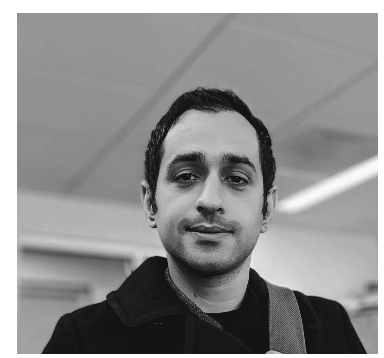
Equilibrium Characterization for Data Acquisition Games

Zachary Schutzman





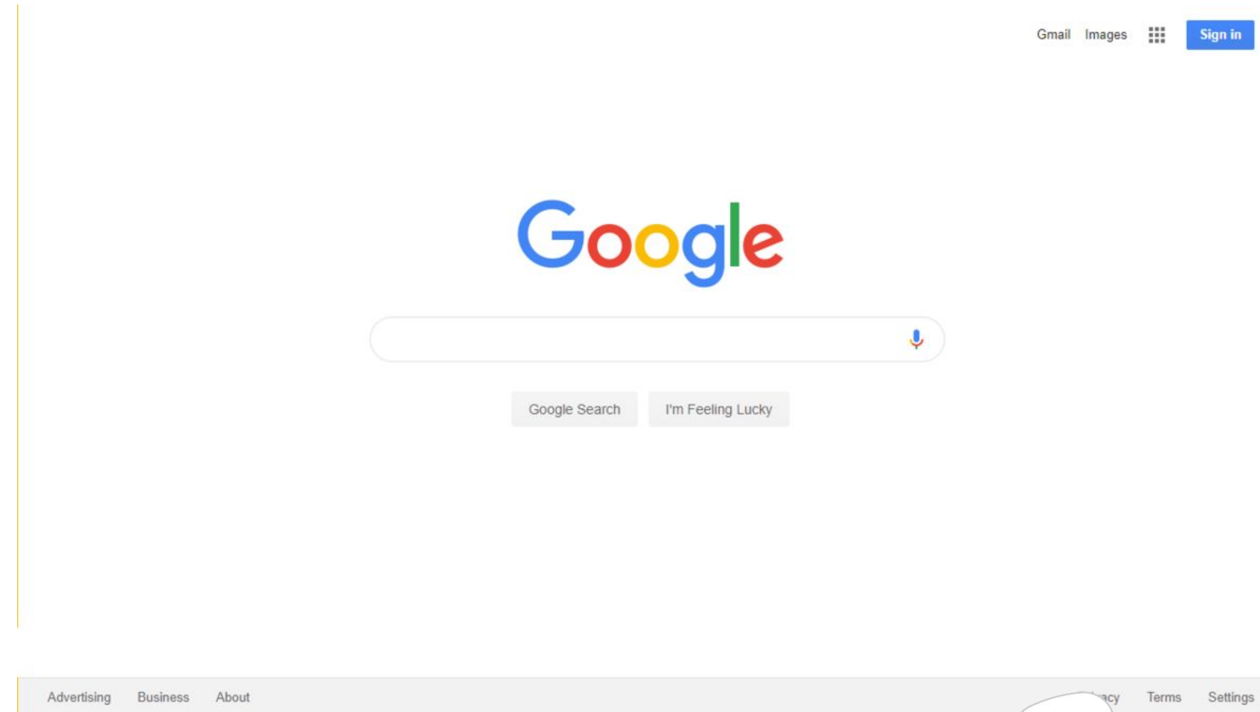

with Jinshuo Dong, Hadi Elzayn, Shahin Jabbari, Michael Kearns

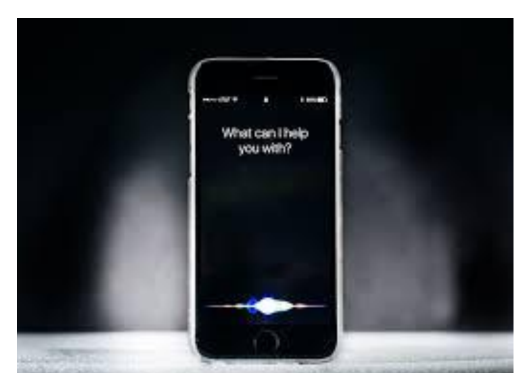
Penn
UNIVERSITY of PENNSYLVANIA

IJCAI 2019

Motivation

- Modern services are built on data and ML
- Classical economic models need to be adapted





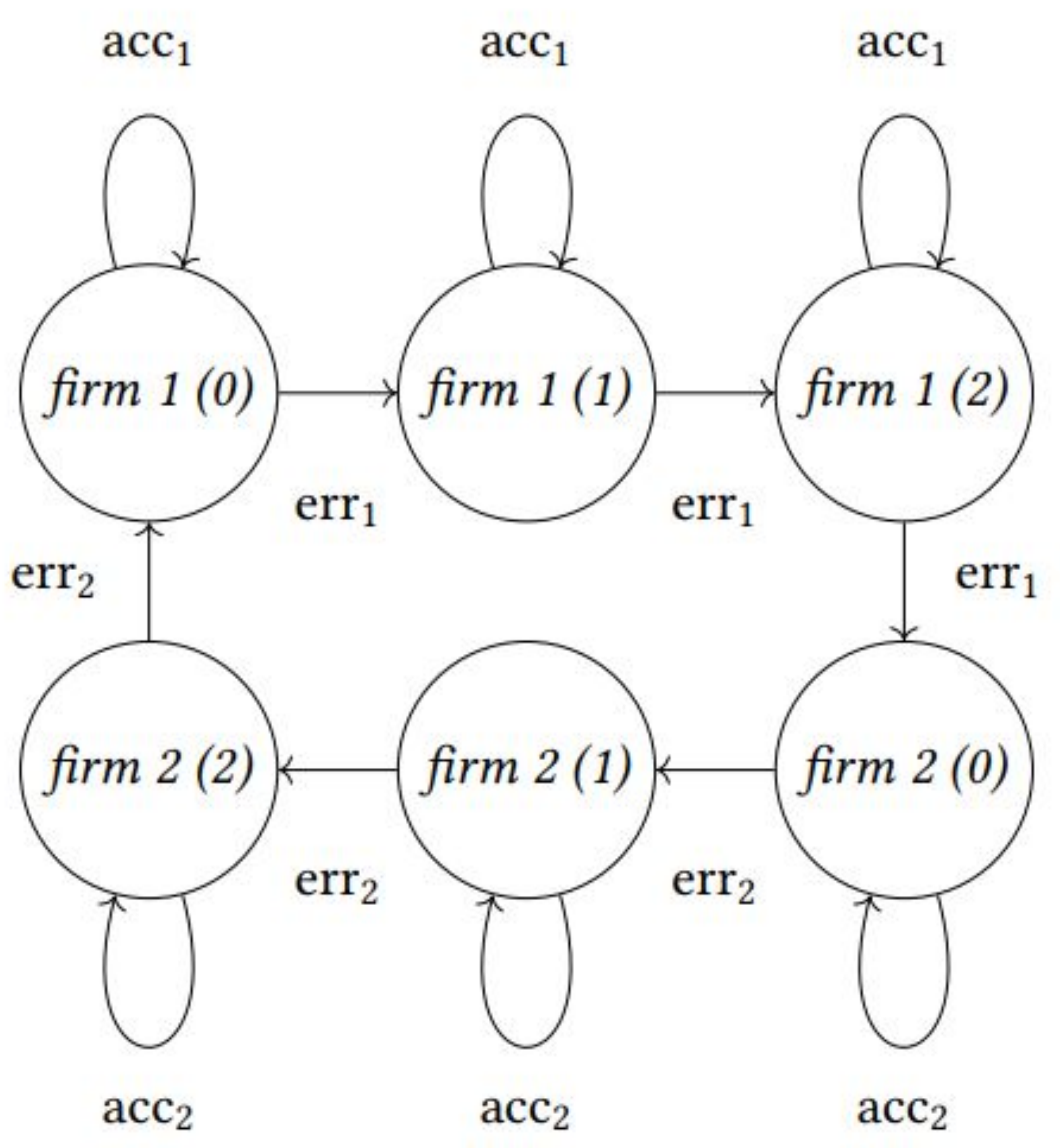
Setting

- Two firms provide a similar service. Throughout, we assume that Firm 1 has more data than Firm 2

- Each firm already has some data and captures a certain share of the market

- There is a new corpus of n data points available at a price p

Data and Market Share



- A user makes queries of a service until *a* mistakes are made, then switches
- The relative errors of the firms' models and this "competition" parameter *a* determine the relative market shares

Model Selection

Problem: Firms need to jointly choose a **learning model** and a buy/don't buy **action** in the game.

How do we reason about this (extremely large) strategy space?

Reduction from Learning Theory

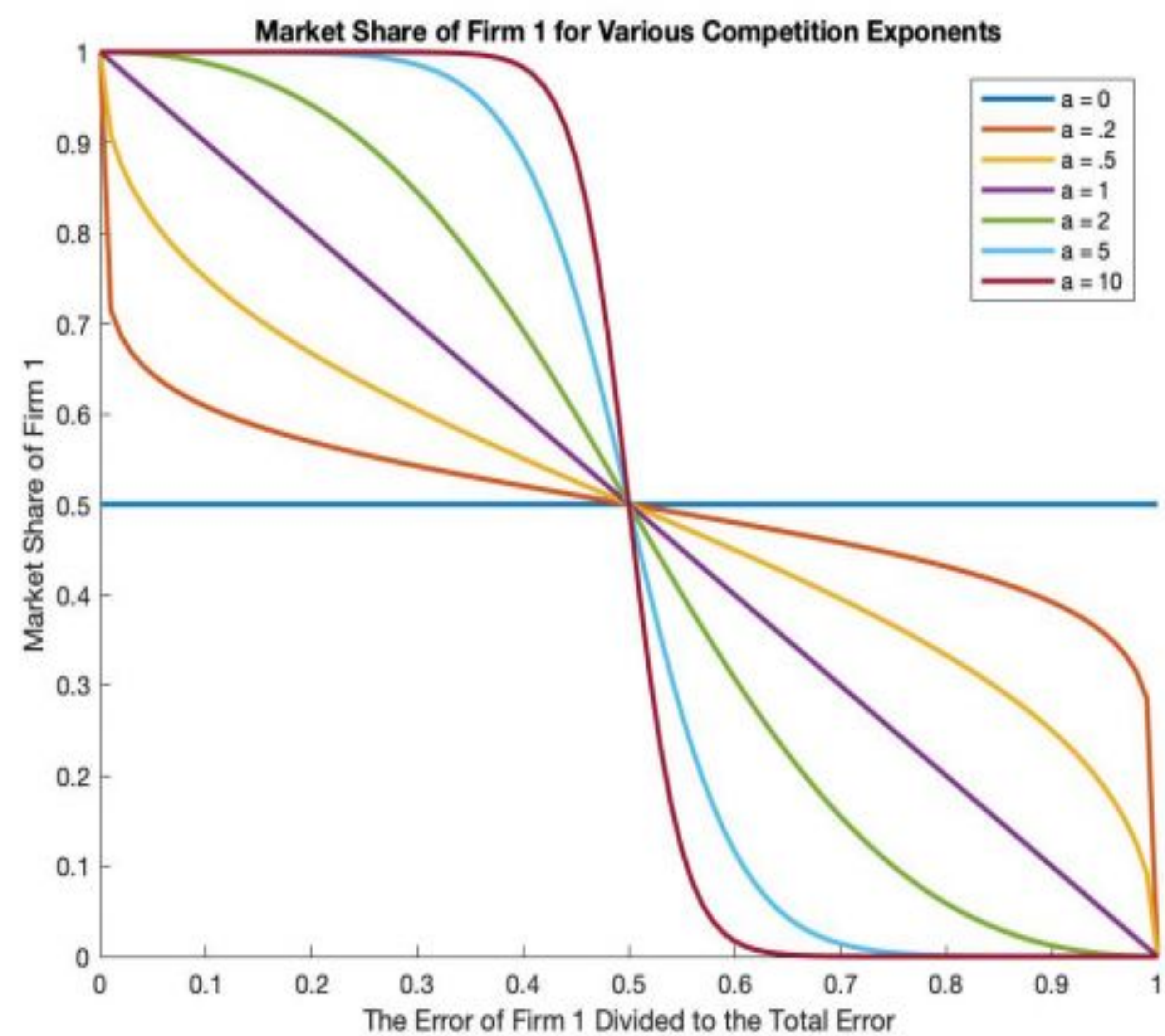
For the class of neural nets with *d* nodes, given m training samples, the generalization error is at most $c_1/m + c_2/d$ [Barron, 1994]

- For an amount of data *m*, there is an optimal choice of d to minimize error!

  Here d is $\Theta(1/\sqrt{m})$, generally $\Theta(m^{-r})$ for some *r* called the **learning rate**

Market Shares

- We can write the relative market share of Firm 1 as $\mu_1 = m_1^b/(m_1^b + m_2^b)$
- b = a*r where a is the competition exponent and -r is the learning rate



The Simplified Game

- Firms choose to buy the new data or not based only on the price and how market shares will change
- The firms face the following payoff matrix:

Firm 1/Firm 2	Buy (B)	Not Buy (NB)
Buy (B)	$\frac{1}{2}(\mu_1(m_1+n, m_2, b)$ $+ \mu_1(m_1, m_2+n, b)$ $-p)$	$\mu_1(m_1+n, m_2, b) -$ p
Not Buy (NB)	$\mu_1(m_1, m_2+n, b)$	$\mu_1(m_1, m_2, b)$

Equilibrium Characterization

There are **three regimes** to consider in analyzing the equilibria of this game:

- If the price is **too high**, both firms always decline to buy the data

- If the price is **too low**, both firms always try to buy the data

- In the **intermediate range**, there are three equilibria

Price Thresholds

$$A = \frac{(m_1+n)^b}{(m_1+n)^b + {m_2}^b} - \frac{{m_1}^b}{{m_1}^b + (m_2+n)^b},$$

$$C = \frac{(m_1+n)^b}{(m_1+n)^b + {m_2}^b} - \frac{{m_1}^b}{{m_1}^b + {m_2}^b},$$

$$D = \frac{(m_2+n)^b}{{m_1}^b + (m_2+n)^b} - \frac{{m_2}^b}{{m_1}^b + {m_2}^b}.$$

- A/2 is the expected change in μ_1 when moving from (NB,B) to (B,B)
- C is the change in μ_1 when moving from (NB,NB) to (B,NB)
- D is the same for μ_2

Price Thresholds

- The **lower threshold** is max(C,D)
- The **upper threshold** is A

- A/2 is the expected change in μ_1 when moving from (NB,B) to (B,B)
- C is the change in μ_1 when moving from (NB,NB) to (B,B)
- D is the same for μ_2

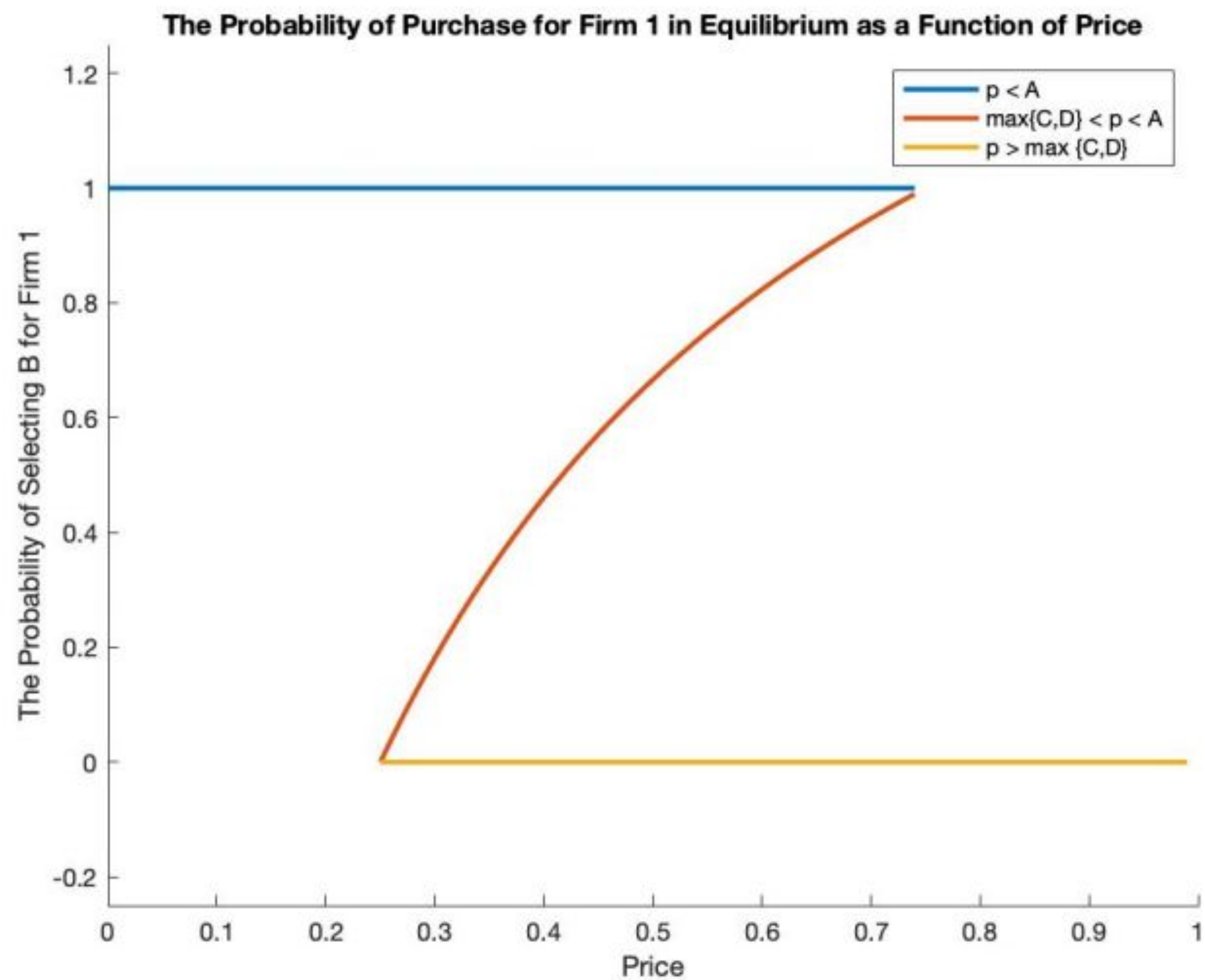
Intermediate Prices

When p is in the **middle range** there are **three equilibria**:

1. Both firms **buy the data**
2. Both firms **decline to buy the data**
3. A unique **mixed strategy Nash equilibrium**

Three Equilibria

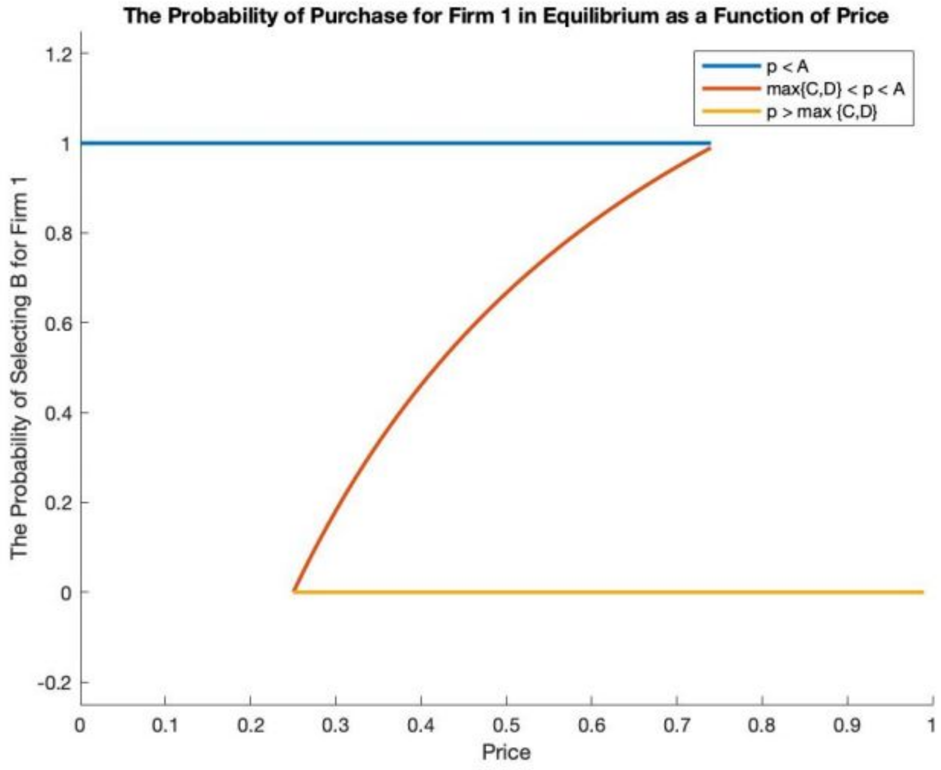
- In the mixed equilibrium, Firm 2 puts a **higher weight** on buying than Firm 1 does
- For both firms, the probability of buying is **increasing** in the price p



A Data “Arms Race”

- Both firms prefer **neither buys the data**

- Both firms prefer **having the data** rather than the other firm having it



Impact on Market Shares

- For any choice of parameters, Firm 2 is **more likely** to get the new data than Firm 1
- The market tends **away from monopoly**

Impact on Consumers

- Users prefer Firm 1 to improve its already superior product
  - (B,NB) ≥ (B,B) ≥ (NB,B) ≥ (NB,NB)
- Note (B,NB) is **never** a pure strategy equilibrium outcome and is an **unlikely** mixed strategy outcome
- Preferences of users and equilibrium outcomes **do not align**

Thank you!

ianzach@seas.upenn.edu