Manipulative consumers

Michael Richter

Baruch College, CUNY and

Royal Holloway, University of London

Nikita Roketskiy

University College London

Christmas Microeconomic Theory Conference in Bonn
December 2023

Research question

- Sellers use consumer data for pricing (and product design)
- Consumers can manipulate their records at a cost

How much is data worth to the seller?

- Value of data $\sim$ price dispersion
- Richer data is worth less

Related works

Manipulable data:

- Ball (2021), Frankel and Kartik (2019, 2022) - inference from manipulable data
- Eliaz and Spiegler (2021), Caner and Eliaz (2021) - IC estimators
- Deneckere and Severinov (2017), Severinov and Tam (2019), Perez-Richet and Skreta (2022), Dana, Larsen and Moshary (2023), Tan (2023), Moreno de Barreda and Safonov (2023) - m./test design
- Bonatti and Cisternas (2019), Bhaskar and Roketskiy (2021) - consumer history and price discrimination

Market segmentation:

- Hidir and Vellodi (2020) - IC market segmentation
- Liang and Madsen (2021) - profiling and incentivizing effort
- Eilat, Eliaz and Mu (2020) - restricting informativeness of a price discrimination

Value of data/Privacy:

- Dubé and Misra (2021) - value of personalized pricing
- Bergemann and Bonatti (2015), Bergemann, Bonatti and Smolin (2018) Segura-Rodriguez (2019) - data brokers
- Bonatti, Huang and Villas-Boas (2023)

Consumers

- $C = [0, 1]$ - cont. of consumers
- $\tau : C \to \{t_\ell, t_h\}$ - valuation
- $s(i, q) = \tau(i)q - \frac{q^2}{2}$, - surplus of $i \in C$
- $d = t_h - t_\ell$

Monopolistic seller

- produces a variety of vertically differentiated products, quality q (at “zero” cost)
- menu pricing $p(q)$
- can condition the menu on observables $\alpha(i)$
- no commitment to the data practices

Consumer data

- $\omega : C \to \{0,1\}^K$ - consumer attributes
- $\alpha : C \to \{0,1\}^K$ - consumer data
- $\alpha(i)$ is chosen at a cost $\frac{\|\alpha(i)-\omega(i)\|}{K}c$
- $\tau(i)$ is correllated with $\omega(i)$
- $m(\cdot)$ is measure of ℓ-consumers
- $n(\cdot)$ is measure of h-consumers
- two assumptions (A1, A2) on these measures

Market segments

- Seller uses consumer data to price the products: a consumer faces prices that depend on her attributes.
- A combo of 2nd and 3rd degree price discrimination:
  - Each market segment $S \in \mathfrak{S}$ gets its own optimal menu.
  - Firm estimates the consumer demand within the segment.
- Market segment labels $\mathfrak{S}$
- Firm regresses attributes to market segments:

$$R : \mathfrak{A} \to \mathfrak{S}$$

Optimal menu in segment S

Demand statistics:

$$h(S) = \frac{n}{m}(\{i \in C : R(\alpha(i)) = S\})$$

Consumer surplus (per H-consumer):

$$U_h(S) = \max\{0, 2d(t_\ell - h(S)d)\}$$

Profit (per ℓ-consumer in S):

$$\rho(S) = h(S)(t_\ell + d)^2 + [\max\{0, t_\ell - h(S)d\}]^2$$

A1

$$\bar{h} \in \left[\frac{c}{2d^2}, \frac{t_\ell}{d} - \frac{c}{2d^2}\right].$$

Value of consumer data

Aggregating profit across segments

Proposition:

$$\pi(S) - \pi^* = d^2 \underbrace{\sum_{\mathbf{a}} m(\mathbf{a}) \left[h(R(\mathbf{a})) - \bar{h}\right]^2}_{\text{Var}\left[h(S(\cdot))\right]} = \frac{1}{4} \sum_{\mathbf{a}} m(\mathbf{a}) \left[p_h(R(\mathbf{a})) - \bar{p}_h\right]^2$$

Corollary: $h(R(\mathbf{a}))$ is a mean-preserving contraction of $h(\mathbf{a})$ hence use all info

$$R(\mathbf{a}) = \mathbf{a}$$

Value = explained variation

- Seller does a non-parametric regression of h on $\mathbf{a}$.
- Part of variation in "premium" demand explained by the data:

$$\sum_{\mathbf{a}} m(\mathbf{a}) \left[h(\mathbf{a}) - \bar{h} \right]^2$$

is the value of consumer data for the seller.

Attributes

Each consumer is endowed with a vector of K binary attributes (personal data):

$$\omega(i) \in \{0,1\}^K$$

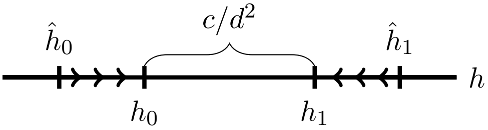
Consumer can change the values of any attributes at a cost. If consumer i sets her attributes to $\alpha(i) \in \{0,1\}^K$ she pays

$$\frac{\|\alpha(i) - \omega(i)\|}{K} c.$$

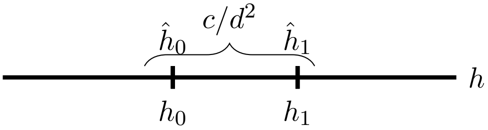
Consumers manipulate their attributes privately before they see the prices.

Incentives to manipulate data, “no-arbitrage constraints”

mixed strategy



no changes to attributes



For any $\mathbf{a}, \mathbf{b} \in \{0,1\}^K$:

$$|h(\mathbf{a}) - h(\mathbf{b})| \leq \frac{c}{d^2} \frac{||\mathbf{a} - \mathbf{b}||}{K}$$

Value of consumer data

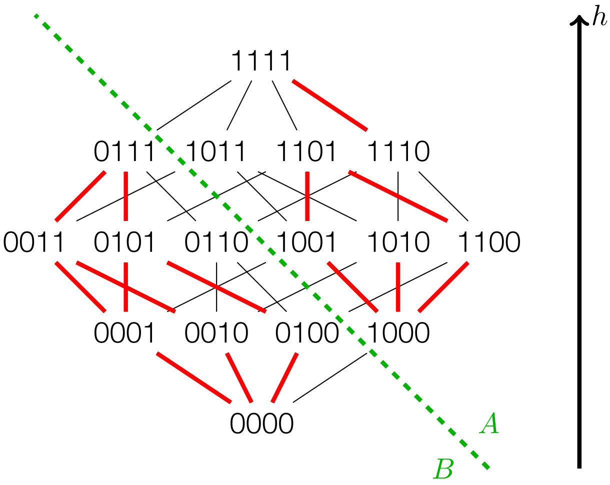
- Value depends on correlation between data and type
- Observed data depends on consumer attributes
- We look at the **seller's best-case scenario:**

$$\max_{h(\cdot)} \sum_{\mathbf{a}} m(\mathbf{a})[h(\mathbf{a}) - \bar{h}]^2$$

$$s.t. \sum_{\mathbf{a}} m(\mathbf{a})[h(\mathbf{a}) - \bar{h}] = 0$$

$$|h(\mathbf{a}) - h(\mathbf{b})| \leq \frac{c}{d^2} \frac{||\mathbf{a} - \mathbf{b}||}{K}, \text{ for all } \mathbf{a}, \mathbf{b} \in \{0, 1\}^K$$

Binding constraints



$$\sum_{\mathbf{a}} m(\mathbf{a})[h(\mathbf{a}) - \bar{h}]^2 =$$

$$\sum_{\mathbf{a}\in A} m(\mathbf{a})[h(\mathbf{a}) - h_A]^2 +$$

$$\sum_{\mathbf{a}\in B} m(\mathbf{a})[h(\mathbf{a}) - h_B]^2 +$$

$$m(A)[h(A) - \bar{h}]^2 +$$

$$m(B)[h(B) - \bar{h}]^2$$

Lemma The graph of binding constraints is connected.

New attributes, new information

A2

There exist marginal probabilities $\mu_i : \{0,1\} \to \mathbb{R}_+, i = 1, .., K$, such that for any vector of attributes **a** :

$$m(\mathbf{a}) = \bar{m} \prod_{i=1}^{K} \mu_i(\mathbf{a}_i)$$

.

This assumption allows for induction on the number of attribiutes.

The main result

If A1 and A2, then the value of consumer data is

$$D = \frac{1}{K}\bar{m}\left[\frac{c}{2d}\right]^2 \frac{\sum_{j=1}^{K} \mu_j(0)\mu_j(1)}{K}$$

Scope for manipulation

The spread of ℓ-consumers across attribute values:

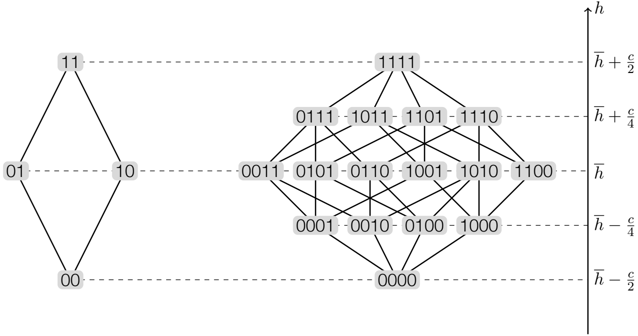
$$\frac{\sum_{j=1}^{K} \mu_j(0)\mu_j(1)}{K}$$

If ℓ-consumers are concentrated, it is easy for the h-consumers to blend in.

The effect of increasing K

$$\frac{1}{K}\bar{m}\left[\frac{c}{2d}\right]^2$$

$$\lim_{K\to\infty}\frac{1}{2^K}\binom{K}{K/2} = 1$$



Opaque use of data

As in Frankel and Kartik (2019, 2022) and Ball (2021):

If firm can **commit** to using single **unspecified** attribute then the value of consumer data is

$$D' = \bar{m} \left[\frac{c}{2d} \frac{\sum_{j=1}^{K} \sqrt{\mu_j(0)\mu_j(1)}}{K} \right]^2$$

Conclusion

- Value of information is measured by the spread in demand (or prices)
- Adding new (non-duplicating) variables to the data, increases both informational content and manipulation opportunities–the latter erodes value