Lecture 5

Markets, Mechanisms and Machines

David Evans and Denis Nekipelov

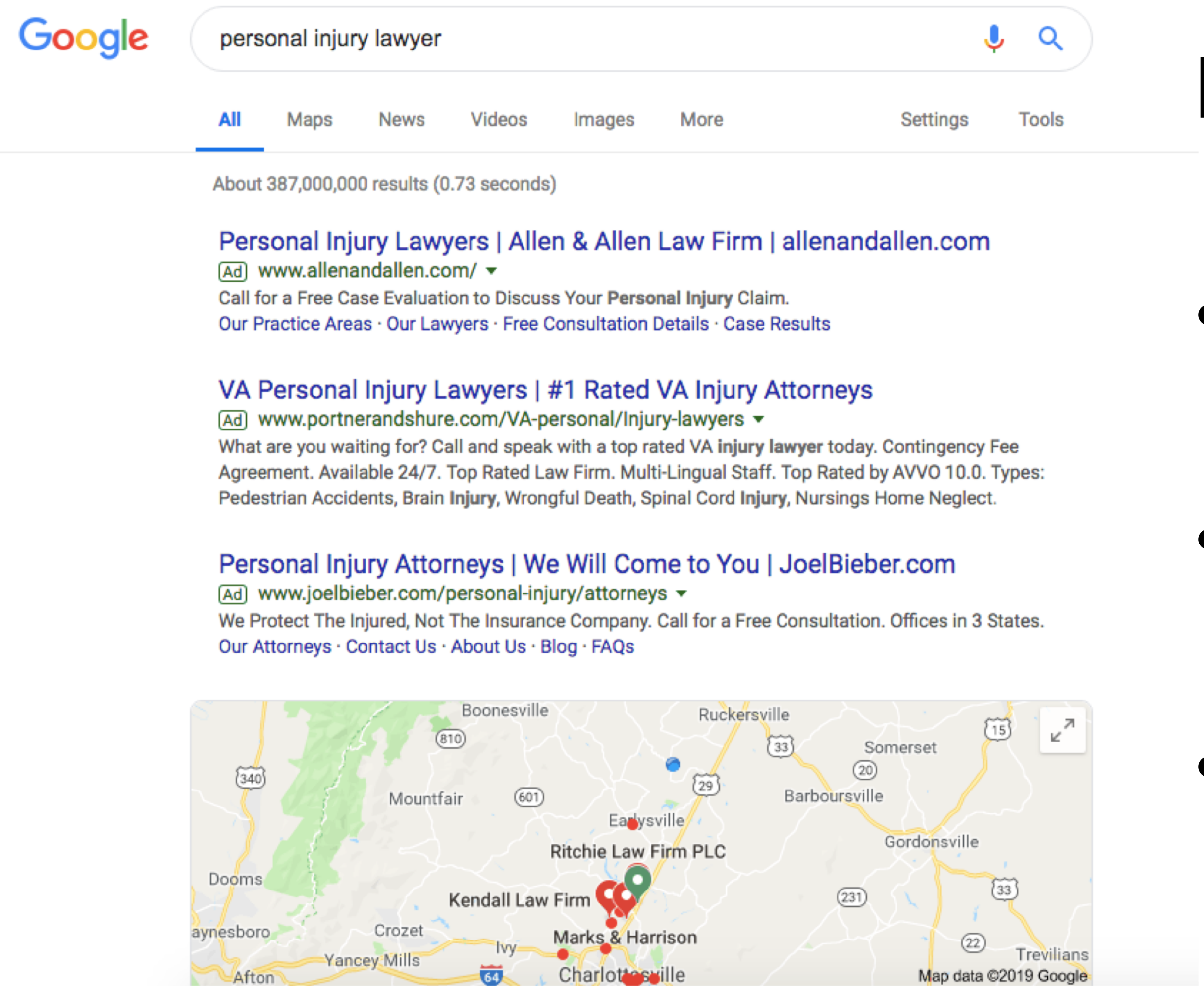
Empirical risk minimization in practice

- Empirical risk minimization as a component of a large system
  - Allocation decisions
  - Pricing decisions

- Outcomes of statistical learning
  - Inform decisions
  - Indicate if system needs to be changed/updated

Online advertising

Ad displays based on causal model of relevance



Effectiveness evaluated by:

- The fact of ad display
- User engagement (“likes,” comments)
- User clicks

Treatment effects

- General model: policy (ad display) leading to specific outcome (user click)
- Statistical model: study how application of policy (treatment) changes the distribution of outcome
  - Treatment status: $D = 0$ or 1
  - Treatment outcome: random variable $Y(1)$ (if $D=1$) or random variable $Y(0)$ (if $D=0$)
- This is a very popular model for statistical analysis of policies

Treatment effects

- Economists and social scientists are often interested in seeing how a population responds to a treatment or program.
- We already talked about effect of ad displays, other examples include the effectiveness of job training programs, the impact of a minimum wage, and the effect of school vouchers.
- Program evaluation concerns estimating the effect of such programs, policies, and interventions (i.e. treatments).

Treatment effects

- Ideally, we'd like to compare the same individual with and without treatment (i.e. compute Y(1)-Y(0) for each individual), but this is clearly impossible.
- An alternative is to run a randomized experiment. How?
- Construct a random sample of *n* individuals and assign treatment $D_i = D$ to a randomly selected subset of them (the treatment group) and keep the remaining subset as a control (the control group).
  - Select *n* users and show an ad to a random subset of them
  - Select *n* unemployed people, give some job training to a random subsample and compare employment outcomes.
  - select *n* kids, give school vouchers to a random subsample and compare education outcomes.
  - select *n* sick people, give a new drug to a random subsample and compare health outcomes.

Endogeneity

- When a regressor is correlated with regression error, we call it *endogenous*
- Endogeneity most commonly arises when variables that had to be included in the regression were omitted
- E.g. when the regression model is

$$Y=a_0+a_1X_1+a_2X_2+u$$

but we fail to include X_2 and instead only include X_1, then coefficients estimated from model

$$Y=b_0+b_1X_1+v$$

will not produce accurate estimator for a_0 and a_1

Endogeneity

- The reason is that if $Y=a_0+a_1X_1+a_2X_2+u$ is the true causal model, then in the short model $Y=b_0+b_1X_1+v$ is a part of the residual v
- Therefore, if X_1 and X_2 are correlated, then X_1 is correlated with v in the short model and b_1 is not an accurate representation of a_1
- We can verify that the OLS estimator for a_1 in the short model is biased

Endogeneity

- In this case

$$E\left[\hat{b}_1\right] = a_1 + E\left[\frac{(X_{1i} - \overline{X_1})(X_{2i} - \overline{X_2})}{(X_{1i} - \overline{X_1})^2}\right] \neq a_1$$

- The term after a_1 is called the omitted variable bias
- Regression of Y on X_1 does not characterize causal effect of Y on X_1
- **<u>Note</u>**: *If X_1 is fully independent from all other variables then b1 does estimate a causal effect!*

Treatment effects

- Causal parameter of interest: *average treatment effect* (ATE)

  a = E[Y(1)-Y(0)] = E[Y | D=1]-E[Y | D=0]

- If the treatment is randomly assigned, D_i will be uncorrelated with the omitted variables by construction. Why?
- As a result, we can estimate ATE via regression

  $Y = a_0 + aD + v$

Treatment effects

- The estimator produced from a linear regression of Y on D is called the differences estimator since it is the difference between the sample average outcome of the treatment group and the sample average outcome of the control group.
- Since an ideal randomized experiment eliminates any correlation between D and all other variables that can cause changes in Y, the differences estimator is unbiased
- Of course, real world experiments are not always ideal, can lead to endogeneity of D

Internal and external validity

- We say that model is internally valid if its estimates converge in probability to true causal parameters in the given population.
- We say that the model is externally valid if its causal predictions extend to other populations
- Threats to internal validity include
  - Failure to randomize (so randomize next time!)
  - Failure to follow treatment protocol (e.g. non-compliance)
  - Attrition (e.g. most able drop out of job training because they get a job)

Internal and external validity

- We say that model is internally valid if its estimates converge in probability to true causal parameters in the given population.
- We say that the model is externally valid if its causal predictions extend to other populations
- Threats to external validity include
  - Nonrepresentative sample
  - Nonrepresentative program/policy
  - General Equilibrium effects
  - Treatment vs. eligibility effects

Causal inference with randomized experiments

- In an ideal randomized experiment, the causal effect can be estimated with the differences estimator
- If treatment is random, the differences estimator is unbiased, but may have high variance (i.e. not statistically efficient).
- Also, if some of the problems on the previous 2 slides are present, the differences estimator may be biased.
- Due to these potential problems, it may be necessary to add additional regressors to the differences estimator:

$$Y = a_0 + aD + b_1W_1 + \ldots + b_kW_k + v$$

where $W_1,\ldots,W_k$ are variables measuring k individual characteristics of each person in the sample.

Causal inference with randomized experiments

- The OLS estimator for *a* in the model

$Y = a_0 + aD + b_1W_1 + \ldots + b_kW_k + v$

where $W_1,\ldots,W_k$ are variables measuring k individual characteristics of each person in the sample is called the "differences estimator with additional regressors".

- Here, D is the treatment variable and the W's are called "control variables".
- So what's a control variable?
- A control variable is a variable that's included to control for a possibility that, if omitted from the regression, would lead to biased estimator for the coefficient of interest
- That is, controls are not of interest per se, but are in the regression to control for omitted variables that could be correlated with the variable we're actually interested in.

Imperfect randomized experiments

- Inclusion of control variables allows us to relax the requirement that treatment assignment D must be fully random
- If the treatment assignment is independent from the outcome conditional on the control variables, then model

$Y = a_0 + aD + b_1W_1 + \ldots + b_kW_k + v$

still produces a correct estimator for the ATE

- However, the assignment of treatment may be different depending on the values of control variables W
  - We may treat individuals with different characteristics differently
  - There could be restrictions and budget constraints that do not allow treat the entire population

Why include additional regressors

$Y = a_0 + aD + b_1W_1 + \ldots + b_kW_k + v$

1. To improve efficiency
   - Including the additional determinants of Y reduces the variance of the error term.
2. To check for randomization
   - If treatment is correlated with the W's, models with and without W's will give different results, indicating a problem.
3. To adjust for "conditional" randomization
   - If randomization depends on W's , we can control for their influence by including them.

Quasi /Natural experiments

- For cost, ethical, and practical reasons, ideal randomized experiments are rare in the social sciences.
- However, the methods we've seen so far can carry over to nonexperimental settings.
- In a quasi-experiment (aka a natural experiment), variations in laws or circumstances or accidents of nature are treated "as if" they induce random assignment to treatment and control groups.
    - e.g. differences in laws, location, timing, birthdates, etc.

Quasi/Natural experiments

There are two types of quasi-experiments

1. Those in which whether (or not) an individual receives treatment is viewed “as if” it's randomly determined.
2. Those in which the “as if” random variation is only a partial determinant of treatment.

- Examples of the first type include
  - Using randomization from experiments not related to the experiment of interest (e.g. Google constantly experiments with various user-facing features)
  - ”Shock events” (weather phenomena, civil unrest events, etc.)
- Examples of the second type include
  - Men's birthdays partially determining draft lottery numbers
  - Distance to a cardiac cath lab partially determining cath rates

A trashy example

JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT **28**, 241–255 (1995)

House Prices during Siting Decision Stages: The Case of an Incinerator from Rumor through Operation[1]

KATHERINE A. KIEL

Northeastern University, Boston, Massachusetts 02115

AND

KATHERINE T. MCCLAIN

Pennsylvania State University, University Park, Pennsylvania 16802

A trashy example

The dataset compiled for this study consists of 2593 single-family home sales in North Andover, Massachusetts, from January 1974 through May 1992. North Andover is located approximately 20 miles north of Boston near several major highways and has a total area of 27.85 square miles. Only houses within North Andover are considered because tax treatment, city services, and host benefits are unique to the town.

As the town's landfill moved toward capacity in the late 1970s, an incinerator which would turn refuse into electricity was proposed. First mention of the facility in the North Andover *Citizen* was in 1978. The facility is located in the northwest corner of the town.

A trashy example

Pre-rumor	1974–1978
Rumor	1979–1980
Construction	1981–1984
Online	1985–1988
Ongoing Operation	1989–1992.

Although the first mention of the plant occurred in 1978, it was late in the year, and some time must elapse for such information to become widespread. Therefore, data from 1978 was assigned to the pre-rumor stage and 1979 data to the rumor stage. Estimation results support this choice. The choice on where to separate the early operation stage and the ongoing operation stage was more abstract, but 4 years was considered sufficient time for local residents to become aware of the advantages and disadvantages of living near the incinerator.

A trashy example

- The difference regression for the houses sold in 1981 is

  Price = 101.3 – 30.7 D

where D is the variable that is equal to 1 if the house is “close” (within 3 miles) of the incinerator site

- Average selling prices for houses near the incinerator are $30.7k lower than those that aren’t

A trashy example

- This does not imply that the incinerator *caused* housing prices to drop
- The regression for the houses sold in 1978 produces

  Price = 82.5 – 18.8D
- Even before there was talk of building the incinerator, houses near to its future were significantly cheaper than houses that were not
  - Is this surprising?

Difference-in-differences estimator

- The problem here is that treatment (i.e. houses near incinerator) and control (i.e. houses far from incinerator) are different from the very beginning
- Introduction of incinerator changed disparity between the values of two types of houses and this is exactly what we need to measure
- To do this, we need to use time variation: measure the difference between prices of houses near and far from incinerator **before** it was known that it was arriving and difference between prices of houses near and far from incinerator **after**
- The difference between these two gaps would produce the actual impact of the incinerator construction and is called the *difference-in-difference estimator*

Difference-in-differences estimator

- Diff-in-diffs estimator is constructed using sample averages

$$\hat{a}^{dif-in-dif} = (\bar{Y}^{t,2} - \bar{Y}^{t,1}) - (\bar{Y}^{c,2} - \bar{Y}^{c,1})$$

- This can be implemented in the regression model when data is available for 2 periods and G is a binary variable equal to 0 in period 1 and 1 in period 2

$$Y = a_0 + aDG + b_1 D + b_2 G + v$$

A trashy example

- We can use diff-in-diffs estimator to eliminate "pre-treatment" differences in the prices

Average House Prices (in 1000's of 1978 $'s)

	1978	1981	Change
Treatment (Near Incin.)	63.7	70.6	6.9
Control (Far from Incin.)	82.5	101.3	18.8

- $\hat{a}^{dif-in-diffs}$ = 6.9 – 18.8 = -11.9

Difference-in-differences estimator

- Because quasi-experiments typically don't have true randomization, there can be systematic differences between the treatment and control groups
- As before, the Diffs-in-Diffs estimator can easily be extended to include additional control variables

$$Y = a_0 + a DG + b_1 D + b_2 G + b_3 W_1 + \ldots + b_{k+2} W_k + v$$

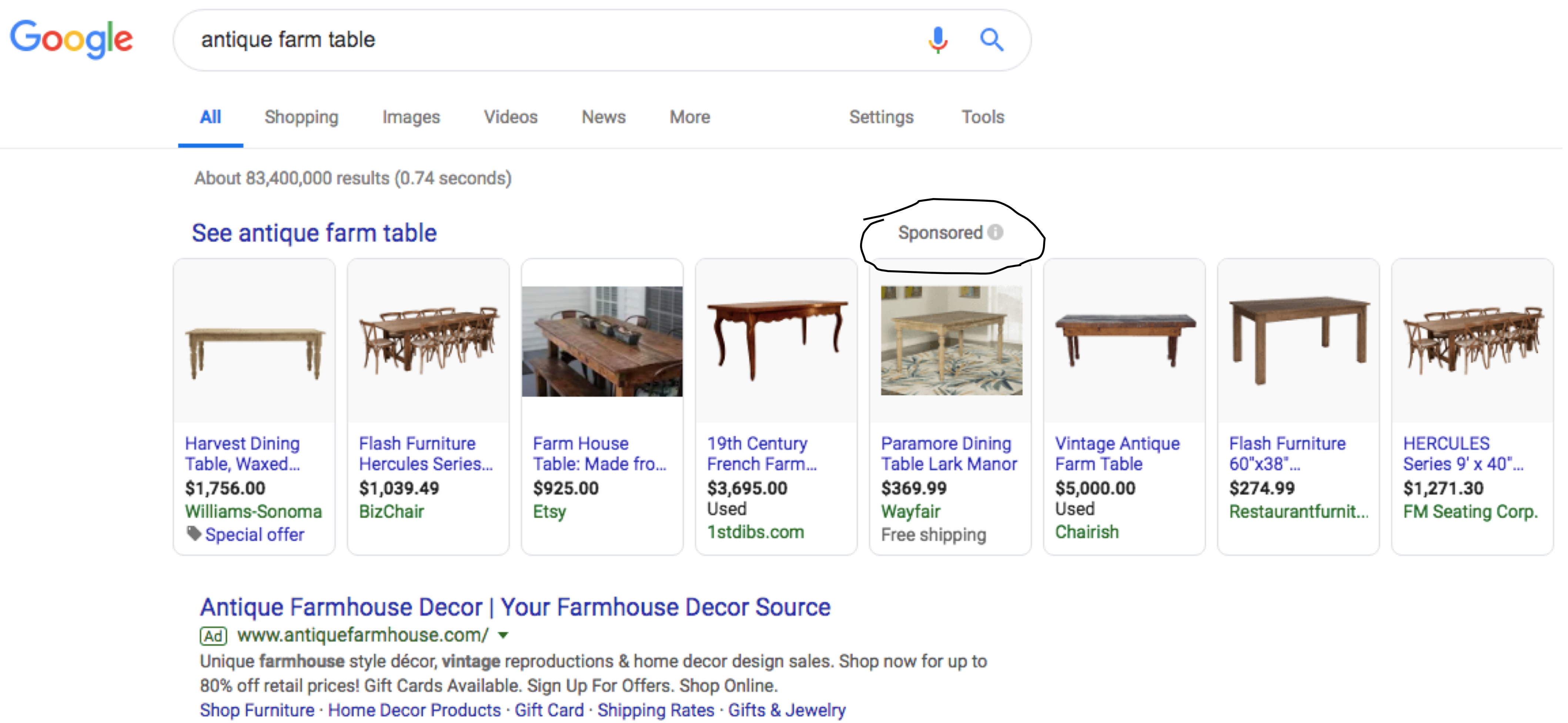
which can also improve eciency.

A trashy example

	Prerumor$_a$ 1974–1977	Rumor$_a$ 1978–1980	Prerumor$_b$ 1974–1978	Rumor$_b$ 1979–1980	Construction 1981–1984	Online 1986–1988	Ongoing 1989–1992
CONST	3.59**	3.47**	3.54**	3.79**	3.29**	2.72**	3.06**
	(5.29)	(7.46)	(6.86)	(5.84)	(9.73)	(8.18)	(9.26)
AGE	−0.79E-02**	−0.76E-02**	−0.79E-02**	−0.71E-02**	−0.58E-02**	−0.29E-02**	−0.65E-02**
	(−5.59)	(−7.41)	(−7.06)	(−5.50)	(−5.35)	(−3.11)	(−6.90)
AGESQ	0.33E-04**	0.29E-04**	0.33E-04**	0.27E-04**	0.27E-04**	0.89E-05**	0.31E-04**
	(4.71)	(5.21)	(5.74)	(4.01)	(3.74)	(1.78)	(6.08)
AREA	0.10E-03**	0.22E-03**	0.14E-03**	0.21E-03**	0.18E-03**	0.18E-03**	0.17E-03**
	(3.05)	(8.33)	(5.11)	(5.56)	(8.95)	(10.27)	(10.63)
BATH	0.14**	0.08**	0.14**	0.08**	0.22**	0.16**	0.15**
	(3.37)	(2.91)	(4.27)	(1.97)	(8.38)	(5.99)	(5.72)
ROOM	0.07**	0.05**	0.07**	0.04**	0.01	0.05**	0.02
	(4.23)	(2.79)	(4.88)	(1.87)	(1.01)	(3.81)	(1.24)
LAND	0.17E-05**	0.44E-06	0.13E-05**	0.32E-06	0.55E-07	0.35E-06	0.15E-06
	(2.67)	(1.25)	(3.23)	(0.43)	(0.19)	(0.92)	(0.74)
LN DIST	0.014	0.042	0.020	0.743E-02	0.071*	0.122**	0.107**
	(0.19)	(0.82)	(0.36)	(0.10)	(1.95)	(3.30)	(3.03)
INTST	0.43E-05	0.16E-04**	0.57E-05	0.28E-04**	0.22E-04**	0.12E-04**	0.17E-04**
	(0.41)	(2.39)	(0.73)	(2.85)	(3.83)	(1.99)	(3.31)
INTSTSQ	0.16E-09	−0.49E-09**	−0.19E-09	−0.74E-09**	−0.73E-09**	−0.47E-09**	−0.64E-09**
	(−0.58)	(−3.07)	(−0.95)	(−3.33)	(−5.03)	(−3.33)	(−5.18)
LAKE	−0.17**	0.07	−0.12**	0.09	0.17E-02	0.10**	0.04
	(−2.11)	(1.53)	(−2.09)	(1.48)	(0.04)	(2.54)	(1.03)

Sales and online advertising

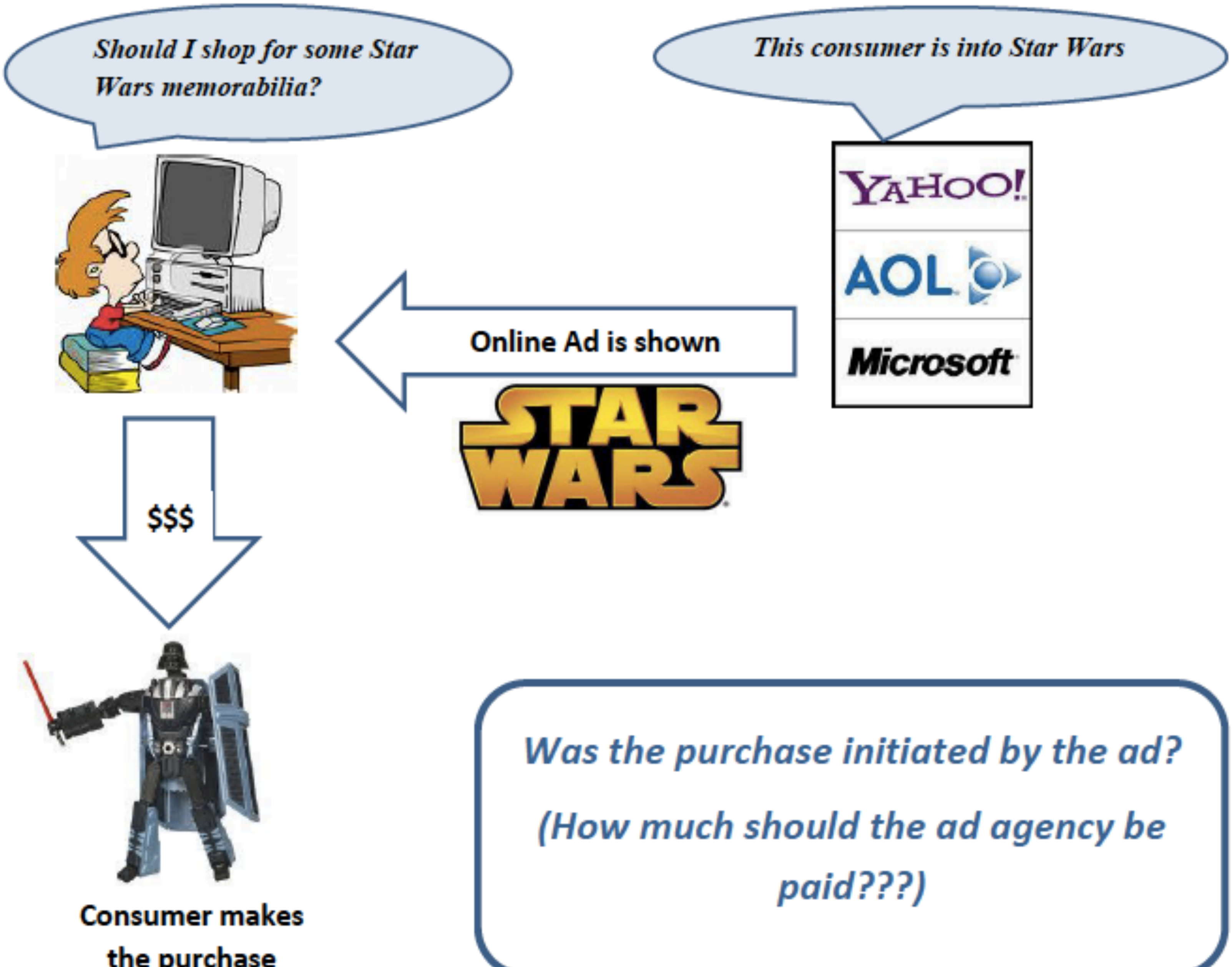
Many online retail platforms rely on search advertising



Sales and online advertising

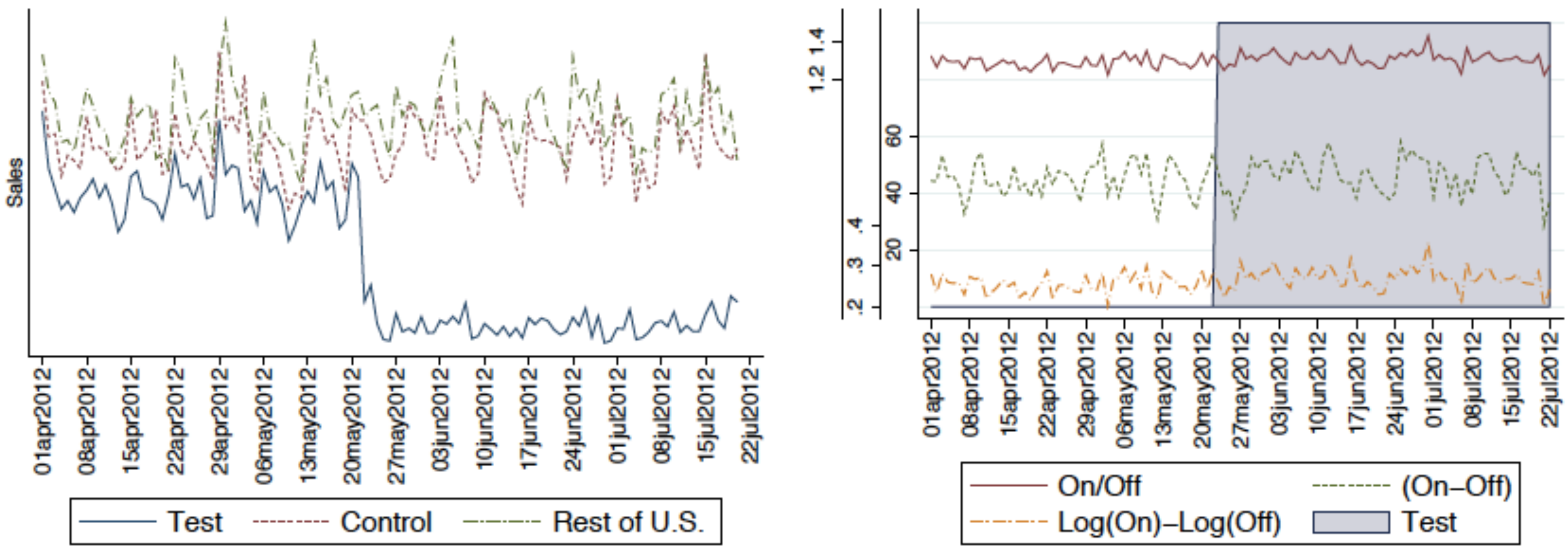
- Ads are sold through auctions (much more coming soon, stay tuned!)
- Large platforms bid on 100’s of millions of keywords
- Looking for “instant gratification:” consumer sees the ad and buys a searched item on the platform
  - Is ad exposure exogenous here?

Sales and online advertising



Sales and online advertising

- In 2012 eBay stops ads in several randomly selected submarkets and selects a “close to identical” set of control submarkets where ads are running



(a) **Attributed Sales by Region**

(b) **Differences in Total Sales**

Reported in Blake, Nosko, Tadelis (2017)

Sales and online advertising

- Diff-in-diffs (region selected for experiment + before and after)

Table 2: Diff-in-Diff Regression Estimates

	Daily		Totaled	
	(1) Log Sales	(2) Log Sales	(3) Log Sales	(4) Log Sales
Interaction	0.00659	0.00659	0.00578	0.00578
	(0.00553)	(0.00555)	(0.00572)	(0.00572)
Experiment Period	-0.0460***		0.150***	0.150***
	(0.00453)		(0.00459)	(0.00459)
Search Group	-0.0141		-0.0119	
	(0.168)		(0.168)	
DMA Fixed Effects		Yes		Yes
N	23730	23730	420	420

Standard errors, clustered on the DMA, in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Reported in Blake, Nosko, Tadelis (2017)