability

How does v_q^* depend on q?

v_q^* goes beyond a *marginal* interpretation $\rightsquigarrow$ WTP for discrete changes in q

Proposition (Stability)

There exists finite collection $\{Q_1, \ldots, Q_K\}$ of open sets in $\mathbb{R}_+^\Omega$ s.t.:

- $\bigcup Q_k$ has full measure
- (v_q^*, λ_q^*) is **unique** and **constant** in $q \in Q_k$ for each k

How does v_q^* depend on q?

v_q^* goes beyond a *marginal* interpretation $\rightsquigarrow$ WTP for discrete changes in q

Proposition (Stability)

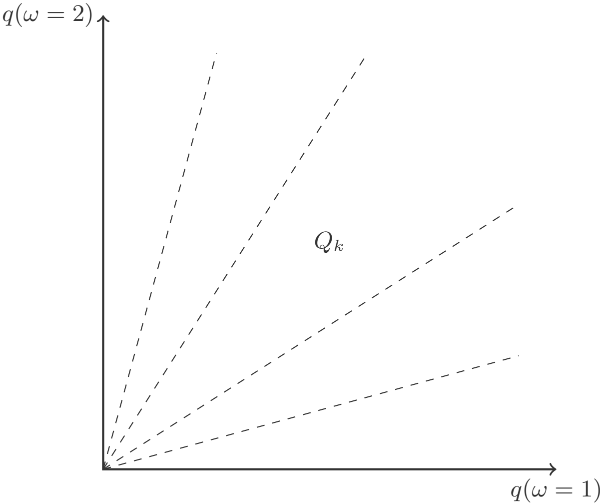
There exists finite collection $\{Q_1, \ldots, Q_K\}$ of open sets in $\mathbb{R}_+^\Omega$ s.t.:

- $\bigcup Q_k$ has full measure
- (v_q^*, λ_q^*) is **unique** and **constant** in $q \in Q_k$ for each k

Note : v_q^* constant in Q_k even though x_q^* changes

Proof idea : algebraic representation of extreme points & optimality

Example



Proposition

For $\pi \leq \frac{1}{2}$,

$$v_q^*(\omega) = \begin{cases} (1-\pi)\omega & \text{if } \omega < a_q \\ \pi a_q + (1-\pi)(\omega - a_q) & \text{if } \omega \geq a_q; \end{cases}$$

Moreover, $t_q^*(\omega) > 0$ for $\omega < a_q$ and $t_q^*(\omega) \leq 0$ for $\omega \geq a_q$

For $\pi \geq \frac{1}{2}$ we have $v_q^*(\omega) = u_q^*(\omega) = \pi\omega$ for all ω