The Price of Data

Simone Galperti
UC San Diego

Aleksandr Levkun
UC San Diego

Jacopo Perego
Columbia University

October 2020

Motivation

Data has become essential **input** in modern economies

Few formal markets for data; often data collected “for free” (Posner-Weyl ’18)

Question: what is the **individual** value of a datapoint? $\rightarrow$ **price**

- value that **each** datapoint in database **individually** generates for its owner? $\rightsquigarrow$ WTP for additional datapoint
- **drivers** of prices?
- effects of **privacy** concerns?
- compensating data **sources** for their data?

This Paper

Simple insight:

- **data pricing** problem intimately related to how owner **uses** data, given objective
  - combine data as inputs to *produce* actionable *information*
  - to **make** own decisions or to **influence** others' decisions

  $\Rightarrow$ **data usage: mechanism**/**information** design problem

- when carefully formulated, pricing and usage problems are in a special mathematical relationship: **duals**

Goals for today

1. formalize data usage–pricing relationship + novel interpretation
2. (preliminary) characterization of price determinants and properties
3. showcase properties through examples

Related Literature

	This Paper
Mechanism Design. Myerson ('82, '83) ...	– formulation of data usage
Information Design. Kamenica & Gentzkow ('11), Bergemann & Morris ('16,'19) ...	– subclass of data usage
Duality & Correlated Equilibrium. Nau & McCardle ('90), Nau ('92), Hart & Schmeidler ('89), Myerson ('97)	– duality to characterize CE – feasible mechanisms for principal
Duality & Bayesian Persuasion. Kolotilin ('18), Dworczak & Martini ('19), Dizdar & Kovac ('19), Dworczak & Kolotilin ('19)	– dual not as a solution method, but as focus of analysis – independent economic question – games, mechanisms
Markets for Information. Bergemann & Bonatti ('15), Bergmann, Bonatti, Smolin ('18), Posner & Weyl ('18), Bergemann & Bonatti ('19)	– individual prices of data
Information Privacy. Acquisti, Taylor, Wagman ('16), Ali, Lewis, Vasserman ('20), Bergemann, Bonatti, Gan ('20), Acemoglu, Makhdoumi, Malekian, Ozdaglar, ('20)	– formal method for assessing effects of privacy on value of data

illustrative example

Internet Platform (Bergemann et al. '15)

Internet platform owns **data** (cookies) about each potential buyer of product of monopolistic seller (MC=0)

Database: big list (continuum) of datapoints = buyer ID and valuation

- share μ of datapoints has valuation $\omega_0 = 1$
- share $1 - \mu$ of datapoints has valuation $\omega_0 = 2$

Platform **mediates interaction** between each buyer and seller:

- bins buyers into market segments (information production)
- discloses segments to seller for setting price a
- objective: maximize buyers' surplus

Internet Platform

Questions: what price $p(\omega_0)$

- would capture **individual value** that ω_0-datapoint has for platform?
- would/should platform be **willing to spend** to add one datapoint with valuation ω_0 to database?

Broadly refer to these questions as **data-pricing problem**

$p(\omega_0)$ **not** interpreted as monetary transfer to buyers for their data

- important, yet distinct issue (later)

Internet Platform

Given optimal segmentation, let $v^*(\omega_0)$ be **realized** surplus of ω_0-buyer

Question: does it make sense to set $p(\omega_0) = v^*(\omega_0)$?

Extreme cases: $\mu = 1 \Rightarrow v^*(1) = 0$ and $\mu = 0 \Rightarrow v^*(2) = 0$

If $\mu \in (0, 0.5)$, optimal market segmentation

	s'	s''	$v^*(\omega_0)$
$\omega_0 = 1$	1	0	0
$\omega_0 = 2$	$\frac{\mu}{1-\mu}$	$1 - \frac{\mu}{1-\mu}$	$\frac{\mu}{1-\mu}$
$\to a(s)$	1	2	

Idea: 1-buyers 'help' platform achieve positive surplus with **some** 2-buyers

Punchline: v^* misses this, so not good measure for $p(\omega_0)$

Internet Platform

If $\mu \in (0.5, 1)$, optimal market segmentation

	s'	s''	$v^*(\omega_0)$
$\omega_0 = 1$	1	0	0
$\omega_0 = 2$	$\frac{\mu}{1-\mu}$	$1 - \frac{\mu}{1-\mu}$	$\frac{\mu}{1-\mu}$
$\rightarrow$ price	1	2	

Idea: 1-buyers 'help' platform achieve positive surplus with **some** 2-buyers

Our approach will yield $p^*(1) = 1 > v^*(1)$ and $p^*(2) = 0 < v^*(2)$

- 1-datapoints useful $\rightsquigarrow$ induce seller to set **suboptimal** price for **2**-buyers
- 1-datapoints **scarce** 'input' in database ($\mu < 0.5$)

model

Overview

Principal (she) mediates economic interaction between group of agents (he) — e.g., buyer-seller trade

$\rightsquigarrow$ general formulation : **Bayes incentive problem** á la Myerson ('82,'83)

Each interaction characterized by **data** — e.g., buyer's valuation

Principal **uses** data to mediate interaction — e.g., segmentation

Question: what is **value** for principal of **individual** data characterizing each interaction she can mediate?

Standard Primitives

Parties: principal $i = 0$, agents $i \in I = \{1, \ldots, n\}$

Action privately controlled by party i: $a_i \in A_i$

$\rightsquigarrow A = A_0 \times \cdots \times A_n$

Piece of data privately and directly accessed by party i: $\omega_i \in \Omega_i$

$\rightsquigarrow \Omega = \Omega_0 \times \cdots \times \Omega_n$

Payoff function of party i: $u_i : A \times \Omega \to \mathbb{R}$

$\Rightarrow$ every $\omega = (\omega_0, \ldots, \omega_n)$ pins down one **type** of economic interaction the principal can mediate

Letting $\mu \in \Delta(\Omega)$, assume $\Gamma = (I, (\Omega, \mu), (A_i, u_i)_{i=0}^n)$ is common knowledge

Principal as Data User

Myerson's principal can **commit** to mediating interaction by

- **eliciting** agents' private data
- setting **rules/incentives** agents face: A_0 (mechanism)
- sending **signals** to affect agents' private actions: A_i (information)

As usual, focus on direct mechanisms $x : \Omega \to \Delta(A)$ that satisfy IC

- **honesty**: optimal for each agent to report ω_i truthfully
- **obedience**: optimal for each agent to follow recommended a_i

$\Rightarrow$ **data-usage** problem involves

- production technologies $=$ IC mechanisms
- inputs $=$ data $\omega \in \Omega$
- objective $= \sum_\omega u_0(a, \omega) x(a|\omega) \mu(\omega)$

Datapoints and Databases

Frequentist interpretation:

- population of distinct economic interactions between agents (e.g., monopolist-buyer trade for **all** buyers in market)
- Ω = set of types of interactions
- each interaction of type ω = **datapoint** of type ω
- population = **database**
- $\mu(\omega)$ = **stock** of ω-datapoints as share of total quantity in database
- principal commits **ex ante** to how she mediates **all** interactions (ex: all monopolist-buyer trades)

Incentive compatibility $\Rightarrow$ **as if**

- principal **already** owns database with entire datapoints (e.g., platform owns all buyers' valuations even if elicitation needed)
- **but** restricted to using IC mechanisms

The Notion of A Price

Data-pricing problem: given μ, find function

$$p : \Omega \to \mathbb{R}$$

s.t. $p(\omega)$ reflects principal's willingness to pay for **replacing/adding marginal** ω-datapoint to those already in database

Interpretation:
- derivation of **demand functions** for each $\omega \in \Omega$
- each demand depends on overall μ, as mechanisms $\sim$ **non-separable production** technology

Some Other Examples

Internet platform mediating competing firms (Armstrong-Zhou '19)

- platform's own data about buyers' demand
- firms' internal data from market intelligence

Auctions with(out) information design (Bergemann-Pesendorfer '07; Daskalakis et al. '16)

- data from bidders' reports about their valuations
- auctioneer's own data about features of item for sale

Navigation app routing drivers (Kremer et al. '14, Das et al. '17, Liu-Whinston '19)

- app's own data about overall traffic conditions
- drivers' data about desired destination and road conditions

data-pricing formulation

Omniscient Principal

Important case: principal's data fully reveals all parties' data (**omniscient**)

1. simpler to develop concepts and intuitions
2. in many instances (Posner-Weyl '18), principal already knows agents' data and can use it without their consent (akin to no privacy protection)
3. benchmark for problem where principal has to **elicit** agents' data with their consent (akin to privacy protection)

Data Usage Formalized

Consider mechanisms x that have to satisfy **only** obedience

Problem $\mathcal{U}$

$$V_{\mathcal{U}} = \max_x \quad \sum_{\omega,a} u_0(a,\omega)x(a|\omega)\mu(\omega)$$

s.t. for all i, ω_i, a_i, and a_i'

$$\sum_{\omega_{-i},a_{-i}} \Big(u_i\big(a_i,a_{-i},\omega\big) - u_i\big(a_i',a_{-i},\omega\big)\Big)x\big(a_i,a_{-i}|\omega\big)\mu(\omega) \geq 0$$

Question: what is the proper share of $V_{\mathcal{U}}$ to attribute to ω? $\to p(\omega)$

One approach: define **direct value** of ω as $v^*(\omega) = \sum_a u_0(a,\omega)x^*(a|\omega)$

Clearly, $\sum_\omega \mu(\omega)v^*(\omega) = V_{\mathcal{U}}$. But v^* may give incorrect shares/prices ...

An Alternative Approach

Using primitives Γ, we can define a data-pricing problem

Principal designs for each agent i, a_i, and ω_i

$$\ell_i(\cdot|a_i,\omega_i) \in \Delta(A_i) \qquad \text{and} \qquad q_i(a_i,\omega_i) \in \mathbb{R}_{++}$$

Problem $\mathcal{P}$

$$V_{\mathcal{P}} = \min_{\ell,q} \quad \sum_{\omega} p(\omega)\mu(\omega)$$

s.t. for all ω,

$$p(\omega) = \max_{a \in A} \left\{ u_0(a,\omega) + \sum_i T_{\ell_i,q_i}(a,\omega) \right\}$$

$$T_{\ell_i,q_i}(a,\omega) = q_i(a_i,\omega_i) \sum_{a_i' \in A_i} \Big(u_i(a_i,a_{-i},\omega) - u_i(a_i',a_{-i},\omega) \Big) \ell_i(a_i'|a_i,\omega_i)$$

Dual Relationship

Why is $\mathcal{P}$ the 'right' data-pricing problem?

Lemma

Problem $\mathcal{P}$ is equivalent to the **dual** of Problem $\mathcal{U}$. By strong duality,

$$\sum_{\omega} v^*(\omega)\mu(\omega) = V_{\mathcal{U}} = V_{\mathcal{P}} = \sum_{\omega} p^*(\omega)\mu(\omega)$$

Dual Relationship

Why is $\mathcal{P}$ the 'right' data-pricing problem?

Lemma

Problem $\mathcal{P}$ is equivalent to the **dual** of Problem $\mathcal{U}$. By strong duality,

$$\sum_{\omega} v^*(\omega)\mu(\omega) = V_{\mathcal{U}} = V_{\mathcal{P}} = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- $p(\omega)$ corresponds to $\mathcal{U}$-constraint

$$\sum_{a} x(a|\omega) \qquad = 1 \qquad \forall\omega$$

Dual Relationship

Why is $\mathcal{P}$ the 'right' data-pricing problem?

Lemma

Problem $\mathcal{P}$ is equivalent to the **dual** of Problem $\mathcal{U}$. By strong duality,

$$\sum_{\omega} v^*(\omega)\mu(\omega) = V_{\mathcal{U}} = V_{\mathcal{P}} = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- $p(\omega)$ corresponds to $\mathcal{U}$-constraint

$$\sum_{a} x(a|\omega)\mu(\omega) = \mu(\omega) \qquad \forall \omega$$

Dual Relationship

Why is $\mathcal{P}$ the 'right' data-pricing problem?

Lemma

Problem $\mathcal{P}$ is equivalent to the **dual** of Problem $\mathcal{U}$. By strong duality,

$$\sum_{\omega} v^*(\omega)\mu(\omega) = V_{\mathcal{U}} = V_{\mathcal{P}} = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- $p(\omega)$ corresponds to $\mathcal{U}$-constraint

$$\sum_{a} \chi(\omega, a) = \mu(\omega) \qquad \forall \omega$$

Dual Relationship

Why is $\mathcal{P}$ the 'right' data-pricing problem?

Lemma

Problem $\mathcal{P}$ is equivalent to the **dual** of Problem $\mathcal{U}$. By strong duality,

$$\sum_{\omega} v^*(\omega)\mu(\omega) = V_{\mathcal{U}} = V_{\mathcal{P}} = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- $p(\omega)$ corresponds to $\mathcal{U}$-constraint

$$\sum_{a} \chi(\omega, a) = \mu(\omega) \qquad \forall \omega$$

- $p(\omega)$ captures shadow **price** of **stock** $\mu(\omega)$ of ω-datapoints

Dual Relationship

Why is $\mathcal{P}$ the 'right' data-pricing problem?

Lemma

Problem $\mathcal{P}$ is equivalent to the **dual** of Problem $\mathcal{U}$. By strong duality,

$$\sum_{\omega} v^*(\omega)\mu(\omega) = V_{\mathcal{U}} = V_{\mathcal{P}} = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- $p(\omega)$ corresponds to $\mathcal{U}$-constraint

$$\sum_{a} \chi(\omega, a) = \mu(\omega) \qquad \forall \omega$$

- $p(\omega)$ captures shadow **price** of **stock** $\mu(\omega)$ of ω-datapoints
- $p(\omega) =$ principal's WTP for **marginal** ω-datapoint in database

Dual Relationship

Why is $\mathcal{P}$ the 'right' data-pricing problem?

Lemma

Problem $\mathcal{P}$ is equivalent to the **dual** of Problem $\mathcal{U}$. By strong duality,

$$\sum_{\omega} v^*(\omega)\mu(\omega) = V_{\mathcal{U}} = V_{\mathcal{P}} = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- $p(\omega)$ corresponds to $\mathcal{U}$-constraint

$$\sum_{a} \chi(\omega, a) = \mu(\omega) \qquad \forall \omega$$

- $p(\omega)$ captures shadow **price** of **stock** $\mu(\omega)$ of ω-datapoints
- $p(\omega) =$ principal's WTP for **marginal** ω-datapoint in database
- $\mathcal{P}$-variables (ℓ, q) correspond to $\mathcal{U}$-obedience constraints

A Normative Interpretation

$\mathcal{P}$ offers rigorous way of assessing individual price of each datapoint, viewed as **input** in mechanism-information-design problem

A classic interpretation of duality: (Dorfman, Samuelson, Solow '58)

- reminiscent of operations of frictionless **competitive market**
- competition among data users forces to offer data sources full value to which their data give rise
- competition among data sources drives data prices down to minimum consistent with this full value

$\rightsquigarrow$ **normative** meaning to p^*

- takes into account **full** value that each datapoint generates in database
- a benchmark for actual markets for data

back to example

Internet Platform

Seller's profit:

$u_1(a,\omega_0)$	$a=1$	$a=2$
$\omega_0=1$	1	0
$\omega_0=2$	1	2

Buyer's surplus:

$u_0(a,\omega_0)$	$a=1$	$a=2$
$\omega_0=1$	0	0
$\omega_0=2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell,q} \quad \sum_{\omega_0} p(\omega_0)\mu(\omega_0)$$

s.t. for all ω_0,

$$p(\omega_0) = \max_{a\in A}\left\{u_0(a,\omega_0) + T_{\ell,q}(a,\omega_0)\right\}$$

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad \sum_{\omega_0} p(\omega_0)\mu(\omega_0)$$

s.t. for all ω_0,

$$p(\omega_0) = \max_{a \in A} \left\{ u_0(a, \omega_0) + T_{\ell, q}(a, \omega_0) \right\}$$

Internet Platform

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad \sum_{\omega_0} p(\omega_0)\mu(\omega_0) = p(1)\mu + p(2)(1 - \mu)$$

s.t. for all ω_0,

$$p(\omega_0) = \max_{a \in A} \left\{ u_0(a, \omega_0) + T_{\ell, q}(a, \omega_0) \right\}$$

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell,q} \quad p(1)\mu + p(2)(1-\mu)$$

$$\text{s.t.} \quad p(1) = \max\left\{u_0(1,1) + T_{\ell,q}(1,1), u_0(2,1) + T_{\ell,q}(2,1)\right\}$$

$$p(2) = \max\left\{u_0(1,2) + T_{\ell,q}(1,2), u_0(2,2) + T_{\ell,q}(2,2)\right\}$$

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad p(1)\mu + p(2)(1-\mu)$$

$$\text{s.t.} \quad p(1) = \max\left\{q(1)\ell(2|1), -q(2)\ell(2|1)\right\}$$

$$p(2) = \max\left\{1 - q(1)\ell(2|1), q(2)\ell(1|2)\right\}$$

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad p(1)\mu + p(2)(1 - \mu)$$

$$\text{s.t.} \quad p(1) = \max\left\{q(1)\ell(2|1), -q(2)\ell(2|1)\right\} = q(1)\ell(2|1)$$

$$p(2) = \max\left\{1 - q(1)\ell(2|1), q(2)\ell(1|2)\right\}$$

Internet Platform

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad p(1)\mu + p(2)(1-\mu)$$

$$\text{s.t.} \quad p(1) = q(1)\ell(2|1)$$

$$p(2) = \max\left\{1 - q(1)\ell(2|1), q(2)\ell(1|2)\right\}$$

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad p(1)\mu + p(2)(1-\mu)$$

$$\text{s.t.} \quad p(1) = q(1)\ell(2|1)$$

$$p(2) = \max\left\{1 - q(1)\ell(2|1), q(2)\ell(1|2)\right\}$$

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad p(1)\mu + p(2)(1-\mu)$$

$$\text{s.t.} \quad p(1) = q(1)\ell(2|1)$$

$$p(2) = \max\left\{1 - q(1)\ell(2|1), 0\right\}$$

Internet Platform

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad p(1)\mu + p(2)(1-\mu)$$

$$\text{s.t.} \quad p(1) = q(1)\ell(2|1)$$

$$p(2) = \max\left\{1 - q(1)\ell(2|1), 0\right\}$$

Assuming $\mu < \frac{1}{2}$, solution involves setting $q^*(1)\ell^*(2|1) = 1$

Internet Platform

Seller's profit:

$u_1(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	1	0
$\omega_0 = 2$	1	2

Buyer's surplus:

$u_0(a, \omega_0)$	$a = 1$	$a = 2$
$\omega_0 = 1$	0	0
$\omega_0 = 2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell, q} \quad p(1)\mu + p(2)(1-\mu)$$

$$\text{s.t.} \quad p(1) = q(1)\ell(2|1) = 1$$

$$p(2) = \max\left\{1 - q(1)\ell(2|1), 0\right\} = 0$$

Assuming $\mu < \frac{1}{2}$, solution involves setting $q^*(1)\ell^*(2|1) = 1$

Internet Platform

Seller's profit:

$u_1(a,\omega_0)$	$a=1$	$a=2$
$\omega_0=1$	1	0
$\omega_0=2$	1	2

Buyer's surplus:

$u_0(a,\omega_0)$	$a=1$	$a=2$
$\omega_0=1$	0	0
$\omega_0=2$	1	0

Data-pricing problem (seller is the only agent)

$$\min_{\ell,q} \quad p(1)\mu + p(2)(1-\mu)$$

$$\text{s.t.} \quad p(1) = q(1)\ell(2|1) = 1 > v^*(1) = 0$$

$$p(2) = \max\left\{1 - q(1)\ell(2|1), 0\right\} = 0 < v^*(2) = \frac{\mu}{1-\mu}$$

Assuming $\mu < \frac{1}{2}$, solution involves setting $q^*(1)\ell^*(2|1) = 1$

information externalities

Externalities Between Data Entries

Principal **combines** datapoints to produce actionable information

What ω yields depends on which/how other ω' are combined with it

Information externalities between datapoints, which v^* fails to capture

Proposition

Let x^* and (ℓ^*, q^*) be optimal for $\mathcal{U}$ and $\mathcal{P}$. Then

1. $p^*(\omega) > v^*(\omega)$ for some $\omega \iff p^*(\omega') < v^*(\omega')$ for some ω'
2. $p^*(\omega) - v^*(\omega) = \sum_a \Big(\sum_i T_{\ell_i^*, q_i^*}(a, \omega) \Big) x^*(a|\omega)$ for all ω

$1. \Leftarrow$ strong duality: $\sum_\omega [v^*(\omega) - p^*(\omega)]\mu(\omega) = 0$

$2. \Leftarrow$ compl. slackness: $x^*(a|\omega)\left\{p^*(\omega) - v(a,\omega) - \sum_i T_{\ell_i^*, q_i^*}(a,\omega)\right\} = 0$

Externalities Between Data Entries

Why transfer *value* $V_{\mathcal{U}}$ from ω-datapoints to ω'-datapoints?

Definition: Augmented Correlated Equilibrium

$ACE(\Gamma_\omega)$ = distributions $y \in \Delta(A)$ s.t. for all $i \in I$ and $a_i, a_i' \in A_i$,

$$\sum_{a_{-i}} \big(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega)\big)\, y(a_i, a_{-i}) \geq 0$$

Proposition

If $v^*(\omega) > p^*(\omega)$, there must exists a such that $x^*(a|\omega) > 0$ and

$$u_0(a, \omega) > \bar{v}(\omega) = \max_{y \in ACE(\Gamma_\omega)} \sum_a u_0(a, \omega) y(a)$$

Achieve $u_0(a, \omega) > \bar{v}(\omega)$ by pooling ω with $\omega' \rightarrow p^*(\omega') > v^*(\omega')$

In paper: sufficient conditions for $p^* \neq v^*$ and for $p^* = v^*$

Recap

Which datapoints tend to be **less** valuable?

- ω pooled with **other** ω' to produce information that achieves otherwise impossible outcomes for ω

Which datapoints tend to be **more** valuable?

- ω pooled with other ω' to **help** ω' achieve otherwise impossible outcomes

what drives p^*

An **independent** interpretation of $\mathcal{P}$ to understand what drives p^*

Recall :
$$\min_{\ell,q} \quad \sum_{\omega} p(\omega)\mu(\omega)$$
$$\text{s.t.} \quad p(\omega) = \max_{a\in A}\left\{u_0(a,\omega) + \sum_i T_{\ell_i,q_i}(a,\omega)\right\} \quad \forall\omega$$

$\rightarrow$ p ultimately determined by (ℓ, q) through best trade-off between

1. principal's **direct payoff** u_0
2. "transfer" function T_{ℓ_i,q_i} that account for **information externalities**

What are ℓ and q?

Dual Side of Mechanisms-Information Design

Fix (a, ω) and recall $q_i(a_i, \omega_i) \in \mathbb{R}_{++}$, $\ell_i(\cdot|a_i, \omega_i) \in \Delta(A_i)$, and

$$T_{\ell_i, q_i}(a, \omega) = q_i(a_i, \omega_i) \sum_{a_i' \in A_i} \Big(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega)\Big) \ell_i(a_i'|a_i, \omega_i)$$

Principal designs **gambles** against agents **contingent** on (a, ω)

- (ℓ_i, q_i) family of gambles (lottery & stake) contingent on (a_i, ω_i)
- given (a, ω), $\ell_i(?|a_i, \omega_i)$ yields **prize** $u_i(a_i, a_{-i}, \omega) - u_i(?, a_{-i}, \omega)$
- principal **wins** iff $u_i(a_i, a_{-i}, \omega) < u_i(a_i', a_{-i}, \omega)$
  $\leftrightarrow$ had i known (a_{-i}, ω), he would have preferred $a_i' \neq a_i$ (**ex-post mistake**)
- for every ω, value $p(\omega)$ given by best trade-off between $u_0(a, \omega)$ and gambles $\sum_i T_{\ell_i, q_i}(a, \omega)$ across a
- principal commits to (ℓ, q) ex ante $\rightarrow$ average with respect to μ

Structure of Gambles

$\min_{\ell,q} \sum p(\omega)\mu(\omega) \rightsquigarrow$ principal wants to win gambles as much as possible

Constraint 1: Limited Flexibility

gambles against i can be tailored to (a_i, ω_i), but not (a_{-i}, ω_{-i})

$\rightsquigarrow$ **links** between pricing formula of (ω_i, ω_{-i}) and (ω_i, ω'_{-i})

- manifestation in $\mathcal{P}$ of non-separabilities in $\mathcal{U}$ across ω
- still pin down *individual* prices for each ω

$\rightsquigarrow$ **trade-offs** across datapoints: using (ℓ_i, q_i) to lower $p(\omega_i, \omega_{-i})$ may cost raising $p(\omega_i, \omega'_{-i})$

Structure of Gambles

$\min_{\ell,q} \sum p(\omega)\mu(\omega) \rightsquigarrow$ principal wants to win gambles as much as possible

Constraint 2: Agents' Joint Rationality (Nau '92)

$\sim$ agents accept gambles where they lose in (a, ω) only if they win in (a', ω')

Proposition

For every* (ℓ, q), if $\sum_i T_{\ell_i, q_i}(a, \omega) < 0$ for (a, ω), there must exist (a', ω') such that $\sum_i T_{\ell_i, q_i}(a', \omega') > 0$

$\Rightarrow$ **key trade-off** for principal:

winning less important for relatively scarce data (low μ) $\rightsquigarrow$ higher price

Relations between Pricing and Using Data

Optimal (ℓ^*, q^*) for $\mathcal{P}$ has corresponding optimal x^* for $\mathcal{U}$ (and vice versa)

Proposition

Generically, $\ell_i^*(a_i'|a_i, \omega_i) > 0$ if and only if, given ω_i, agent i indifferent between a_i' and recommendation a_i from x^*

$\sim$ only **indifferent agents** under x^* contribute to gap $p^*(\omega) - v^*(\omega)$

Proposition

Generically, $x^*(a|\omega) > 0$ if and only if $p^*(\omega) = u_0(a, \omega) + \sum_i T_{\ell_i^*, q_i^*}(a, \omega)$

$\sim$ all uses of ω-datapoints under x^* yield **same** (maximal) **total** value $p^*(\omega)$

Recap

Which datapoints tend to be more valuable?

1. ω that helps principal trick agents into making ex-post mistakes for some **other** ω'
2. ω relatively scarce in database (i.e., low $\mu(\omega)$)

Which datapoints tend to be less valuable?

1. ω where agents make ex-post mistakes with help of some **other** ω'
2. ω relatively abundant in database (i.e., high $\mu(\omega)$)

example II

Mediated Cournot Competition in e-Commerce

To illustrate, operator (principal) manages online marketplace

Two firms (agents), each chooses to participate or not: produce $a_i \in \{0, 1\}$

Profits: $u_i(a_i, a_{-i}, \omega_0) = \left(\omega_0 - \sum_i a_i\right) a_i$

Demand strength: $\Omega_0 = \{\underline{\omega}_0, \bar{\omega}_0\}$, $\mu(\underline{\omega}_0) = \mu(\bar{\omega}_0) = \frac{1}{2}$

Operator maximizes total production: $u_0(a, \omega) = \sum_i a_i$

Mediated Cournot Competition in e-Commerce

Firms have own data about demand strength: $\Omega_i = \{\underline{\omega}_i, \bar{\omega}_i\}$

$\underline{\omega}_0$	$\underline{\omega}_2$	$\bar{\omega}_2$
$\underline{\omega}_1$	γ^2	$\gamma(1-\gamma)$
$\bar{\omega}_1$	$\gamma(1-\gamma)$	$(1-\gamma)^2$

$\bar{\omega}_0$	$\underline{\omega}_2$	$\bar{\omega}_2$
$\underline{\omega}_1$	$(1-\gamma)^2$	$\gamma(1-\gamma)$
$\bar{\omega}_1$	$\gamma(1-\gamma)$	γ^2

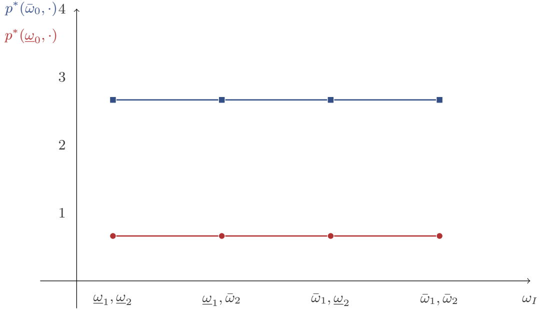
where $1/2 < \gamma < 1$

Data usage: given ω, convey info to influence a_1 and a_2

Data pricing: find $p(\omega) = p(\omega_0, \omega_1, \omega_2)$ for all ω

Today, assume $\omega_0 \in \{0, 3\}$

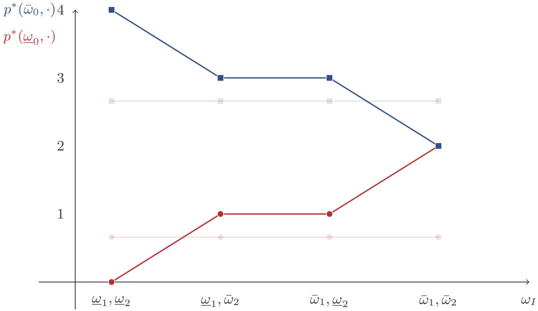
Optimal Prices p^*



Case 1: firms' data gives weak signal, $\gamma < \underline{\gamma}$

- prices independent of (ω_1, ω_2)
- $\bar{\omega}_0$ is more valuable than $\underline{\omega}_0$
  - $p^*(\underline{\omega}_0, \omega_1, \omega_2) < v^*(\underline{\omega}_0, \omega_1, \omega_2)$ and $p^*(\bar{\omega}_0, \omega_1, \omega_2) > v^*(\bar{\omega}_0, \omega_1, \omega_2)$
  - gambles: $q_i^*(1, \underline{\omega}_i)\ell_i^*(0|1, \underline{\omega}_i) = q_i^*(1, \bar{\omega}_i)\ell_i^*(0|1, \bar{\omega}_i) > 0$, for all i

Optimal Prices p^*



Case 2: firms' data gives strong signal, $\gamma > \bar{\gamma}$

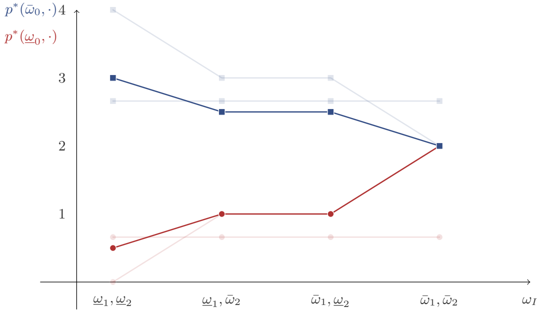
- pessimistic firms $\rightsquigarrow$ pooling harder $\rightsquigarrow$ larger externality

$$p^*(\underline{\omega}_0, \underline{\omega}_1, \underline{\omega}_2) < v^*(\underline{\omega}_0, \underline{\omega}_1, \underline{\omega}_2) < v^*(\bar{\omega}_0, \underline{\omega}_1, \underline{\omega}_2) < p^*(\bar{\omega}_0, \underline{\omega}_1, \underline{\omega}_2)$$

- optimistic firms $\rightsquigarrow$ always produce $\rightsquigarrow$ no externalities

$$p^*(\underline{\omega}_0, \bar{\omega}_1, \bar{\omega}_2) = v^*(\underline{\omega}_0, \bar{\omega}_1, \bar{\omega}_2) = v^*(\bar{\omega}_0, \bar{\omega}_1, \bar{\omega}_2) = p^*(\bar{\omega}_0, \bar{\omega}_1, \bar{\omega}_2)$$

- gambles: $q_i^*(1, \underline{\omega}_i)\ell_i^*(0|1, \underline{\omega}_i) > 0 = q_i^*(1, \bar{\omega}_i)\ell_i^*(0|1, \bar{\omega}_i)$, for all i



Case 3: firms' data gives intermediate signal, $\underline{\gamma} < \gamma < \bar{\gamma}$

- pessimistic firms $\rightsquigarrow$ pooling harder $\rightsquigarrow$ larger externality

$$p^*(\underline{\omega}_0, \underline{\omega}_1, \underline{\omega}_2) < v^*(\underline{\omega}_0, \underline{\omega}_1, \underline{\omega}_2) < v^*(\bar{\omega}_0, \underline{\omega}_1, \underline{\omega}_2) < p^*(\bar{\omega}_0, \underline{\omega}_1, \underline{\omega}_2)$$

- optimistic firms $\rightsquigarrow$ always produce $\rightsquigarrow$ no externalities

$$p^*(\underline{\omega}_0, \bar{\omega}_1, \bar{\omega}_2) = v^*(\underline{\omega}_0, \bar{\omega}_1, \bar{\omega}_2) = v^*(\bar{\omega}_0, \bar{\omega}_1, \bar{\omega}_2) = p^*(\bar{\omega}_0, \bar{\omega}_1, \bar{\omega}_2)$$

- gambles: $q_i^*(1, \underline{\omega}_i)\ell_i^*(0|1, \underline{\omega}_i) > 0 = q_i^*(1, \bar{\omega}_i)\ell_i^*(0|1, \bar{\omega}_i)$, for all i

prices under privacy

Privacy Protection

Suppose principal has to incentivize agents to report their private data

Incentives:

- directly from how principal **commits** to use data (no monetary transfers)
- in some settings, monetary transfer as part of mechanisms

Formally, mechanisms in problem $\mathcal{U}$ must satisfy **honesty** and obedience

Question: How are prices affected by need to elicit data?

Data Usage with Elicitation

Elicitation does not change mathematical structure of problem

Problem $\mathcal{U}^e$

$$V_{\mathcal{U}} = \max_x \quad \sum_{\omega, a} u_0(a,\omega) x(a|\omega)\mu(\omega)$$

s.t. for all i, ω_i, and $\delta_i : A_i \to A_i$

$$\sum_{a_i, a_{-i}, \omega_{-i}} u_i\big(a_i, a_{-i}, \omega\big) x\big(a_i, a_{-i}|\omega_i, \omega_{-i}\big)\mu(\omega_i, \omega_{-i}) \geq$$

$$\sum_{a_i, a_{-i}, \omega_{-i}} u_i\big(\delta_i(a_i), a_{-i}, \omega\big) x\big(a_i, a_{-i}|\omega_i, \omega_{-i}\big)\mu(\omega_i, \omega_{-i})$$

Data Usage with Elicitation

Elicitation does not change mathematical structure of problem

Problem $\mathcal{U}^e$

$$V_{\mathcal{U}^e} = \max_x \quad \sum_{\omega, a} u_0(a, \omega) x(a|\omega)\mu(\omega)$$

s.t. for all i, ω_i, ω_i', and $\delta_i : A_i \to A_i$

$$\sum_{a_i, a_{-i}, \omega_{-i}} u_i\big(a_i, a_{-i}, \omega\big) x\big(a_i, a_{-i}|\omega_i, \omega_{-i}\big)\mu(\omega_i, \omega_{-i}) \geq$$

$$\sum_{a_i, a_{-i}, \omega_{-i}} u_i\big(\delta_i(a_i), a_{-i}, \omega\big) x\big(a_i, a_{-i}|\omega_i', \omega_{-i}\big)\mu(\omega_i, \omega_{-i})$$

Principal chooses, for each player i and ω_i,

$$\hat{\ell}_i(\cdot|\omega_i) \in \Delta(\Omega_i \times D_i) \qquad \text{and} \qquad \hat{q}_i(\omega_i) \in \mathbb{R}_{++}$$

Problem $\mathcal{P}^e$

$$V_{\mathcal{P}^e} = \min_{\hat{\ell},\hat{q}} \quad \sum_{\omega} p(\omega)\mu(\omega)$$

s.t. for all ω,

$$p(\omega) = \max_{a\in A} \left\{ u_0(a,\omega) + \sum_i T_{\hat{\ell}_i,\hat{q}_i}(a,\omega) \right\}$$

Data Pricing with Elicitation

Data pricing with vs without elicitation:

- transfer function $T_{\hat{\ell}_i, \hat{q}_i}$ now involves **richer gambles** $(\hat{\ell}, \hat{q})$
- principal can win against agent when
  1. deviating from obedience is ex-post beneficial (as before)
  2. deviating from honesty is ex-post beneficial (new)
  3. both (new)

Work in progress:

- $p(\omega)$ incorporates difficulty to honestly elicit ω: new externalities
- compare $p(\omega)$ under omniscient and elicitation $\rightsquigarrow$ insights into effects on value of data (e.g., effects of privacy protection)
- compare $p(\omega)$ under elicitation with monetary transfer (if any) to agents for their data $\rightsquigarrow$ are they properly rewarded?

back to example II

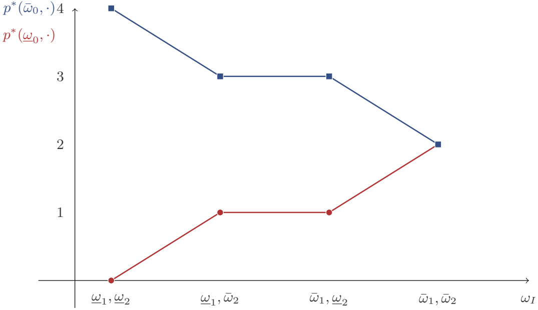
Cournot Competition with Elicitation

Clearly, **total** value $V_{\mathcal{U}}$ decreases with elicitation. What about **individual** p^*?

Cournot Competition with Elicitation

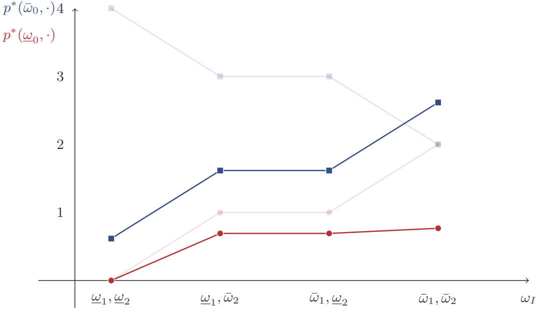
Clearly, **total** value $V_{\mathcal{U}}$ decreases with elicitation. What about **individual** p^*?

$\boxed{\gamma > \bar{\gamma}}$



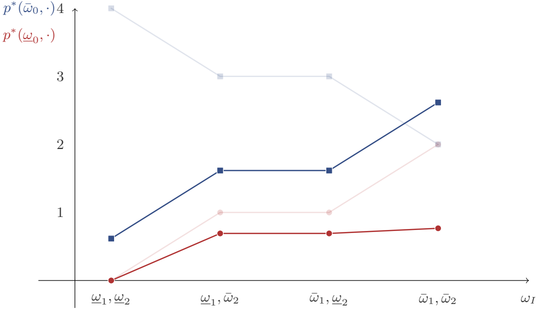
Clearly, **total** value $V_{\mathcal{U}}$ decreases with elicitation. What about **individual** p^*?

$\boxed{\gamma > \bar{\gamma}}$



Clearly, **total** value $V_{\mathcal{U}}$ decreases with elicitation. What about **individual** p^*?

$\boxed{\gamma > \bar{\gamma}}$



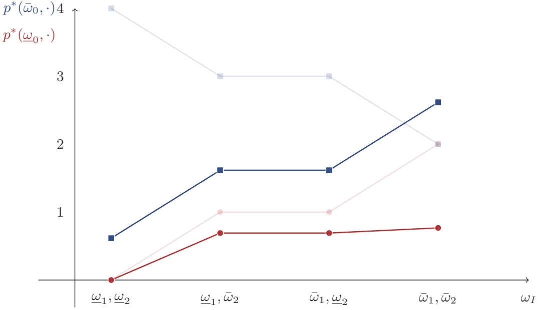
1. elicitation $\rightsquigarrow$ **qualitative** change in $p(\bar{\omega}_0, \omega_1, \omega_2)$

- $\bar{\omega}_i$ tempted to mimic $\underline{\omega}_i$ to get more informative recommendation
- ω induces temptation to lie $\to$ suffers negative externality (gambles)
- x^* **distorted** to make mimicking $\underline{\omega}_i$ less attractive, despite $\bar{\omega}_0$

Cournot Competition with Elicitation

Clearly, **total** value $V_{\mathcal{U}}$ decreases with elicitation. What about **individual** p^*?

$\boxed{\gamma > \bar{\gamma}}$



2. $p^*(\bar{\omega}_0, \bar{\omega}_1, \bar{\omega}_2)$ **higher** than in omniscient case

– mimicking gamble $\hat{\ell}_i(\underline{\omega}_i, \cdot | \bar{\omega}_i) > 0 \to$ loss for principal if $\omega_0 = \bar{\omega}_0$

– $(\bar{\omega}_0, \bar{\omega}_1, \bar{\omega}_2)$ only data left with full participation under $x^* \rightsquigarrow$ value $\uparrow$

extensions

Next Steps

Robust data usage:

- **robust mechanisms** that do not rely on agents' higher-order beliefs
- for example, ex-post equilibrium $\to$ LP and similar data pricing

Restrictions on data usage:

- mechanism x can **depend only on parts of datapoint** ω
- for example, auctioneer can use data to influence bidders' valuations, but not to directly run the auction (Bergemann-Pesendorfer '07)
- formulated as linear constraints on $x \to$ LP and similar data pricing

Value of **more precise data** for each mediated interaction:

- ω_0' is more precise data than ω_0 about buyer's valuation for seller's product (e.g., longer cookie history) $\rightsquigarrow$ databases (Ω, μ) and (Ω', μ')
- individual value of extra data $= p^*(\omega_0') - p^*(\omega_0)$

summary

Summary

A theory of how to price datapoints in a database to reflect their individual value

Basic insight:

- data-usage problem = mechanism-information design problem
- data-pricing problem = its dual

Preliminary analysis reveals:

- prices take into account information externalities across datapoints
- valuable data: scarce + helps trick agents into making mistakes
- rigorous method to assess effects of privacy protection: can have significant impact and increase prices of some types of data