Data Sellers

Ernest Liu
Princeton

Song Ma
Yale

Laura Veldkamp
Columbia

WHAT INDICATES MARKET POWER IN DATA MARKETS?

▶ Market power in data markets has been a cause for concern. What are indicators of market power?

  ▶ Marginal cost is typically close to zero. Markups near infinite.
  ▶ Profits could reflect efficiency or high value of data.
    Not necessarily bad for conusmers.

▶ Key model features:

  ▶ Data is a durable good, with depreciation.
    Data is still relevant tomorrow, e.g., to train our AI algorithms.

  ▶ Data also exhibits strategic substitutability.
    It's less valuable when others know it.

▶ Test model predictions in online data marketplaces.

FINDINGS

▶ Firms that sell data and cannot commit to sell only a few copies will earn little.

  ▶ This is bad news for consumers: Such firms also invest little in data quality.

▶ Data can be sold as a subscription: A tool for firms to restore their commitment power.

  ▶ Firms profit more and consumers get higher quality data.

▶ Firms with financial constraints may prefer data sales because the cash flows are front-loaded.
Small markets or rapid depreciation make commitment less problematic.

▶ Empirical evidence: Data sellers choose sales when they have little financing, when data depreciates quickly and when the market is small. Otherwise subscriptions.

▶ We use the model to show when data subscriptions indicate market power problems (low consumer surplus).

OUTLINE

A MODEL OF DATA SALES

▶ A continuum of consumers, of varieties and of goods-producing firms.

▶ 2 goods producers are randomly selected to Bertrand compete for each variety.

▶ They earn a profit $\hat{\pi}(x)$, *if they have more data than their competitor.* (makes data more valuable when others do not have it.)

▶ A single data seller chooses data quality x at $t = 0$ and chooses how many copies to sell n_t, at each date t.

▶ Maximize expected profits, with discount rate β.

MODEL MAIN RESULTS

▶ If selling data means selling the right to use the data forever,
  ▶ Data sellers compete with the future selves.
    Knowing that a seller will sell more copies of the data tomorrow, today's
    buyer expects its future value to be low. Erodes firms' revenue.

▶ If firms instead sell subscriptions, with costly updates,
  ▶ Firms sell less, because a drop in data value means a decline in current
    revenue. Less scope to steal value from previous buyers.

  ▶ With high data depreciation, firms can achieve full-commitment revenue.

FOUR HYPOTHESES TO TEST

1. Financially constrained data sellers sell, instead of subscriptions.
   Financial constraints $\rightarrow$ higher effective discount rates for firms.
   Data sales generate more cash quickly.

2. Rapidly depreciating data is sold.
   Depreciation makes data sold in the past less relevant.
   That alleviates commitment problems.

3. Sell data if market is small.
   If there are fewer buyers, the commitment problems are less severe.

4. Data sellers' revenue grows slowly.
   Adding too many customers quickly undermines data value
   (substitutability).
   If customers are added quickly, data prices will be low.

OUTLINE

Word cloud of data description text on Datarade.

EXAMPLE: A DATARADE SELLER

Top data categories on the Datarade platform.

Category	Product Count	Percentages
B2B Contact Data	530	14.39%
B2B Leads Data	527	14.31%
B2B Marketing Data	522	14.17%
Company Data	513	13.93%
B2B Email Data	457	12.41%
Firmographic Data	303	8.23%
B2B Decision Maker Data	302	8.20%
Point Of Interest POI Data	268	7.28%
Business Website Data	241	6.54%

TESTING H1: FINANCIALLY CONSTRAINED SELL DATA

Dependent Variable: Data Transaction Type

	One-Off-Purchase	Monthly Licensing	Yearly Licensing
ln(# Rounds)	-0.075***	0.092***	0.109***
	(0.016)	(0.014)	(0.015)
R-squared	0.006	0.012	0.014
ln(Funding $)	-0.003*	0.012***	0.012***
	(0.001)	(0.001)	(0.001)
R-squared	0.001	0.023	0.019
Observations:	3,683		

2 rounds of financing, instead of 1, reduces the probability of data purchase 5.2%, from the base rate of 46.4% (an 11.2% decline).

Testing H2: Rapidly depreciating data is sold

	(1) One-Off-Purchase	(2) Monthly Licensing	(3) Yearly Licensing
Depreciation	0.027*** (0.003)	-0.012*** (0.003)	-0.011*** (0.003)
Observations	3,667	3,667	3,667
R-squared	0.025	0.006	0.004

Moving up one decile of depreciation rates makes data sale +2.7pp more likely (5.8% increase).

	(1) One-Off-Purchase	(2) Monthly Licensing	(3) Yearly Licensing
Market Size	-0.015 (0.018)	0.144*** (0.016)	0.182*** (0.017)
Observations	3,683	3,683	3,683
R-squared	0.000	0.022	0.030

A 10% larger market size is associated with 1.4pp more monthly licensing (a 5.1% increase) .

TESTING H4: SLOW DATA REVENUE GROWTH

	(1) ΔVenture Value	(2) ΔGoogle Trends Index
Data Provider	-0.526**	-0.066*
	(0.216)	(0.035)
Observation-level	Company-Round	Search Term-YearMonth
Observations	54,124	19,741
R-squared	0.024	0.012
Fixed Effects	Year State	Year-Month

Data providers have a 53% lower venture value growth and 6.6% lower google growth for data firms, relative to non-data comparables.

Source: Crunchbase and PitchBook. ΔVenture Value is % change of the current financing round from the previous round.

OUTLINE

CALIBRATING PARAMETERS
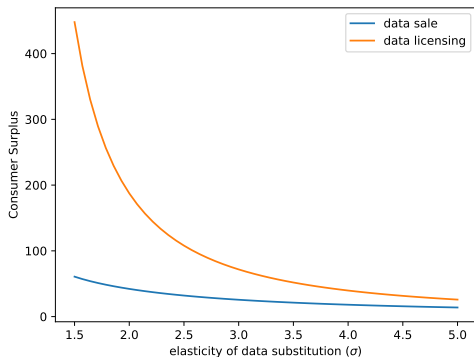
▶ How do monopoly power and the incentive to invest in quality data trade-off in consumer surplus?

▶ Which type of data transaction is better for consumers?

▶ 2 structural parameters: time preference β and goods substitutability σ. Calibrate $\beta = 0.96$ to match riskless rate.

▶ Explore various levels of σ.

CONSUMER SURPLUS: SALES VS. SUBSCRIPTIONS

▶ Data sale is better only for very elastic goods (nearly perfect substitutes).

▶ Market power of licensing creates incentives for data production that ultimately benefit consumers.

▶ Data monopolies do not necessarily imply strong market power.
The inability of firms to commit to restrict future data sales undermines
their own profit.

▶ Real world data markets show signs of this commitment problem, especially
when firms face financial constraints in large markets with durable data.

▶ Monopoly profits create incentives to invest in quality data.
Consumer welfare might be better off with data markets that sustain
market power.

DATARADE SUMMARY STATISTICS

Pricing Model	count	mean	std	25%	50%	75%
One-Off-Purchase (0/1)	3683	0.464	0.499	0	0	1
Monthly Licensing (0/1)	3683	0.277	0.448	0	0	1
Yearly Licensing (0/1)	3683	0.384	0.487	0	0	1
Usage-Based (0/1)	3683	0.370	0.483	0	0	1