The Price of Data

Simone Galperti
UC San Diego

Aleksandr Levkun
UC San Diego

Jacopo Perego
Columbia University

August 2020

Overview

Data has become an essential commodity in modern economies

A few markets for data have emerged, where data sources are compensated for the data they generate

This paper: A theory of how to **individually** price the **entries** of a dataset so as to reflect their proper value

Overview

Data has become an essential commodity in modern economies

A few markets for data have emerged, where data sources are compensated for the data they generate

This paper: A theory of how to **individually** price the **entries** of a dataset so as to reflect their proper value

Our questions:

- **Normative:** How much does each **entry** contribute to the total value of the dataset for its owner?
- **Operational:** What is owner's WTP for an additional data entry?
- What drive these prices and how can we compute them?
- How are these prices affected by **privacy** concerns?

Overview

Our approach leverages a **simple insight**:

- The **data-pricing problem** is intimately related to *how the dataset is used by its owner to achieve a given goal*
- When carefully formulated, the two problems are in a *special mathematical relationship*

Overview

Our approach leverages a **simple insight**:

- The **data-pricing problem** is intimately related to the **information design problem**
- When carefully formulated, the two problems are in a *special mathematical relationship*

Overview

Our approach leverages a **simple insight**:

- The **data-pricing problem** is intimately related to the **information design problem**
- When carefully formulated, the two problems are in a **dual relationship**

Overview

Our approach leverages a **simple insight**:

- The **data-pricing problem** is intimately related to the **information design problem**
- When carefully formulated, the two problems are in a **dual relationship**

Goal for Today's Talk

1. Formalize relationship + data-pricing problem
2. Preliminary characterization of price determinants and properties
3. Showcase properties through examples

Modeling Ingredients

A standard and flexible framework:

- Finite static games with incomplete information

Data **entries** and the **dataset**:

- A "data entry" is a state of the world:

  Payoff state + players' private signals about it

- The "dataset" consists of all entries + their frequencies

Designer may use entries :

- Without players' consent **(no privacy)**
- Only with players' consent **(privacy)**

Preliminary Results

Pricing formula

- Individual price for each data entry despite info-design problem being non-separable across states

What drives the prices?

- (1) Designer's payoff + (2) Designing information equivalent to gambling against players (*novel interpretation* for dual variables)

Properties

- Price captures externalities that each data entry may exert on others
- Price captures dependencies between dimensions of each data entry

The effects of privacy protection

- It lowers value of dataset, but can increase price of some entries

Related Literature

Information Design. Kamenica & Gentzkow ('11), Bergemann & Morris ('16, '19), ...

Duality & Correlated Equilibrium. Nau & McCardle ('90), Nau ('92), Hart & Schmeidler ('89), Myerson ('97)

Duality & Bayesian Persuasion. Kolotilin ('18), Dworczak & Martini ('19), Dizdar & Kovac ('19), Dworczak & Kolotilin ('19)

Markets for Information. Bergemann & Bonatti ('19) Bergemann & Bonatti ('15), Bergmann, Bonatti, Smolin ('18)

Information Privacy. Ali, Lewis, and Vasserman ('20), Bergemann, Bonatti, and Gan ('20), Acemoglu, Makhdoumi, Malekian, and Ozdaglar, ('20), Acquisti, Taylor, Wagman ('16) illustrative example

A Monopolist's Problem (Bergemann et al. '15)

example

Monopolist sells to potential buyers (assume MC=0)

Monopolist does not directly observe buyers' valuation

A dataset contains data about the potential buyers:

- A share $\mu > \frac{1}{2}$ of the entries has valuation $\omega = 2$
- A share $1 - \mu$ of the entries has valuation $\omega = 1$

A **data intermediary** owns the dataset; can use it without buyers' consent

Monopolist sets price a and can discriminate depending on the information she receives

A Monopolist's Problem (Bergemann et al. '15)

Suppose monopolist receives this information about the potential buyer

	s'	s''
$\omega = 1$	1	0
$\omega = 2$	$\frac{1-\mu}{\mu}$	$1 - \frac{1-\mu}{\mu}$

Monopolist would set

$$a(s) = \begin{cases} 1 & \text{for "segment" } s' \\ 2 & \text{for "segment" } s'' \end{cases}$$

The total **consumer surplus** is $V^* = 1 - \mu$ and for each buyer ω

$$v^*(\omega) = \begin{cases} 0 & \text{if } \omega = 1 \\ \frac{1-\mu}{\mu} & \text{if } \omega = 2 \end{cases}$$

Our Questions:

- What price $p(\omega)$ would/should the data intermediary be **willing to pay** to add one more buyer with valuation ω to her dataset?
- What price $p(\omega)$ would "properly" compensate buyer ω for role that her data plays to achieve V^*?

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

Our Questions:

- What price $p(\omega)$ would/should the data intermediary be **willing to pay** to add one more buyer with valuation ω to her dataset?
- What price $p(\omega)$ would "properly" compensate buyer ω for role that her data plays to achieve V^*?

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

We do **not** interpret $p(\omega)$ as monetary incentive to give up data

- Important, yet distinct issue

model

Data Entries and Dataset

Finite set of **players** $I = \{1, \dots, n\}$

Finite set of **payoff states** Ω_0

Finite set of **private types** $\Omega_I = \Omega_1 \times \ldots \times \Omega_n$, players' own data

Common prior belief $\mu \in \Delta(\Omega)$, where $\Omega = \Omega_0 \times \Omega_I$

We refer to (Ω, μ) as a **dataset** and to each ω as a **data entry**

Base Game and Information

Each player i has finite set of actions A_i. Let $A = A_1 \times \ldots \times A_n$

Utility function $u_i : A \times \Omega_0 \to \mathbb{R}$

Base game $G = \Big(I, (\Omega, \mu), (A_i, u_i)_{i \in I}\Big)$

An **information structure** is $\pi : \Omega \to \Delta(S_1 \times \ldots \times S_n)$, with S_i finite $\forall i$

BNE(G, π) set of Bayes-Nash equilibria for (G, π)

Designer provides information via π to players

Objective is $v : A \times \Omega_0 \to \mathbb{R}$

We consider two cases:

1. **Omniscient** design. Designer already owns dataset and can use it without players' consent (akin to no privacy protection)
2. Design w/ **Elicitation**. Designer has to obtain players' data and needs their consent (akin to privacy protection)

We begin by analyzing the data-pricing problem under omniscient design data-pricing problem

The Notion of a Price

The **data-pricing problem** consists in finding a function

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

s.t. $p(\omega)$ reflects the "proper" value that ω generates **for the designer**

p should depend on **how data entries are used** to produce information

We think of data entries ω's as inputs into a **production problem** whose output is **information**:

$$\pi : \Omega \to \Delta(S)$$

Data-pricing problem $\iff$ Data-use problem

Build on the information-design literature:

- How to optimally use data to produce information so as to maximize a given objective

For each π, define

$$V(\pi) = \max_{\sigma \in \text{BNE}(G,\pi)} \sum_{\omega,s,a} v(a,\omega_0)\left(\prod_{i\in I} \sigma(a_i|\omega_i,s_i)\right)\pi(s|\omega)\mu(\omega)$$

The **information-design problem** consists of $V^\star = \max_\pi V(\pi)$

Question

- What is the proper share of V^* to attribute to ω? $\to p(\omega)$

Direct Value of Data

One possible approach to answer this question:

1. Find solution of ID problem π^* and σ^*
2. Compute **direct value** of ω. This is the expected payoff from ω

$$v^*(\omega) = \sum_s v(a, \omega_0)\sigma^*(a|s, \omega_I)\pi^*(s|\omega)$$

Clearly, $\sum_\omega \mu(\omega)v^*(\omega) = V^*$

Direct Value of Data

One possible approach to answer this question:

1. Find solution of ID problem π^* and σ^*
2. Compute **direct value** of ω. This is the expected payoff from ω

$$v^*(\omega) = \sum_s v(a, \omega_0)\sigma^*(a|s, \omega_I)\pi^*(s|\omega)$$

Clearly, $\sum_\omega \mu(\omega)v^*(\omega) = V^*$

Does $v^*(\omega)$ capture the share of $V^\star$ that is attributable to ω?

Not quite! it fails to capture that ω may play a role in the payoff that is generated by another ω'

An Alternative Approach

The information-design problem can be formulated as a linear program

Let $x : \Omega \to \Delta(A)$ be an **outcome** for G

By Bergemann and Morris (2016), "feasibility" of x is equivalent to a set of **obedience conditions** which are linear constraints on x.

Problem $\mathcal{P}$ (Bergemann and Morris, 2016, 2019)

$$\max_{x} \quad \sum_{\omega, a} v(a, \omega_0) x(a|\omega) \mu(\omega)$$

s.t. for all i, ω_i, a_i, and a_i'

$$\sum_{\omega_{-i}, a_{-i}} \Big(u_i(a_i, a_{-i}, \omega_0) - u_i(a_i', a_{-i}, \omega_0) \Big) x(a_i, a_{-i}|\omega) \mu(\omega) \geq 0$$

Data-Pricing Problem

Using same primitives (G, v), we can define a data-pricing problem

Designer chooses, for each player i, a_i, and ω_i

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

Problem $\mathcal{D}$ (Data-Pricing Problem)

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

s.t. for all ω,

$$p(\omega) = \max_{a \in A} \Big\{ v(a,\omega_0) + \sum_i T_{\ell_i,q_i}(a,\omega) \Big\}$$

Where:

$$T_{\ell_i,q_i}(a,\omega) = q_i(a_i,\omega_i) \sum_{a_i' \in A_i} \Big(u_i(a_i,a_{-i},\omega_0) - u_i(a_i',a_{-i},\omega_0)\Big)\ell_i(a_i'|a_i,\omega_i)$$

Information-design and data-pricing problems are connected:

Lemma

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

$$\sum_{\omega} v^*(\omega)\mu(\omega) = \sum_{\omega} p^*(\omega)\mu(\omega)$$

Dual Relationship

Information-design and data-pricing problems are connected:

Lemma

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

$$\sum_{\omega} v^*(\omega)\mu(\omega) = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- Price $p(\omega)$ in $\mathcal{D}$ corresponds to $\mathcal{P}$-constraint

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

Dual Relationship

Information-design and data-pricing problems are connected:

Lemma

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

$$\sum_{\omega} v^*(\omega)\mu(\omega) = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- Price $p(\omega)$ in $\mathcal{D}$ corresponds to $\mathcal{P}$-constraint

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

Dual Relationship

Information-design and data-pricing problems are connected:

Lemma

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

$$\sum_{\omega} v^*(\omega)\mu(\omega) = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- Price $p(\omega)$ in $\mathcal{D}$ corresponds to $\mathcal{P}$-constraint

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

Dual Relationship

Information-design and data-pricing problems are connected:

Lemma

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

$$\sum_{\omega} v^*(\omega)\mu(\omega) = \sum_{\omega} p^*(\omega)\mu(\omega)$$

- Price $p(\omega)$ in $\mathcal{D}$ corresponds to $\mathcal{P}$-constraint

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

- Thus, $p(\omega)$ captures the shadow **price** of relaxing $\mu(\omega)$
- Designer's WTP for one more ω in the dataset
- The $\mathcal{D}$-variables (ℓ, q) correspond to $\mathcal{P}$-obedience constraints

A Normative Interpretation

Problem $\mathcal{D}$ as a rigorous way of assessing the individual price of each state, viewed as data input in the information-design problem

A classic interpretation: Dorfman, Samuelson, Solow (1958)

- Reminiscent of the operations of a frictionless **competitive market**
- Competition among data intermediaries forces to offer data sources the full value to which their data give rise
- Competition among data sources drives data prices down to the minimum consistent with this full value

A Normative Interpretation

Problem $\mathcal{D}$ as a rigorous way of assessing the individual price of each state, viewed as data input in the information-design problem

A classic interpretation: Dorfman, Samuelson, Solow (1958)

- Reminiscent of the operations of a frictionless **competitive market**
- Competition among data intermediaries forces to offer data sources the full value to which their data give rise
- Competition among data sources drives data prices down to the minimum consistent with this full value

Thus, these prices have a **normative** interpretation

- $p^*(\omega)$ takes into account information externalities that ω generates
- A possible benchmark to be used in actual markets for data

back to example

A Monopolist's Problem

Monopolist's Profit:

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

Consumer Surplus:

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

Information-design problem finds π^* and *direct values* are

$$v^*(\omega_0) = \begin{cases} 0 & \text{if } \omega_0 = 1 \\ \frac{1-\mu}{\mu} & \text{if } \omega_0 = 2 \end{cases}$$

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

s.t. for all ω_0,

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

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

s.t. for all ω_0,

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

A Monopolist's Problem

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

s.t. for all ω_0,

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

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

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

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

A Monopolist's Problem

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

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

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

A Monopolist's Problem

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

$$\begin{aligned} \min_{\ell, q} \quad & p(1)(1-\mu) + p(2)\mu \\ \text{s.t.} \quad & p(1) = \max\left\{q(1)\ell(2|1), -q(2)\ell(2|1)\right\} = q(1)\ell(2|1) \\ & p(2) = \max\left\{1 - q(1)\ell(2|1), q(2)\ell(1|2)\right\} \end{aligned}$$

A Monopolist's Problem

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

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

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

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

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

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

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

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

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

A Monopolist's Problem

example

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

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

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

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

Monopolist's Profit:

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

Consumer Surplus:

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

Data-Pricing Problem

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

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

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

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

Splitting The Surplus

Therefore, we obtain **prices**

$$p^*(1) = 1 \qquad p^*(2) = 0$$

Whereas, **direct values** are

$$v^*(1) = 0 \qquad v^*(2) = \frac{1-\mu}{\mu}$$

Splitting The Surplus

Therefore, we obtain **prices**

$$p^*(1) = 1 \qquad p^*(2) = 0$$

Whereas, **direct values** are

$$v^*(1) = 0 \qquad v^*(2) = \frac{1-\mu}{\mu}$$

Discussion:

- Designer not willing to pay for additional entry $\omega_0 = 2$, despite the only with positive direct value $v^*(2) > 0$
- Designer willing to pay for additional entry $\omega_0 = 1$ into dataset
- Why? Buyer $\omega_0 = 1$ receives no surplus, yet her data plays key role to generate surplus for $\omega_0 = 2$
- This externality cannot be captured by v^*, but it is by p^*

information externalities

Externalities Between States

Gap between v^* and p^* is not a special feature of the example

Direct values from $\mathcal{P}$ "misprice" data entries as it fails to incorporate the possible **information externalities** that exist between states

We characterize these externalities:

Proposition

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

$$p^*(\omega) - v^*(\omega) = T^*(\omega) \qquad \forall\omega$$

where $T^*(\omega) = \sum_a \Big(\sum_i T_{\ell_i^*, q_i^*}(a, \omega) \Big) x^*(a|\omega)$. Moreover,

$$p^*(\omega) > v^*(\omega) \iff p^*(\omega') < v^*(\omega')$$

Externalities Between States

To gain intuition, let

$$\Omega^*_- = \{\omega : v^*(\omega) > p^*(\omega)\} \qquad \Omega^*_+ = \{\omega : v^*(\omega) < p^*(\omega)\}$$

Why transfer of *value* V^* from states in Ω_- to states in Ω_+?

Proposition

If $\omega \in \Omega^*_-$, there must exists a such that $x^*(x|\omega) > 0$ and

$$v(a, \omega_0) > \bar{v}(\omega_0) = \max_{\sigma \in CE(G_{\omega_0})} \sum_a v(a, \omega_0)\sigma(a)$$

Designer achieves $v(a, \omega_0) > \bar{v}(\omega_0)$ by pooling $\omega \in \Omega^*_-$ with other states, specifically those in Ω^*_+

Converse. If x^* involves no pooling — it can be implemented by a **fully revealing** π — then there is no externality and $p^* = v^*$

what drives p^*

An interpretation to understand how the prices are determined

Recall that:

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

s.t. for all ω,

$$p(\omega) = \max_{a \in A} \left\{ v(a, \omega_0) + \sum_i T_{\ell_i, q_i}(a, \omega) \right\}$$

The price of ω ultimately determined by (ℓ, q) through:

1. **Designer**'s payoff v
2. The "transfer" function T_{ℓ_i, q_i}, which depends on **player**'s i utility u_i

Fix player i and outcome realization (a, ω):

$$T_{\ell_i, q_i}(a, \omega) = q_i(a_i, \omega_i) \sum_{a_i' \in A_i} \Big(u_i(a_i, a_{-i}, \omega_0) - u_i(a_i', a_{-i}, \omega_0) \Big) \ell_i(a_i' | a_i, \omega_i)$$

Interpretation of (ℓ_i, q_i) as **bets** against player i contingent on (a_i, ω_i):

- $\ell_i(\cdot | a_i, \omega_i) \in \Delta(A_i)$ is a lottery offered to the player
- Prizes of such lottery given by $u_i(a_i, a_{-i}, \omega_0) - u_i(a_i', a_{-i}, \omega_0)$
- Designer puts stake $q_i(a_i, \omega_i) > 0$ into such lottery

Player wins if $T_{\ell_i, q_i}(a, \omega) > 0$ and loses if $T_{\ell_i, q_i}(a, \omega) < 0$

– If loses, she would have been better off playing some $a_i' \neq a_i$ given (a_{-i}, ω_0) (ex post mistake)

What drives the choice of these bets? Recall, $\min_{\ell,q} \sum p(\omega)\mu(\omega)$

- Designer's overall goal is to win against players as much as possible

What drives the choice of these bets? Recall, $\min_{\ell,q} \sum p(\omega)\mu(\omega)$

- Designer's overall goal is to win against players as much as possible

However, designer faces two kinds of constraints

1. Links Between States

- Bets for i can be tailored to (a_i, ω_i), but not (a_{-i}, ω_{-i})
- This creates trade-offs across states, as the best bet for (ω_i, ω_{-i}) may not be the same as the best bet for (ω_i, ω'_{-i})
- Thus, pricing formulas are linked across states, yet they still pin down prices *state-by-state*
- This structure is constraining because bets are chosen ex ante with commitment, just like x in problem $\mathcal{P}$

What drives the choice of these bets? Recall, $\min_{l,q} \sum p(\omega)\mu(\omega)$

- Designer's overall goal is to win against players as much as possible

However, designer faces several constraints

2. Player's Incentives

- **Result:** If designers wins at (a, ω), she must lose at some other (a', ω')
- Represents counterpart in $\mathcal{D}$ of *Bayesian rationality* in $\mathcal{P}$ (Nau '92)
  - Intuitively, if i accepts a losing bet at some (a, ω), she must receive a winning bet at some other (a', ω')
- **Result:** Optimal bets must induce player's indifference
  - Generically, $T_{\ell_i, q_i}(a, \omega) \neq 0$ *if and only if* i is indifferent in $\mathcal{P}$ conditional on (a_i, ω_i), between a_i and the lottery

example II

A Simple Cournot Game

To illustrate, we consider a data-pricing problem with **strategic interactions** and **private information**

Two firms, each sets a production quantity $a_i \in \{0, 1\}$

Profits are given by $u_i(a_i, a_{-i}, \omega_0) = \Big(\omega_0 - \sum_i a_i\Big) a_i$

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

Designer maximizes total production, $v(a, \omega_0) = \sum_i a_i$

A Simple Cournot Game

Firms are privately informed about demand ω_0: $\Omega_i = \{\underline{\omega}_i, \bar{\omega}_i\}$

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

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

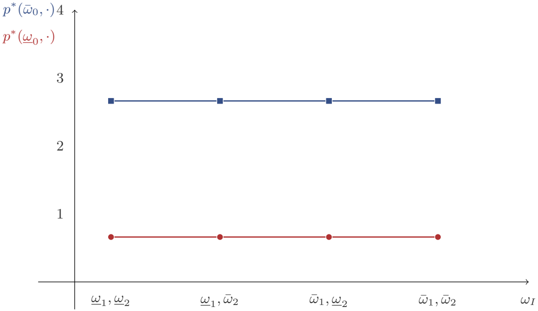
where $1/2 < \gamma, \eta < 1$

The data-pricing problem finds $p(\omega) = p(\omega_0, \omega_1, \omega_2)$, for all ω

Duality as a **solution method** to analytically find optimal p^* and x^*

Today, show results for $\gamma = \eta$ and $\omega_0 \in \{0, 3\}$

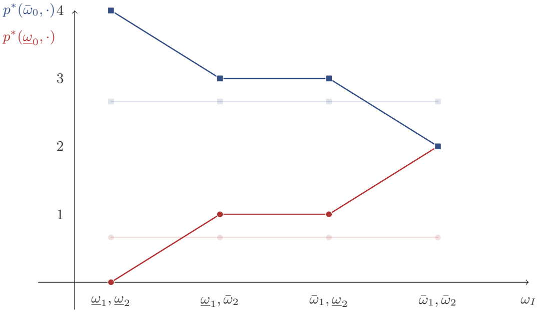
OPTIMAL PRICES p^*



Case 1: Suppose players' private information is poor, $\eta = \gamma < \underline{\phi}$

- Prices are independent of (ω_1, ω_2)
- State $\bar{\omega}_0$ is more valuable than $\underline{\omega}_0$
  - Bets: $q_i^*(1, \underline{\omega}_i)\ell_i^*(0|1, \underline{\omega}_i) = q_i^*(1, \bar{\omega}_i)\ell_i^*(0|1, \bar{\omega}_i) > 0$, for all i
  - $\Rightarrow T^*(\underline{\omega}_0, \omega_I) < 0$ and $T^*(\bar{\omega}_0, \omega_I) > 0$

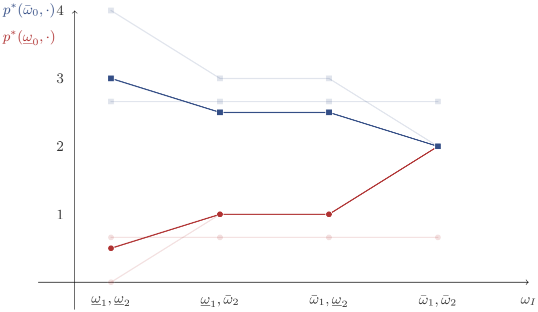
Optimal Prices p^*



Case 2: High informativeness, $\eta = \gamma > \bar{\phi}$

- If firms are pessimistic, pooling becomes harder, larger externality
$$p^*(\underline{\omega}_0,\underline{\omega}_1,\underline{\omega}_2) < v^*(\underline{\omega}_0,\underline{\omega}_1,\underline{\omega}_2) < v^*(\bar{\omega}_0,\underline{\omega}_1,\underline{\omega}_2) < p^*(\bar{\omega}_0,\underline{\omega}_1,\underline{\omega}_2)$$
- If optimistic firms always produce. No externalities
$$p^*(\underline{\omega}_0,\bar{\omega}_1,\bar{\omega}_2) = v^*(\underline{\omega}_0,\bar{\omega}_1,\bar{\omega}_2) = v^*(\bar{\omega}_0,\bar{\omega}_1,\bar{\omega}_2) = p^*(\bar{\omega}_0,\bar{\omega}_1,\bar{\omega}_2)$$
- Bets: $q_i^*(1,\underline{\omega}_i)\ell_i^*(0|1,\underline{\omega}_i) > 0 = q_i^*(1,\bar{\omega}_i)\ell_i^*(0|1,\bar{\omega}_i)$, for all i

Optimal Prices p^*



Case 2: High informativeness, $\eta = \gamma > \bar{\phi}$

- If firms are pessimistic, pooling becomes harder, larger externality

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

- If optimistic firms always produce. No externalities

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

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

The example illustrates another property of p^*,

- While each state can be priced individually, p^* is not in general **additively separable**
- That is, there is no $\hat{p}_0$ and $\hat{p}_i$ for all i, , s.t.

$$p^*(\omega_0, \omega_1, \dots, \omega_n) = \hat{p}_0(\omega_0) + \sum_i \hat{p}_i(\omega_i)$$

Why? v may not be separable in a_i and players interact strategically

Summary

- Price of one entry depends on other entries: $p^*(\omega) \neq v^*(\omega)$
- Price captures dependencies between dimensions of each data entry

prices under privacy

Suppose designer has to incentivize players to disclose their private data

Incentives come directly from how designer commits to use the data

- No monetary transfers (very important, yet distinct issue)
- Role of commitment

Formally, the incentive-compatible use of data means considering as the primal $\mathcal{P}$ an information-design problem **with elicitation**

Question:

- How are prices affected by the need to elicit the data?

Adding elicitation does not alter the mathematical structure of the problem

Problem $\mathcal{P}$ (Bergemann and Morris, 2019)

$$\max_x \quad \sum_{\omega,a} v(a,\omega_0)x(a|\omega)\mu(\omega)$$

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

$$\sum_{a_i,a_{-i},\omega_{-i}} u_i\big(a_i,a_{-i},\omega_0\big)x\big(a_i,a_{-i}|\omega_i,\omega_{-i}\big)\mu(\omega_i,\omega_{-i}) \geq$$

$$\sum_{a_i,a_{-i},\omega_{-i}} u_i\big(\delta_i(a_i),a_{-i},\omega_0\big)x\big(a_i,a_{-i}|\omega_i,\omega_{-i}\big)\mu(\omega_i,\omega_{-i})$$

Adding elicitation does not alter the mathematical structure of the problem

Problem $\mathcal{P}$ (Bergemann and Morris, 2019)

$$\max_x \sum_{\omega,a} v(a,\omega_0)x(a|\omega)\mu(\omega)$$

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

$$\sum_{a_i,a_{-i},\omega_{-i}} u_i\big(a_i,a_{-i},\omega_0\big)x\big(a_i,a_{-i}|\omega_i,\omega_{-i}\big)\mu(\omega_i,\omega_{-i}) \geq$$

$$\sum_{a_i,a_{-i},\omega_{-i}} u_i\big(\delta_i(a_i),a_{-i},\omega_0\big)x\big(a_i,a_{-i}|\omega_i',\omega_{-i}\big)\mu(\omega_i,\omega_{-i})$$

Data-Pricing Problem with Elicitation

Designer chooses, for each player i and ω_i, a pair:

$$\left(\hat{\ell}_i(\cdot|\omega_i), \hat{q}_i(\omega_i)\right) \in \Delta(\Omega_i \times D_i) \times \mathbb{R}_{++}$$

and solves:

Problem $\mathcal{D}$ (Data-Pricing Problem)

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

s.t. for all ω,

$$p(\omega) = \max_{a \in A} \left\{ v(a, \omega_0) + \sum_i T_{\hat{\ell}_i, \hat{q}_i}(a, \omega) \right\}$$

Where transfer function $T_{\hat{\ell}_i, \hat{q}_i}$ is now a richer object

Data-Pricing Problem with Elicitation

Data-Pricing problem with vs without elicitation:

- Identical objective and similar pricing formulas with richer set of bets $(\hat{\ell}, \hat{q})$ against players
- Designer can win against player when:
  1. Deviating from obedience is ex-post beneficial (as in before)
  2. Deviating from truth telling is ex-post beneficial (new)
  3. Both (new)

Directions:

- The price of a state must incorporate difficulty to truthfully eliciting it: new externalities
- Comparing prices under omniscient and under elicitation offers insights into effects of IC on value of data:

  E.g. how price of data is affected by privacy protection

back to example

Cournot: Prices p^* with Elicitation

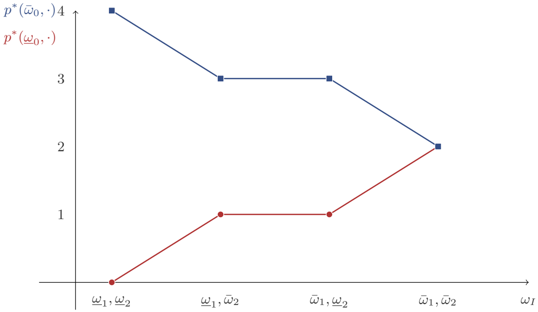
Revisit oligopoly example with elicitation: Fix some $\eta = \gamma > \bar{\phi}$

Clearly, value of data V^* decreases with elicitation. What about prices?

Cournot: Prices p^* with Elicitation

Revisit oligopoly example with elicitation: Fix some $\eta = \gamma > \bar{\phi}$

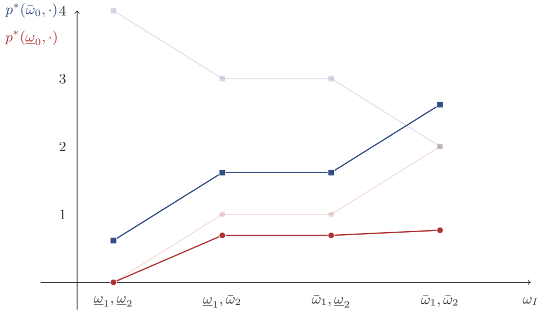
Clearly, value of data V^* decreases with elicitation. What about prices?



Cournot: Prices p^* with Elicitation

Revisit oligopoly example with elicitation: Fix some $\eta = \gamma > \bar{\phi}$

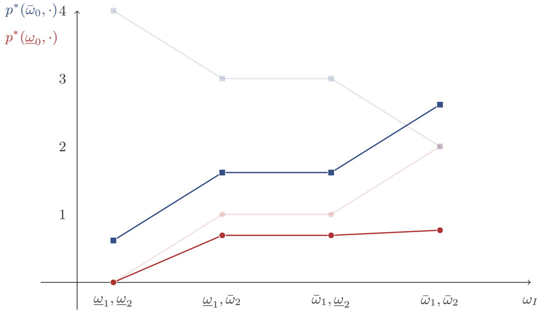
Clearly, value of data V^* decreases with elicitation. What about prices?



Cournot: Prices p^* with Elicitation

Revisit oligopoly example with elicitation: Fix some $\eta = \gamma > \bar{\phi}$

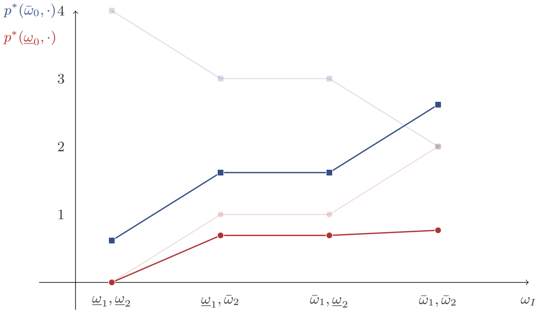
Clearly, value of data V^* decreases with elicitation. What about prices?



1. The need for elicitation induces qualitative change in $p(\bar{\omega}_0, \omega_I)$
   - $\bar{\omega}_i$ has incentive to mimic $\underline{\omega}_i$ to receive better information
   - If state induces temptation to lie, it suffers a negative externality
   - Recommendation x^* distorted to make mimicking less attractive

Revisit oligopoly example with elicitation: Fix some $\eta = \gamma > \bar{\phi}$

Clearly, value of data V^* decreases with elicitation. What about prices?



2. Despite V^* is lower, some prices increase: $p^*(\bar{\omega}_0, \bar{\omega}_1, \bar{\omega}_2)$
   - Information rent for $(\bar{\omega}_0, \bar{\omega}_1, \bar{\omega}_2)$ which is paid by other states

conclusion

SUMMARY

A theory of how to price entries of a dataset to reflect their values

- Basic insight: leverage duality with information design, how to optimally use the data

Our **preliminary analysis** of the properties of the price of data reveals:

- Prices account for externalities across states
- ...and between dimensions of each data entry
- Privacy protection significantly affects prices and can even increase the price of some data entries