Manipulative consumers

Michael Richter

Royal Holloway, University of London

Baruch College

Nikita Roketskiy

University College London

23rd CEPR-JIE Conference on Applied Industrial Organization

Question

- Seller uses consumer data to price its products
- Consumers can manipulate their records at a cost

Question

- Seller uses consumer data to price its products
- Consumers can manipulate their records at a cost

How much is data worth to the seller?

Question

- Seller uses consumer data to price its products
- 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), Dana, Larsen and Moshary (2023) - 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) - mechanism/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:

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

The model

Demand and supply

Consumers:

- C - cont. of consumers
- $\tau : C \to \{t_\ell, t_h\}$ - valuation
- $\alpha : C \to \{0,1\}^K$ - data
- $m(C)$ is measure of L
- $n(C)$ is measure of H
- $d = t_h - t_\ell$

Demand and supply

Consumers:

- C - cont. of consumers
- $\tau : C \to \{t_\ell, t_h\}$ - valuation
- $\alpha : C \to \{0,1\}^K$ - data
- $m(C)$ is measure of L
- $n(C)$ is measure of H
- $d = t_h - t_\ell$

Seller:

- menu pricing (q, p)
- menu cond. on data

Demand and supply

Consumers:

- C - cont. of consumers
- $\tau : C \to \{t_\ell, t_h\}$ - valuation
- $\alpha : C \to \{0,1\}^K$ - data
- $m(C)$ is measure of L
- $n(C)$ is measure of H
- $d = t_h - t_\ell$

Transaction:

- $s(i,q) = \tau(i)q - \frac{q^2}{2}$, where $i \in C$
- consumer gets $s(i,q) - p$
- seller gets p

Seller:

- menu pricing (q,p)
- menu cond. on data

Demand and supply

Consumers:

- C - cont. of consumers
- $\tau : C \to \{t_\ell, t_h\}$ - valuation
- $\alpha : C \to \{0,1\}^K$ - data
- $m(C)$ is measure of L
- $n(C)$ is measure of H
- $d = t_h - t_\ell$

Seller:

- menu pricing (q, p)
- menu cond. on data

Transaction:

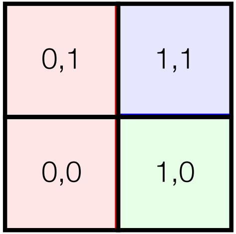
- $s(i,q) = \tau(i)q - \frac{q^2}{2}$, where $i \in C$
- consumer gets $s(i,q) - p$
- seller gets p

$\boxed{\tau(i) \text{ is correllated with } \alpha(i)}$

Market segments

- Seller uses consumer data to price the products: a consumer faces prices that depend on her attributes.
- Market segment labels $\mathfrak{S}$
- 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.
- Firm regresses attributes to market segments:

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



Optimal menu in segment S

Demand suff. statistics:

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

Optimal menu in segment S

Demand suff. statistics:

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

Consumer surplus (per H-consumer):

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

Optimal menu in segment S

Demand suff. 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):

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

Optimal menu in segment S

Demand suff. 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):

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

Assume:

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

Value of consumer data

Proposition:

$$\pi(S) - \pi^* = \frac{d^2}{2} \sum_{\mathbf{a}} m(\mathbf{a}) \left[h(R(\mathbf{a})) - \bar{h} \right]^2$$

Value of consumer data

Proposition:

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

Value of consumer data

Proposition:

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

Value of consumer data

Proposition:

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

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 **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):

$$\mathbf{a} \in \{0, 1\}^K$$

Consumer can change the values of any k attributes at a cost

$$\frac{k}{K}c.$$

Consumers manipulate their attributes privately before they see the prices.

Each attribute carries its own information

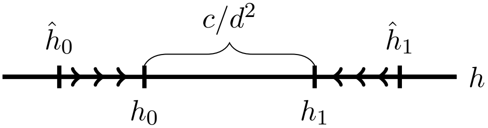
A1: there exist marginal probabilities $\mu_i : \{0,1\} \to \mathbb{R}_+, i = 1, .., K$, such that for any vector of attributes **a** :

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

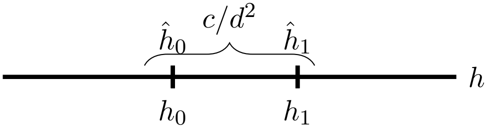
.

Incentives to manipulate data, “no-arbitrage constraints”

mixed strategy



no changes to attributes



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

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

The two opposing forces

- Correlation is valued by the seller,

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

The two opposing forces

- Correlation is valued by the seller, but consumers respond by doing arbitrage

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

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

The two opposing forces

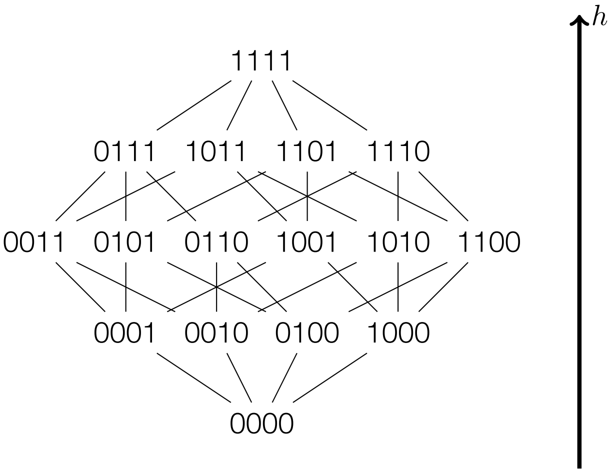
- Correlation is valued by the seller, but consumers respond by doing arbitrage
- We look at the **seller's best-case scenario:**

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

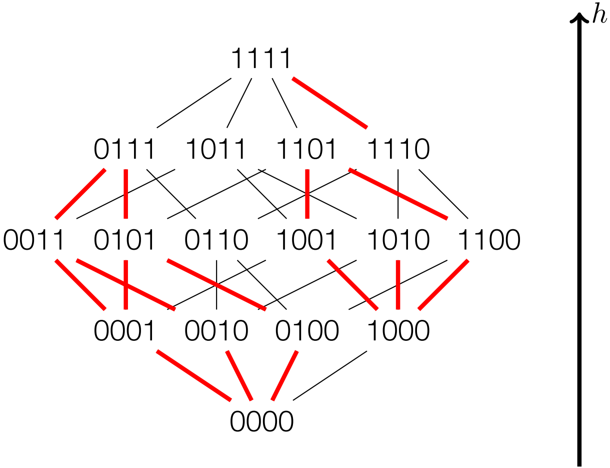
$$s.t. \sum_{\mathbf{a}} m(\mathbf{a})[h(\mathbf{a}) - \bar{h}] = 0$$

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

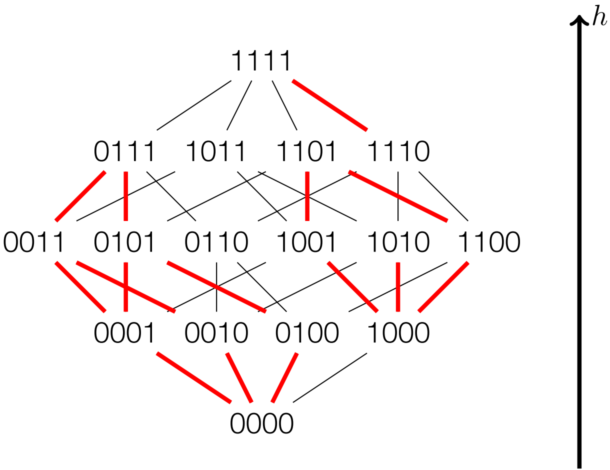
Binding constraints $|h(\mathbf{a}) - h(\mathbf{b})| \leq \frac{c}{d^2} \frac{||\mathbf{a}-\mathbf{b}||}{K}$



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

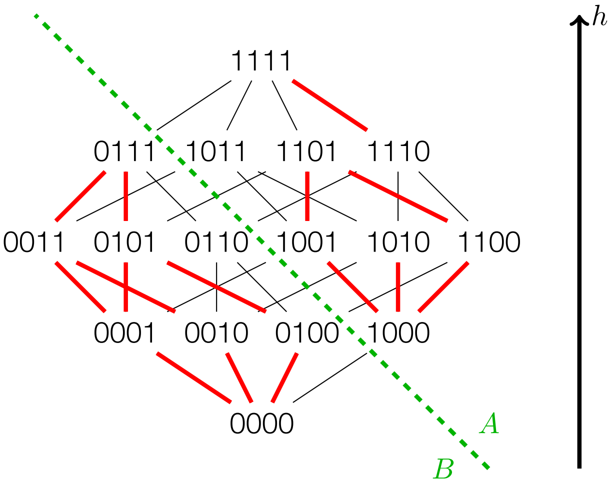


Binding constraints $|h(\mathbf{a}) - h(\mathbf{b})| \le \frac{c}{d^2}\frac{||\mathbf{a}-\mathbf{b}||}{K}$



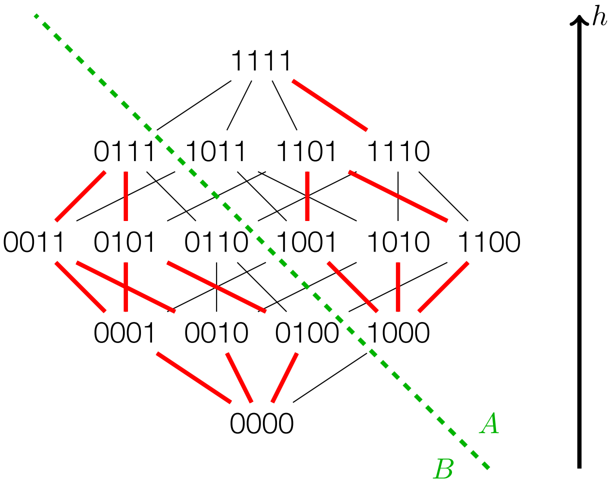
Lemma The graph of binding constraints is connected.

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



Lemma The graph of binding constraints is connected.

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



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

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

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

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

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

Lemma The graph of binding constraints is connected.

Vanishing value

Proposition:

If A1, then the value of consumer data is

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

Vanishing value

Proposition:
If A1, then the value of consumer data is

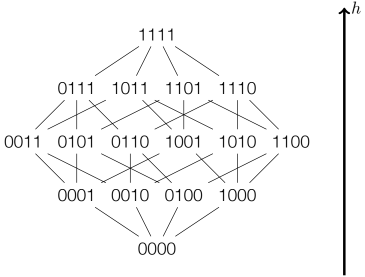
$$D = \frac{1}{K}\bar{m}\left[\frac{c}{2d}\right]^2 \frac{\sum_{j=1}^{K} \mu_j(0)\mu_j(1)}{K}$$

- D vanishes with $K \to \infty$,
- Increases in the average variance of attributes for low types.

Vanishing value

Proposition:
If A1, then the value of consumer data is

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



- D vanishes with $K \to \infty$,
- Increases in the average variance of attributes for low types.

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

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

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

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

- $D' >> D$
- Transperancy of data usage erodes value

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