Markets for Information

Dirk Bergemann    Alessandro Bonatti    Tan Gan
University of Pennsylvania

December 2018
Introduction

- markets for information ever more relevant to economic activity and welfare
- selling information → providing access to a database (e.g., Bloomberg’s financial data)
- predictions, ratings, recommendations, customized products and services
- challenges and opportunities: surplus creation, market power, distortions, privacy
- at best, we only understand individual aspects of information markets
Introduction

- markets for information ever more relevant to economic activity and welfare
- selling information → providing access to a database (e.g., Bloomberg’s financial data)
- predictions, ratings, recommendations, customized products and services
- challenges and opportunities: surplus creation, market power, distortions, privacy
- at best, we only understand individual aspects of information markets
- our goal: provide a comprehensive perspective
- today: a simple model integrating distinct elements
Buying and Selling Information

- large data brokers (Acxiom, Nielsen, Oracle)
- multiproduct platforms (Amazon, Facebook, Google)

**What is for sale?**

<table>
<thead>
<tr>
<th>Who identifies prospect?</th>
<th>Data Seller</th>
<th>Data Buyer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who</td>
<td>Only Information</td>
<td>Access to Consumer</td>
</tr>
<tr>
<td>Data Seller</td>
<td>original lists</td>
<td>sponsored search</td>
</tr>
<tr>
<td>Data Buyer</td>
<td>data appends</td>
<td>retargeting</td>
</tr>
</tbody>
</table>

- new list of prospects vs. additional info about buyer’s existing customers?
- direct vs. indirect sale of information (Admati & Pfleiderer 1990)
- can the buyer reach the consumer without the data seller?
Direct Sale of Information

- data appends = supplemental information about existing / potential customers
- Nielsen Catalina Solutions and Oracle Datalogix connect an individual’s offline and online purchases with the digital media they consume.
- Oracle ID-Graph tracks firms’ customers across several devices
- Email Intelligence by TowerData appends information to email addresses
- Credit reporting agencies: people-specific queries for risk-mitigation, e.g., Equifax Undisclosed Debt Monitoring tracks new negative information between loan approval and closing.
- ACT services (lookup of individual records).
Celebrate together, with Evite.

CREATE AN INVITATION

- Charity/Fundraisers
- The Big Game
- Valentine’s Day
- Kids’ Activities
- Viewing Party

BROWSE PARTY IDEAS

- This California Guacamole Hummus Dip is the Best of Both Worlds
- Satisfy Your Sweet Tooth with this After-Dinner Coffee Bar
- Dip Into this Chicago Deep Dish Pizza Dip
Indirect Sale of Information

Market for sponsored-search advertising:

- consumer conducts a Google / Amazon search;
- advertisers buy a slot on a keyword search results page;
- advertising slots *de facto* bundled with *original list* of qualified *eyeballs*;
- search engine could sell data about individual searches directly,
- but leverages the data to improve quality of their advertising product.

Far larger market than direct sales...
Where your ads may appear

Sponsored Products
On Amazon devices
A Few Aspects We Know

- how a monopolist should sell information products to heterogeneous buyers

- what are the limits of price discrimination due to information

- how data brokers (and advertising sellers) source consumer-level information.

- how to design and distribute information
Many Aspects We Don’t Know

- as awareness of data-sharing practices increases, users must be compensated (through payments or other terms of service) to reveal their information
- how much must brokers and platforms compensate consumers?
- when should they sell advertisers exclusive access to consumers?
- what determines market structure in the information sector?
- what are the drivers of the equilibrium price of information?
- what (related) implications for competition policy?
More Things We Don’t Know

- what is the size advantage of information platforms, natural monopoly?
- what does size relate to: "knowing many things about some people" or "knowing some things about many people"?
- what is the "markup", price over marginal cost, for information
- what is role of "export and share" your data
- what is the role of combining data, what is role of merging information aggregators
Model of Data Intermediary

- a single data broker, $n$ consumers, a single seller (firm)
Consumer and Firm

- each consumer $i$ has willingness-to-pay $w_i$:
  \[ w_i = \theta + \theta_i \]
  with common and idiosyncratic component: $\theta$ and $\theta_i$

- linear quadratic utility function
  \[ u(w_i, q_i, p) = w_i q_i - pq_i - \frac{1}{2} cq_i^2 \]

- seller maximizes expected revenue
  \[ \pi(p) = \mathbb{E} \left[ p \sum_i q_i \right] \]
each consumer $i$ has private and imperfect information:

$$s_i = w_i + \varepsilon + \varepsilon_i$$

with common and idiosyncratic shock: $\varepsilon$ and $\varepsilon_i$

(symmetric) Gaussian environment

$$\begin{pmatrix} \theta \\ \theta_i \\ \varepsilon \\ \varepsilon_i \end{pmatrix} \sim N \begin{pmatrix} \mu \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\theta^2 & 0 & 0 & 0 \\ 0 & \sigma_{\theta_i}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\varepsilon}^2 & 0 \\ 0 & 0 & 0 & \sigma_{\varepsilon_i}^2 \end{pmatrix}$$
Data Broker

- bilateral contracting with consumer and firm
- data broker offers fee $f_i (I_i)$ to consumer $i$ as function of transmitted information $I_i$:

$$f_i : I_i \rightarrow \mathbb{R}_+$$

- $I_i$ can simply be the entire information of consumer $i$ or some, possibly noisy, statistic of his information
- data broker shares information about consumers with firm and asks fee $g (I)$ as function of transmitted information:

$$g : I \rightarrow \mathbb{R}_+$$
Timing

1. data broker offers *ex ante* payments to consumer $i$ and firm: $\{f_i\}, g$
2. consumers and firm accept or reject offer
3. each consumer $i$ observes $s_i$ and shares it with data broker (given data sharing agreement)
4. broker transmit information to firm
5. firm charges a unit price $p$ and each consumer $i$ purchases quantity $q_i$
A Few Qualifying Comments

- direct sale of information: analysis extends to indirect sale of information, say sponsored search
- one firm: analysis extends to many competing firms
- one data broker: analysis extends to competing data brokers
Information and Price Discrimination

- information allows third degree price discrimination by seller
- with linear demand, additional information reduces social welfare (Robinson (1933), Schmalensee (1981))
- equilibrium price become positively correlated with common value $\theta$
  \[ \Rightarrow \text{quantity covaries less with value } \theta \]
  \[ \Rightarrow \text{profits rise, but total surplus falls} \]
with $n > 1$, each consumer can only demand compensation for marginal information he provides

as private information of agent $i$ becomes more correlated with agent $j$’s, her marginal contribution to the broker’s information is lessened

scope for profitable intermediation is determined by the relative variance of the idiosyncratic and the common component in demand
Basic Intermediation

- suppose broker buys and transmits signals without noise

Proposition

*There exists a threshold* \( \bar{n} \) *such that the broker obtains positive profits if and only if the number of consumers is* \( n > \bar{n} \).

- information is traded even when it decreases social surplus
- threshold \( \bar{n} \) is (weakly) increasing in \( \sigma_{\theta_i}^2 \) and is decreasing in \( \sigma_{\theta}^2 \)
- contracting with (information) externalities
Information Design

- does the broker want to add noise?
- noise lowers the value of the information to the firm
- common noise may reduce the compensation that individual consumers require
- in practice, “withholding informative variables” = “adding noise”
Optimal Intermediation

- data broker can add idiosyncratic or common noise to the signals received from the consumers

Proposition

1. The optimal information structure delivers positive profits if and only if
   \[ n\sigma_{\theta}^2 > \sigma_{\theta_i}^2 \]

2. The data broker never adds idiosyncratic noise.

3. The optimal common noise \( \sigma_{\epsilon}^2 \) is positive for \( \sigma_{\theta_i}^2 \) large enough and \( n \) small enough.

- noise reduces compensation necessary to consumer
- compensation is decreasing in \( n \) and increasing in \( \sigma_{\theta_i}^2 \)
Large Markets

- let \( n \) become large
- individual information becomes less valuable

**Proposition**

1. As \( n \rightarrow \infty \), \( f_i \rightarrow 0 \), \( nf_i \rightarrow k < \infty \).
2. As \( n \rightarrow \infty \), \( g \) grows linearly in \( n \).

- marginal compensation of consumer goes to zero
- provides explanation of frequently noticed absence of consumer compensation for individual data
Data Acquisition

- Larger data sample from any given consumer increases precision of information.
- Re-interpretation of optimal intermediation:
  1. Additional consumer is more valuable than additional data point of same consumer.
  2. Larger optimal data acquisition with larger consumer/data base.
  3. Cost of compensation decreases with the size of the consumer base.
any particular source of information has positive amount of idiosyncratic noise

\[ s_i = t_i + \varepsilon_i \]

data broker can reduce the noise about consumer \( i \) by increasing the number of sources/channels it opens

let \( x \) be the number of available sources

reducing idiosyncratic noise has a direct effect—it increases the value of information and potentially decreases the payment owed to the consumer—

but also an indirect effect
More Sources, More Services, More Data

- broker can “afford” to sell even more precise signals to the firm.

Proposition

1. The constrained optimal amount of common noise $\sigma^*_\epsilon(x)$ is weakly decreasing in $x$.
2. The broker’s gross profit (gross of investment cost) is weakly convex in $x$. 
Data Acquisition without Money and Indirect Sale

- consumers often transmit the information to an intermediary at zero price ...
- ...and in conjunction with access to other benefits
- Facebook does not compensate users for social information but gives free access to an electronic network
- Amazon, Google provide organic search results in addition to sponsored listings
- information is sold indirectly and as a bundle: sponsored search, display advertising, etc.
Data Acquisition without Money and Indirect Sale

- information sold jointly with access to consumer
The Price of Information

- Suppose absent informational concerns by the consumer, there are linear benefit to consumer to generate data $b \cdot \tau_i$

- Data intermediary (platform) designs platform to reveal information incentive compatible

- Can choose constant marginal benefit to implement revenue maximizing intermediation without monetary transfer to consumer

- Consumers may be willing to transfer data for very little compensation/service
Concluding Thoughts

- cost of acquiring information vanish while gains persist as markets grow large
- additional sources/services increase revenue more than linearly
- implications for industry structure of data intermediaries
- what is the outcome of competition among brokers/platforms?
Supply Chains and Intermediation

The transfer of information from the consumer to the intermediary does not happen in one stop, but is itself intermediated. For example, in the world of consumer financial data, it is often the banks and financial institutions who collects the individual data, such as the credit history of a personal account. These firms then forward the data to a credit bureau, and then buy additional data about their own consumers and possibly new prospects. Who captures rents in the supply chain of information goods (Manea, 2018)?
Prices and Quantities

When type $t_i$ is realized, consumer $i$ buys

$$q_i(p) = t_i - p.$$  

The seller solves the following problem

$$\pi = \max_{p} \mathbb{E} \left[ (\sum_{i=1}^{n} t_i - np) p \right].$$

Monopoly price when observing $k \leq n$ signals:

$$p(s) = \frac{1}{2} \frac{\mathbb{E} \left[ \sum_{i=1}^{n} t_i \mid s \right]}{n} = \frac{1}{2} \left[ \alpha(n, k) \sum_{i=1}^{k} s_i \right. + \left. (1 - \alpha(n, k)) \mu \right],$$

and hence under any information structure, the ex ante expected price

$$\mathbb{E}[p] = \mu/2.$$
Fix an information structure. Ex ante, consumer surplus is
\[ CS = \frac{1}{2} \mathbb{E} [(t_i - p)^2] = \frac{1}{2} \mathbb{E} [t_i^2] + \frac{1}{2} \mathbb{E} [p]^2 - Cov [t_i, p] + \frac{1}{2} Var [p]. \]

Broker must compensate consumers for differential surplus.

As a consumer, withholding signal \( \Rightarrow \) price covaries less with your type.

Seller’s profits are given by
\[ \pi = n \cdot \mathbb{E} [p]^2 + n \cdot Var [p]. \]

Broker can charge seller’s entire value of information \( n \cdot Var [p]. \)
Broker’s Problem

Suppose there is only one consumer \((n = 1)\).

Consumer compares surplus levels:

\[
CS_0 = \text{constant} \\
CS_1 = \text{constant} - \frac{3}{2} Var[p]
\]

Value of information

\[
\Delta \pi = Var[p] < \Delta CS
\]

Broker cannot make money!
“Selling wine without bottles” (resell, reuse).
Exclusive sales are difficult to monitor and enforce.
Data quality is hard to verify without access (Arrow’s information paradox).
The cost of revealing information depends on its downstream use.

Indirect sales of information (e.g., bundled with advertising) solve:
the resale & reuse problems by never giving out the data;
the exclusivity problem by committing to a scarce number of slots.
Repeated interaction (& credit card data) can alleviate the quality problem.
“Selling wine without bottles” (resell, reuse).
Exclusive sales are difficult to monitor and enforce.
Data quality is hard to verify without access (Arrow’s information paradox).
The cost of revealing information depends on its downstream use.

Indirect sales of information (e.g., bundled with advertising) solve:
the resale & reuse problems by never giving out the data;
the exclusivity problem by committing to a scarce number of slots.
Repeated interaction (& credit card data) can alleviate the quality problem.
But the data must be first acquired, potentially at a cost....
Sourcing and Intermediation: More Examples

Suppose \( m \) firms want to buy information about \( n \) consumers. Firms play a linear-quadratic game of strategic substitutes or complements.

Information design problem with bilateral contracting: broker adds common and idiosyncratic noise to available consumer signals;
each firm receives a message with common and idiosyncratic noise.

Profit-maximizing information structure: correlation, exclusivity?

Credit bureaus’ information:
Equifax “Work number” sources information from centralized payroll services
Sells employment and income verification to other employers or creditors:
buyer submits a list of customer accounts (or job candidates);
Equifax appends some variables of interest.
Selling to Competing Firms

Relationship to literature on information sharing in oligopoly (Vives, 1984; Jappelli and Pagano, 1993; Raith 1996). Firms need to buy signals. Sharing is endogenously determined by the broker. Exclusive sales through asymmetric noise terms. More general information structures, allows correlation in signals (Bergemann and Morris, 2013).