Data Broker competition and downstream market entry

17th CRESSE Conference, Rhodes, 30 June-2 July 2023

Laura Abrardi, Carlo Cambini and Flavio Pino

Politecnico di Torino

carlo.cambini@polito.it

Motivation

- **Data Brokers (DB)** track consumers online, collect data, and sell the information to firms
- Thousands of data segments for nearly every consumer (FTC 2014)
- The DB's sector is a concentrated industry whose revenue is estimated at USD 200 billion (FTC, 2014; Crain, 2018)
- By controlling the provision of data, DBs can raise entry barriers in downstream retail markets, causing consumer harm
- "Virtually unregulated system"
- How does **competition between DBs** affect **downstream entry and consumer surplus**?

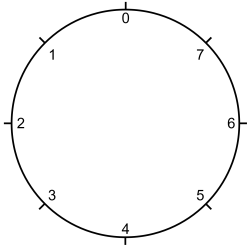
This paper

- **Two competing DBs** sell data to downstream firms, which use data to price discriminate
- We explore two dimensions of data:
  - Quantity of data (extensive margin)
  - Accuracy of data (intensive margin)
- We study how competition between DBs affects **downstream entry** and **welfare**
- We consider different competitive scenarios:
  - **Vertical differentiation**: DBs have the same quantity of data, but of different accuracy
  - **Synergies between datasets**: DBs have data about different (possibly overlapping) sets of consumers

Literature

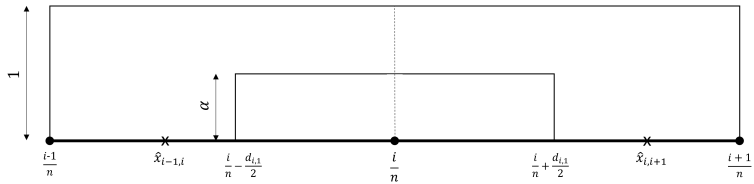
- The literature on the market for information has mostly focused on **monopolistic DBs** or **exogenous data acquisition**
  - Exogenous entry in duopolistic (Montes et al., 2019; Bounie et al., 2021) or oligopolistic markets (Delbono et al., 2021),
  - Salop model with exogenous data (Taylor and Wagman, 2014)
  - Abrardi et al., 2022: oligopoly, endogenous entry
- Few studies on **DB competition**, and assume exogenous number of downstream firms:
  - Competition between DBs increases **consumer surplus** (Bounie et al., 2021)
  - However, **synergies** between DBs' datasets influence DB's incentive to data sharing (Gu et al., 2022)
- We consider a Salop model where data are provided endogenously by competing DBs

The model: Consumers, firms



- Consumers (normalised to 1) are uniformally distributed on a circular city of length 1
- n (endogenous) firms indexed by $i \in \{0, 1, 2, ..., n-1\}$, locations $\frac{i}{n}$
- Marginal costs are 0, and firms pay a cost F if they enter

The model: DBs



- Two DBs, DB_1 and DB_2. DB_k offers data $d_{i,k} \in [0,1]$ to firm i, centered on the firm's location
- Data allows the firm to price discriminate on the market segment $d_{i,k}$
- Data sold by DB_1 and DB_2 have different quality: $d_{i,1}$ allows to identify consumers with probability $\alpha \in [0,1]$; $d_{i,2}$ with probability $\beta\alpha, \beta \in [0,1)$
  - α is the **information accuracy**, β is related to the level of **DB vertical differentiation** (Belleflamme et al., 2020)
- Baseline scenario: DBs' datasets **completely overlap**

Timing

- Stage 1. Firms enter the market and pay the fixed cost F
- Stage 2. DBs simultaneously offer data through Take It Or Leave It (TIOLI) offers. The offer of any DB_k consists of a data partitions $d_{i,k}$ and a data price
- Stage 3. Firms choose which offer they want to accept
- Stage 4. Firms set basic prices $p^{\mathrm{B}}_{i,k}$ for the unidentified consumers
- Stage 5. Firms set tailored prices ${p_{i,k}}^{\mathrm{T}}(x)$ for the consumers identified in position x by the data partition they have purchased

Firms' profits

$$\begin{aligned}\pi_{i,1} &= \alpha \int_{\frac{i}{n}-\frac{d_{i,1}}{2}}^{\frac{i}{n}+\frac{d_{i,1}}{2}} p^{\mathrm{T}}_{i,1}(x)\, dx + (1-\alpha) d_{i,1} p^{\mathrm{B}}_{i,1} + \\ &+ p^{\mathrm{B}}_{i,1}\left(\widehat{x}_{i,i+1} - \widehat{x}_{i-1,i} - d_{i,1}\right) - F\end{aligned}$$

Data have a non-monotonic effect on firms' profits:

- Surplus extraction effect $(+)$
  - The larger the partition, the more consumers are identified, and offered tailored prices
  - The tailored price allows the firm to extract the consumer surplus, increasing firm profits
- Competition effect $(-)$
  - The larger the partition, the more distant are the unidentified consumers, which are offered the basic price, which in turn decrease with data.

Equilibrium prices

In equilibrium, the tailored prices match the rivals' basic prices in utility levels:

$$p_{i,1}^{T*}(x) = \begin{cases} p_{i-1,1}^{B*} + 2tx - \frac{t}{n}(2i-1) & \text{for } x \in [\frac{i}{n} - \frac{d_i}{2}, \frac{i}{n}] \\ p_{i+1,1}^{B*} - 2tx + \frac{t}{n}(2i+1) & \text{for } x \in [\frac{i}{n}, \frac{i}{n} + \frac{d_i}{2}] \end{cases}$$

Where

$$p_{i,1}^{B*} = \frac{t}{n} - 2t \sum_{j=0}^{n-1} \alpha d_{i+j,1} a_j$$

- Note that $p_{i,1}^{B*}$ decreases with α and with any partition $d_{i,1}$.

DBs' offers

- DB_1 has a competitive edge (higher accuracy) over DB_2
  - DB_1 sets his prices equal to firms' willingness to pay for data, namely difference in profits between buying from DB_1 or DB_2: $\pi_{i,1} - \pi_{i,2}$
  - DB_2 exerts competitive pressure on DB_1
- All firms accept the offer of DB_1

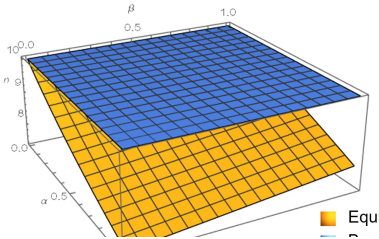
How much data is sold in equilibrium?

- Both DBs offer a partition of the same size d^* to all firms
- d^* is such that some consumers are unidentified, to temper the competitive effect (different from Taylor&Wagman, 2014)
- d^* is increasing in β and decreasing in α
  - Data are strategic complements
  - With more accurate information, less data are required to obtain the same surplus extraction

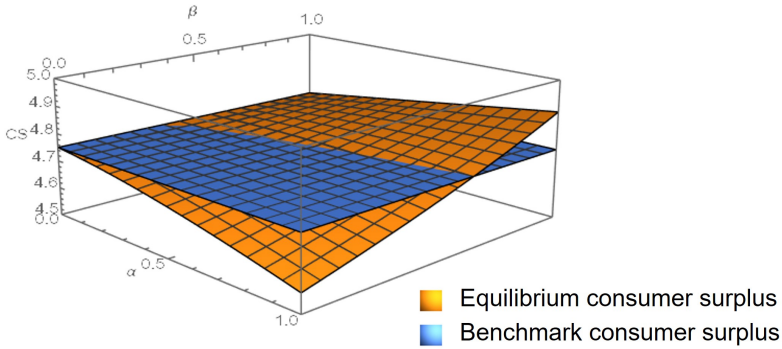
Entry

The number of firms is obtained by imposing the free-entry condition

- Entry is always **lower** than in Salop
  - Data increase downstream competition which dissipates profits
  - Firms pay to purchase data
- Entry increases in β, as more symmetric DBs set a lower data price
- Entry decreases in α, as more accurate data increase downstream competition, reducing profits and the scope for entry;

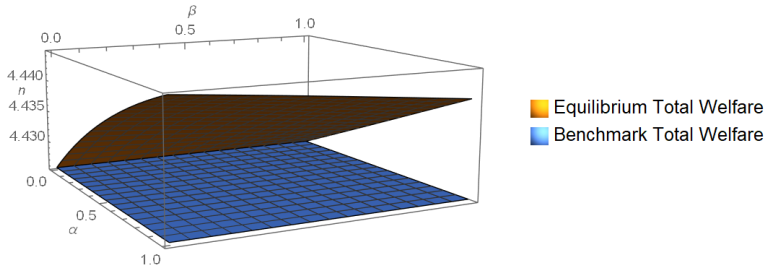


Consumer surplus



- Benchmark: $\alpha = 0$ (no data, no targeting)
- $\beta \to 1$: perfect DBs' competition
- For any level of α, there exists a $\bar{\beta}$ such that $CS^* \geq CS^{BENCH}$ iff $\beta > \bar{\beta}$

Welfare



- **Total welfare** is always higher than in the absence of data ($\alpha = 0$)
- When β increases, welfare decreases due to excessive entry
- Welfare increases with α

Synergic datasets

- Purchasing data from both DBs grants an accuracy γ:
  - If $\gamma > \alpha + \beta\alpha$, **super-additive** data
  - If $\alpha < \gamma \leq \alpha + \beta\alpha$, **sub-additive** data

- In equilibrium both DBs sell data
  - If data are **super-additive**, DBs set prices to completely extract frims' WTP for data. Both DBs sell same sized partitions $d^{*sup} < d^*$
  - If data are **sub-additive**, DB_2 offers bigger partitions than DB_1
- Entry is lower due to the higher price of data

Synergic datasets - Welfare implications

- *TW* **increases**
- The effect on *CS* is ambiguous:
- If data are **super-additive**, the DBs coordinate to price as a monopolist with an accuracy γ: *CS* always **decreases**

- If data are **sub-additive**, CS depends on the **amount of proprietary data** of each DB:
  - A higher γ implies a higher share of proprietary data, which **increases the data price** and **reduces entry**
  - For any level of β, there exists a threshold $\bar{\gamma} \in [\alpha, \alpha + \beta\alpha]$ after which $CS^* < CS^{BENCH}$

Conclusions

- Competition between DBs increases entry vs. the case of a monopolistic DB, but does not eliminate the entry barrier relative to the Salop model
- More accurate information reduces entry due to a higher price of data
- Competition between DBs increases consumer surplus
- If DB competition sufficiently intense, consumer surplus is higher than in the case data are unavailable
- Synergies between DBs' datasets reduces competition between DBs