

# Manipulative consumers

Michael Richter

Baruch College, CUNY and  
Royal Holloway, University of London

Nikita Roketskiy

University College London

IIOC 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) - 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

# Demand and supply

Consumers:

- ▶  $C$  - cont. of consumers
- ▶  $\tau : C \rightarrow \{t_\ell, t_h\}$  - valuation
- ▶  $\alpha : C \rightarrow \{0, 1\}^K$  - 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  $p$

$\tau(i)$ is correlated with $\alpha(i)$
--

# 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:

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

## 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  $\ell$ -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

Aggregating profit across segments

**Proposition:**

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

**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:

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

is the value of consumer data for the seller.



# Attributes

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

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

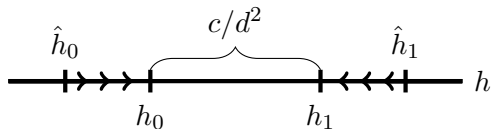
Consumer can change the values of any  $k$  attributes at a cost

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

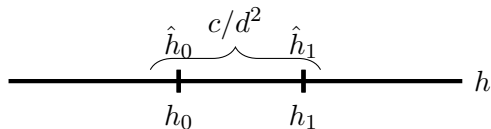
Consumers manipulate their attributes privately before they see the prices.

# Incentives to manipulate data, “no-arbitrage constraints”

mixed strategy



no changes to attributes



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

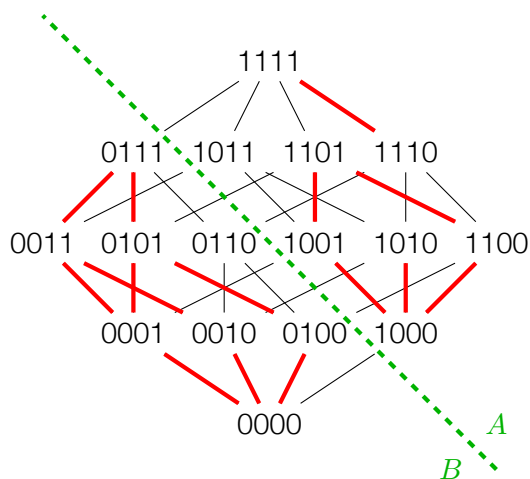
$$|h(\mathbf{a}) - h(\mathbf{b})| \leq \frac{c}{d^2} \frac{\|\mathbf{a} - \mathbf{b}\|}{K}$$

## Value of consumer data

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

$$\begin{aligned} \max_{\{h(\cdot)\}} \quad & \sum_{\mathbf{a}} m(\mathbf{a}) [h(\mathbf{a}) - \bar{h}]^2 \\ \text{s.t.} \quad & \sum_{\mathbf{a}} m(\mathbf{a}) [h(\mathbf{a}) - \bar{h}] = 0 \\ & |h(\mathbf{a}) - h(\mathbf{b})| \leq \frac{c}{d^2} \frac{\|\mathbf{a} - \mathbf{b}\|}{K}, \text{ for all } \mathbf{a}, \mathbf{b} \in \{0, 1\}^K \end{aligned}$$

## 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

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

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

.

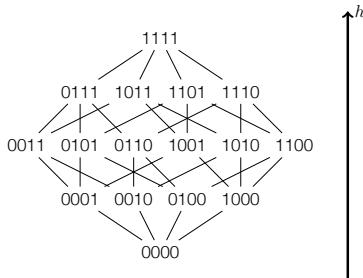
# Vanishing value

## 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  $K \rightarrow \infty$ ,
  - ▶ 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

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

- 
- ▶  $D' \gg D$
  - ▶ Transparency 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