MKT927: INTRO TO QUANTITATIVE MARKETING

Prof. Andrey Fradkin

Lecture 2: Experiments and Advertising

EXPERIMENTS AND POTENTIAL OUTCOMES

THE POTENTIAL OUTCOMES FRAMEWORK (NEYMAN-RUBIN CAUSAL MODEL)

Simplest setup.

- Each unit, i, can either be treated ($D_i = 1$) or not treated ($D_i = 0$) and has an observed outcome $Y_i^{D_i}$.
- The causal/treatment effect of D on Y_i is defined as $\tau_i = Y_i^1 - Y_i^0$.
- The observed outcome is: $Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$.
- The average treatment effect (ATE) is defined as $E[\tau_i]$. Note, this will vary based on the population of interest.

Key identification problem: can't see both Y_i^1 and Y_i^0,

Consumer	Ad (1)	No Ad (0)
1	1	1
2	6	0
3	5	1
4	8	0
5	4	1
6	10	1
7	10	0
8	6	0
9	7	0
10	9	1

Potential Outcomes for Consumers Seeing Ads or Not

Consumer	Ad (1)	No Ad (0)	Treatment Effect
1	2	1	1
2	6	0	6
3	5	1	4
4	8	0	8
5	4	1	3
6	10	1	9
7	10	0	10
8	6	0	6
9	7	0	7
10	9	1	8
Average	**6.7**	**.5**	**ATE = 6.2**

Potential Outcomes for Consumers Seeing Ads or Not

RANDOMIZATION

- “Simple” randomization: flip a coin independently for each person. Problem, can result to numbers of participants being unbalanced.
- “Complete” randomization. If you know how many people are in your group, can ensure exact split.
- “Blocking”. Can try to reduce the variance of the estimate by ensuring consistent proportions of units across treatments.
- “Cluster assignment”. Treat everyone in a group the same way.

DIFFERENCE OF MEANS ESTIMATOR OF ATE

Consumer	**D**	**Ad (1)**	**No Ad (0)**	**Observed**
1	1	**2**	1	2
2	0	6	**0**	0
3	1	**5**	1	5
4	0	8	**0**	0
5	1	**4**	1	4
6	1	**10**	1	10
7	0	10	**0**	0
8	1	**6**	0	6
9	0	7	**0**	0
10	1	**9**	1	9
Average		**6**	**0**	$\widehat{ATE}$ **= 6**

CAUSAL QUANTITIES OF INTEREST

- Average treatment effect, $ATE = E[\tau_i]$
- Average treatment effect on treated, $ATT = E[\tau_i | D_i = 1]$
- Average treatment effect on untreated, $ATU = E[\tau_i | D_i = 0]$
- Conditional ATE, $CATE = E[\tau_i | X_i = x]$
- Intent to treat effect (ITT) vs Complier Average Causal Effect (CACE) in cases with imperfect compliance.

ASSUMPTIONS OF CAUSAL INFERENCE

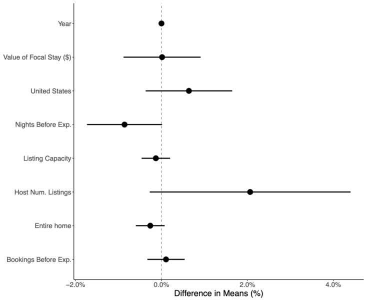
- Stable Unit Treatment Value Assumption (SUTVA) (i.e. no interference, no spillovers).
- Excludability (i.e. nothing else happened at the same time that is not a part of the intended treatment).
- For every experimental paper in this class, I want you to ask if either of these assumptions are violated.

EXPERIMENT DESIGN

- Treatments: Vary one thing at a time. Design to test mechanisms.
- Consider statistical power. Do power simulations based on pilot experiments / priors.
- Consider cluster randomization / blocking where appropriate.
- Pilot experiments are critical.

COVARIATE BALANCE CHECKS

- Why do we do this if we know we randomized?
- Main reason: Don't 100% know randomization was done correctly. This is a diagnostic.
- Also check if the proportions of units in each treatment group correspond to the intended proportions.



RECOMMENDATIONS OF ECKLES

- Report tests of the null hypothesis that treatment was randomized as specified.
- Test should account for clustering / blocking.
- Should certainly not use $p < 0.05$ as a decision criterion here.
- If there is evidence against randomization, authors should investigate.
- They should typically appear in a supplement or appendix — perhaps as Table S1 or Table A1.

UNCERTAINTY QUANTIFICATION

Two sources of uncertainty

- Design-based uncertainty. We are interested in causal effects in the sample. Just by chance we got one randomization vs another.
- Sampling uncertainty. We are interested in causal effects in the population of interest. We are worried that results in our sample don't extrapolate. Just by chance, we may have gotten a sample that is more or less similar to the population.

RANDOMIZATION INFERENCE

- Sharp null hypothesis of interest: All treatment effects equal to 0.
- If this is true, we can impute counterfactuals.
- We can simulate a bunch of randomizations under the null and come up with a distribution of $\widehat{ATE}$ estimates. Other test statistics can also be of interest!
- P-value: The share of simulated test statistics that are greater than the observed test statistic in the true randomization.

WHY RANDOMIZATION INFERENCE?

- Does not require asymptotics.
- We are often not interested in sampling uncertainty. For example, most Prolific/MTurk experiments do not even try to claim that they have representative samples.
- Accommodates a variety of test statistics.
- Even so, rarely done in practice.

REDUCING UNCERTAINTY THROUGH COVARIATES

- The most important thing is to collect the right covariates.
- Pre-treatment outcomes are often very good. For example, on an online platform, the usage of the platform in the month prior to the experiment. (Variations of this are called “CUPED” in industry).
- Make sure that the covariate isn’t affected by the treatment. Or even measured after the treatment happened.

OLD-SCHOOL COVARIATE ADJUSTMENT

Just run a regression.

$$Y_{ij} = \beta X_{ij} + \tau D_i + \epsilon_{ij}$$

Subtle issues arise when treatment effects are heterogeneous or groups have different proportions.

For example see Lin (2013) or Goldsmith-Pinkham et al. (2024) (considers many treatment groups).

HOW I PREFER TO DO REGRESSION ADJUSTMENT (LIN (2013))

Demean covariates, and interact with the treatment.

$$Y_{ij} = \beta X_{ij} + \tau D_i + \tau_x D_i (X_{ij} - \bar{X}) + \epsilon_{ij}$$

DANGERS OF REGRESSION ADJUSTMENT

- Many degrees of freedom, both in covariates and interaction terms.
- Can lead to p-hacking.
- Solution: Pre-register your analysis plan for a specification if you have strong prior that it is the right one and/or state the machine learning method you are going to use.

MACHINE LEARNING AND EXPERIMENTS

- Improve precision.
- Measure heterogeneity.
- Tie one's hands if pre-registered.

No "Gold Standard" for doing this, but many good methods.

POPULAR ML METHODS FOR EXPERIMENTS

We will discuss these when used in specific papers.

- Causal forests (Wager and Athey (2019)).
- Debiased ML + group sorted treatment effects (Chernozhukov and co-authors, many papers).
- MLRATE (Guo et al. (2021)).

WHAT TO ASK ABOUT EVERY EXPERIMENTAL PAPER

- What is the causal question of interest?
- What is the experimental design? Unit of randomization, population of interest, etc.
- Are there any violations of the assumptions of the causal inference framework?
- What is the uncertainty quantification?
- Might it have been p-hacked?
- Does the analysis actually answer the question of interest?

ADVERTISING

WHY ADVERTISE?

Reasons for advertising:

- **Persuasive View:** Advertising changes preferences.
- **Informative View:** Advertising reduces search costs.
- **Complementary View:** Advertising enhances the product's value (e.g., through social signaling).

Bagwell's framing Is advertising good or bad for welfare?

Marketing framing Are firms advertising in a profit-maximizing way?

INFORMATIVE ADVERTISING MODELS

Butters (1977):

- Ads help consumers learn about a firm’s existence and price.
- Market delivers socially optimal advertising.

Grossman-Shapiro (1984):

- Includes product differentiation.
- Can result in excessive or insufficient advertising.

Nelson (1970, 1974b):

- For search goods, ads provide direct information.
- For experience goods, ads signal quality and help match tastes.

ENDOGENOUS SUNK COSTS

Sutton's "Sunk Costs and Market Structure":

- Advertising starts as informative or persuasive.
- Over time, creates entry barriers as market visibility becomes increasingly expensive.
- Related to brand building literature.

ONLINE VS. OFFLINE ADVERTISING

Fundamental Economic Difference:

- Online advertising reduces targeting costs.

Key Issues:

- Ad effectiveness.
- Auctions design.
- Privacy concerns.
- Antitrust issues.

LEWIS & RAO. THE UNFAVORABLE ECONOMICS OF MEASURING THE RETURNS TO ADVERTISING, QJE 2015.

Twenty-five large field experiments with major U.S. retailers and brokerages, each reaching millions of customers and collectively representing $2.8 million in advertising expenditure, reveal that measuring the returns to advertising is exceedingly difficult. The median confidence interval on ROI is over 100% wide, the smallest exceeds 50%. Detailed sales data show that, relative to the per capita cost of the advertising, individual-level sales are incredibly volatile; a coefficient of variation of 10 is common. Hence, informative advertising experiments can easily require more than ten million person-weeks, making experiments costly and potentially infeasible for many firms. Despite these unfavorable economics, randomized control trials represent progress by injecting new, unbiased information into the market. The statistically small impact of profitable advertising amid such noise means that selection bias is a crippling concern for widely-employed observational methods. We discuss how these biases and weak informational feedback from experiments fundamentally impact both advertisers and publishers.

STATISTICAL POWER

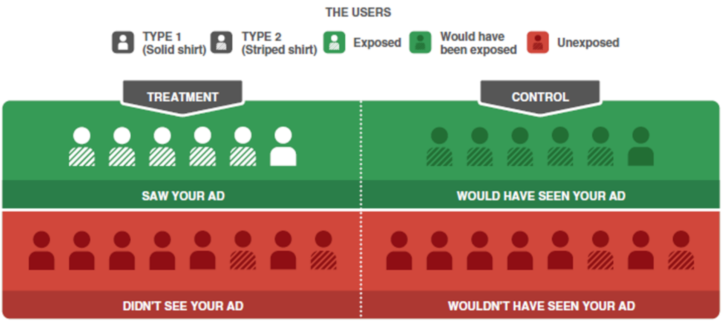
- The power of an experiment is the probability of detecting a true effect if it exists.
- The power of an experiment is determined by the sample size, the effect size, and the standard deviation of the outcome (perhaps residualized).
- Simple power calculation functions are available in all programming languages.
- More complicated experiments can be simulated using simulation-based power analysis.
- Key Challenge: Plugging in plausible values. Running a pilot experiment helps.

IMPLICATIONS OF BETTER MEASUREMENT

- Experiments enable ad effectiveness measurement at scale.
- Repeated experimentation optimizes performance.
- Limitations: Understanding who responds and what is truly being measured.

GHOST ADS

Figure 1
IDEAL EXPERIMENTAL DESIGN



GHOST ADS

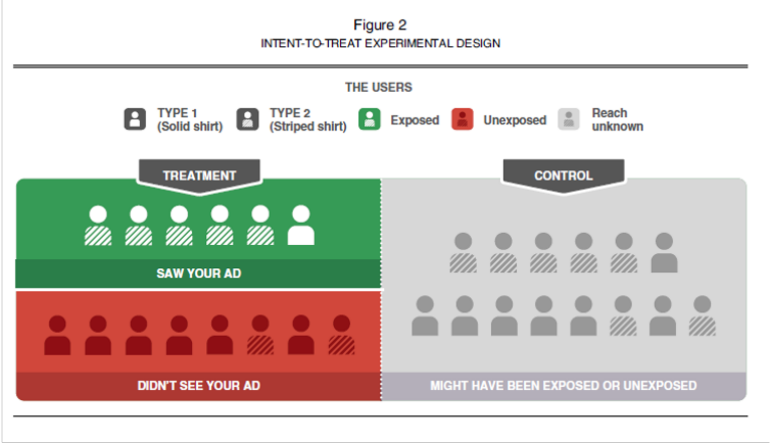


Figure 2
INTENT-TO-TREAT EXPERIMENTAL DESIGN
THE USERS
TYPE 1 (Solid shirt)
TYPE 2 (Striped shirt)
Exposed
Unexposed
Reach unknown
TREATMENT
CONTROL
SAW YOUR AD
DIDN'T SEE YOUR AD
MIGHT HAVE BEEN EXPOSED OR UNEXPOSED

GHOST ADS

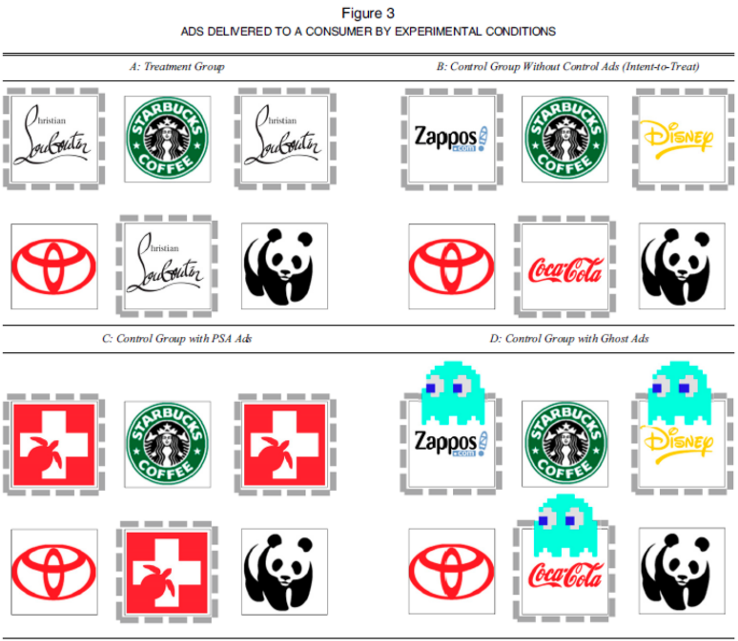
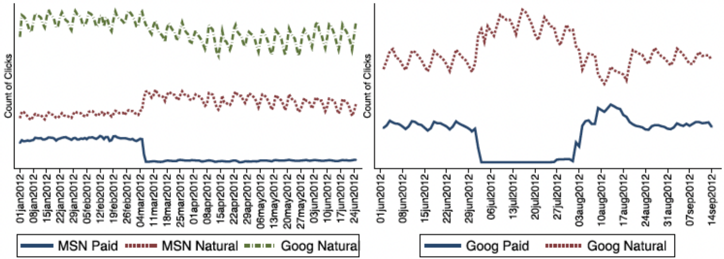


Figure 3
ADS DELIVERED TO A CONSUMER BY EXPERIMENTAL CONDITIONS
A: Treatment Group
B: Control Group Without Control Ads (Intent-to-Treat)
C: Control Group with PSA Ads
D: Control Group with Ghost Ads

GHOST ADS

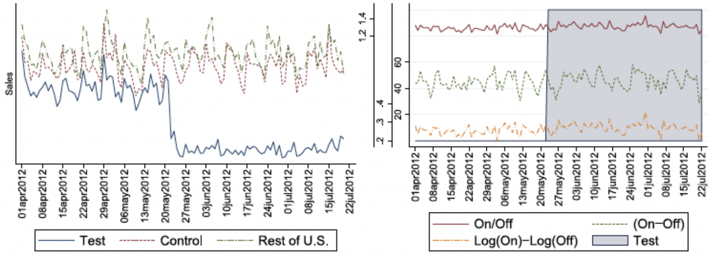
- Provide valid estimates of ATT.
- Hundreds of advertisers use Google’s ghost ad methodology, delivering millions of experimental impressions daily.
- More about other measurements issues in Johnson’s Inferno paper.

BLAKE, NOSKO, AND TADELIS - BRANDED SEARCH ADS



(a) MSN Test

(b) Google Test



(a) Attributed Sales by Region

(b) Differences in Total Sales

FIGURE 3.—Non-brand keyword region test. Panel (a) plots total purchases by users who clicked on an ad prior to purchase, which drops when the test commences in the test areas. Panel (b) plots three different measures of the difference between test and control regions before and after the test. The *y*-axis is shown for the ratio, the log difference, and in differences in thousands of dollars per day, per DMA.

EXPERIMENTAL DESIGN

- Used geographic bid feature to implement it at DMA level.
- Suspended non-brand ads for 30% of DMAs.
- Random subset of DMAs, split evenly to match the serial correlation in the data.
- Controversial, as they took several splits, found one with matching correlation and then flipped a coin once!
- Which unit of randomization would you use?

ANALYZING AN EXPERIMENT WITH A "DIFF-IN-DIFF" SPECIFICATION VS NAIVE OLS

$$ln(Sales_{it}) = \alpha_1 \times ln(Spend_{it}) + \varepsilon_{it}$$

$$ln(Spend_{it}) = \tilde{\alpha}_1 \times AdsOn_{it} + \tilde{\alpha}_2 \times Post_t + \tilde{\alpha}_3 \times Group_i + \varepsilon_{it}$$

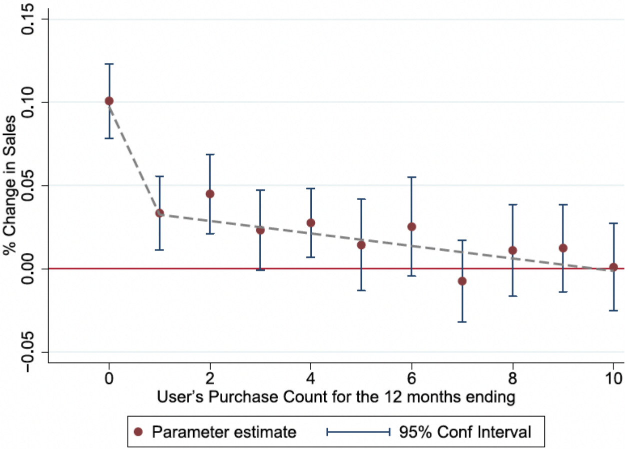
TABLE I

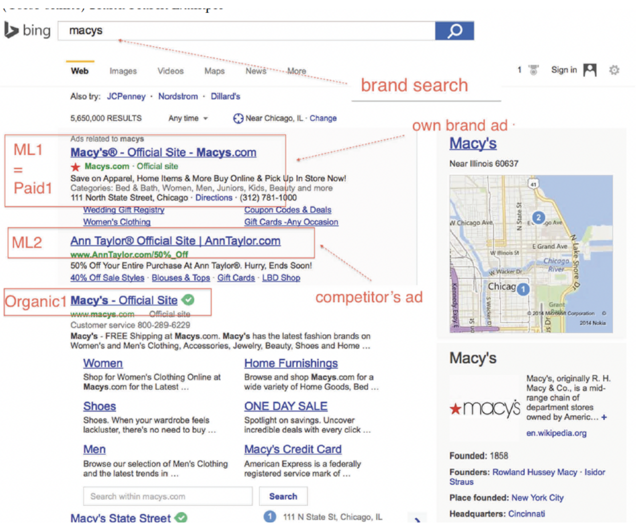
RETURN ON INVESTMENT[a]

	OLS		IV		DnD
	(1)	(2)	(3)	(4)	(5)
Estimated Coefficient	0.88500	0.12600	0.00401	0.00188	0.00659
(Std Err)	(0.0143)	(0.0404)	(0.0410)	(0.0016)	(0.0056)
DMA Fixed Effects		Yes		Yes	Yes
Date Fixed Effects		Yes		Yes	Yes
N	10,500	10,500	23,730	23,730	23,730
$\Delta \ln(Spend)$ Adjustment	3.51	3.51	3.51	3.51	1
$\Delta \ln(Rev)$ (β)	3.10635	0.44226	0.01408	0.00660	0.00659
Spend (Millions of $)	$51.00	$51.00	$51.00	$51.00	$51.00
Gross Revenue (R′)	2,880.64	2,880.64	2,880.64	2,880.64	2,880.64
ROI	4,173%	1,632%	−22%	−63%	−63%
ROI Lower Bound	4,139%	697%	−2,168%	−124%	−124%
ROI Upper Bound	4,205%	2,265%	1,191%	−3%	−3%

HETEROGENEITY

$$Sales_{imt} = \sum_{m=0}^{10} \beta_m \times AdsOn_{imt} \times \theta_m + \delta_t + \gamma_i + \theta_m + \varepsilon_{it}$$

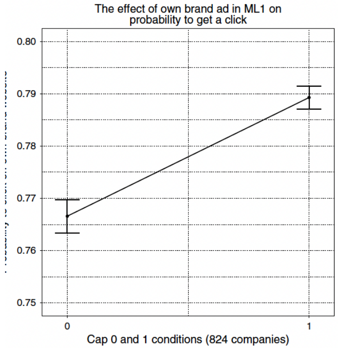


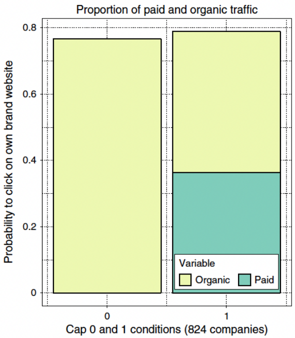

bing
macys
Web
Images
Videos
Maps
News
More
brand search
Sign in
Also try: JCPenney · Nordstrom · Dillard's
5,650,000 RESULTS
Any time
Near Chicago, IL · Change
own brand ad
Ads related to macys
ML1
=
Paid1
Macy's® - Official Site - Macys.com
Macys.com · Official site
Save on Apparel, Home Items & More Buy Online & Pick Up In Store Now!
Categories: Bed & Bath, Women, Men, Juniors, Kids, Beauty and more
111 North State Street, Chicago · Directions · (312) 781-1000
Wedding Gift Registry
Coupon Codes & Deals
Women's Clothing
Gift Cards -Any Occasion
ML2
Ann Taylor® Official Site | AnnTaylor.com
www.AnnTaylor.com/50%_Off
50% Off Your Entire Purchase At Ann Taylor®. Hurry, Ends Soon!
40% Off Sale Styles · Blouses & Tops · Gift Cards · LBD Shop
competitor's ad
Organic1
Macy's - Official Site
www.macys.com Official site
Customer service 800-289-6229
Macy's - FREE Shipping at Macys.com. Macy's has the latest fashion brands on Women's and Men's Clothing, Accessories, Jewelry, Beauty, Shoes and Home ...
Women
Shop for Women's Clothing Online at Macys.com for the Latest ...
Home Furnishings
Browse and shop Macys.com for a wide variety of Home Goods, Bed ...
Shoes
Shoes. When your wardrobe feels lackluster, there's no need to buy ...
ONE DAY SALE
Spotlight on savings. Uncover incredible deals with every click ...
Men
Browse our selection of Men's Clothing and the latest trends in ...
Macy's Credit Card
American Express is a federally registered service mark of ...
Search within macys.com
Search
Macy's State Street
111 N State St, Chicago, IL
Macy's
Near Illinois 60637
W Chicago Ave
E Grand Ave
W Illinois St
Chicago River
Macy's
Macy's, originally R. H. Macy & Co., is a mid-range chain of department stores owned by Americ... +
en.wikipedia.org
macys
Founded: 1858
Founders: Rowland Hussey Macy · Isidor Straus
Place founded: New York City
Headquarters: Cincinnati

METHODOLOGY

- **Field Experiment** on Bing (9 days, 2014):
  - Randomly capped mainline ads (0,1,2,3 vs control=4)
  - User-level randomization
- **Data**:
  - 2,500+ brands with >350 searches
  - 824 consistently advertising brands
- **Measures**:
  - Click probabilities (paid vs organic)
  - Cost Per Incremental Click (CPIC)

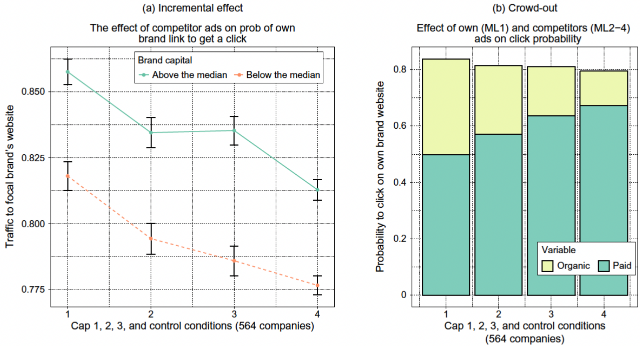
INCREMENTALITY - NO COMPETITOR ADS





INCREMENTALITY - WITH COMPETITOR ADS

Figure 5. (Color online) The Effect of Competitive Ads in Mainline Slots 2–4



COST PER INCREMENTAL CLICK DEPENDS ON COMPETITOR ADS

Our results suggest that advertising by competitors completely changes the story. A single competitor in the top position on the page, on average, steals 18% of clicks from a high traffic brand, but a competitor following a focal brand's ad steals only 1%–2% of clicks. If this difference is due to strong position effects and not selection issues, focal brand ads have a strong ROI. This is because the defense is highly effective (the total CTR returns almost to the case when there is no advertising): Even though the focal brands must pay for 50 clicks to get 18 incremental clicks, their CPC is about 10 times less than they pay on other queries. Putting the pieces together, the implied CPIC is in an attractive range and, indeed, better than usual.

CONCLUSION

- Returns to advertising are context dependent.
- Advertising must be thought of as a competitive game, just like pricing.
- Experiments enable ad effectiveness measurement at scale.
- Next lecture: learning about advertising without experiments.

NEXT LECTURE

- Observational Data - Please read the first half of chapter 9 in The Mixtape, and skim the rest.
- Read Shapiro (2018) and Shapire, Hitch, and Tuchman (2021).