HOW EFFECTIVE IS TARGETED ADVERTISING?

Ayman Farahat and Michael Bailey
Marketplace Architect Yahoo!

July 28, 2011

Thanks Randall Lewis , Yahoo! Research

Agenda

- An Introduction to Measuring Effectiveness

  1. Introduction
     – *Naïve estimation*
  2. Measuring Treatment Effect
     – *Overview of econometrics literature.*
  3. Measuring Targeting Study
     – *Case study: results from a large scale randomized experiment*

- Summary: Measuring the Treatment Effect

Which Advertising Method is More Effective?

Makes $3 profit
for each
lemonade

RON

- Puts an Ad in location X
  - 100 people see the Ad
  - 20 people buy lemonade
  - Cost of reaching a person is $0.20

Conversion rate = 20%
Cost of conversion = (100*0.2)/20 = $1.0

- What if there were no Ad?
  Of the 100 people who would have seen the Ad, 10 buy anyway.
- Return on Advertising Spending
  (20-10)*3 -100*0.2 = $10

TARG

- Puts an Ad in location Y
  - 100 people see the Ad
  - 60 people buy lemonade
  - Cost of reaching a person is $0.40

Conversion rate = 60%
Cost of conversion = (100*0.4)/60 = $.66

- What if there were no Ad?
  Of the 100 people who would have seen the Ad, 50 buy anyway.
- Return on Advertising Spending
  (60-50)*3 -100*0.4 = $5

Issues with Naïve Estimation in Advertising

	Selected	Not Selected
Treated	60%	20%
Not Treated	50%	10%
Lift	20%	50%

- Advertisers show their ads to users who are likely to respond.
- Users who are selected get treated with Ads.
- Users not selected are not treated.
- Estimate lift by comparing two different populations.

OVERVIEW OF ECONOMETRIC TECHNIQUES

Measuring Effectiveness: Data Sources

- The *gold standard* is a randomized experiment.
- Randomized experiment:

  Assignment randomized by Experimenter.
- Natural experiment:

  Assignment randomized by “nature”.
- Observational data:

  Assignment has not been randomized and the experimenter has no control.

Measuring Effectiveness: Econometric Methods

Data	Observational data
Problem	Selection bias Omitted variable bias Simultaneous causality

Method	•Matching Estimator
	    •Propensity Score and matching
	    •Regression Discontinuity
	    •Heckman Correction
	• Regression
	    •2SLS
	    •Instrumental Variable

Endogeneity Issues in Measuring Ad Effectiveness

- Y_{i1} (Y_{i0}) is response of individual "i" to treatment (control).
- D_i is an indicator variable equal to 1 if individual "i" is treated.
- The average treatment effect is $E(Y_{i1} – Y_{i0})$.
- Average Treatment Effect on Treated (ATET): $E(Y_{i1}-Y_{i0}|D_i =1)$.
- The Naïve estimator introduces a bias:

$$
\begin{aligned}
E(Y_i \mid D_{i=1}) - E(Y_i \mid D_{i=0})\} &= \\
&= E(Y_{i1} \mid D_1) - E(Y_{i0} \mid D_1) + E(Y_{i0} \mid D_1) - E(Y_{i0} \mid D_0) \\
&= \underbrace{\{E(Y_{i1} \mid D_1) - E(Y_{i0} \mid D_1)\}}_{ATET} + \underbrace{\{\mathbf{E}(\mathbf{Y_{i0}} \mid \mathbf{D_1}) - E(Y_{i0} \mid D_0)\}}_{\text{selectionbias}}
\end{aligned}
$$

Endogeneity Issues in Measuring Ad Effectiveness

- Advertisers select users who are more likely to respond.
  - Treatment selection is not exogenous.
- The term $E(Y_{i0}|D_1)$ is the response of the users who would have been treated but do not get treated.
- Example: Re-targeting
  - If a user goes to Advertiser's site, is user likely to convert?
  - The probability of user converting without an Ad is high.

Matching Estimator: Overview

- If the treatment assignment is completely random, then we can compare test and control.
- In the case of observational data, we don't have a randomized test and control.
- Given a set of users who have been treated, we create a "matched" or "synthetic" control.
- Compare the test group to the matched control group.
- Examples include Yahoo! Advertiser Analytics (YAA).

Matching Estimator: Example

- Yahoo Advertiser Analytics (YAA) provides advertising insights.
- Advertisers target their Ads based on demographics, techno-graphics, etc.
- Users who meet targeting criteria are shown Ads (treated).
- How effective is Advertising?
  - Construct a control group composed of users who have the same targeting criteria as users who saw Ads.
  - Compare the response of the treated and control group.

Matching Estimator: Mechanics

- There is a set of observable variables "S" that fully capture the heterogeneity between users.
- Examples include demographics, web behavior, etc.
- Within a strata of S, the residual variation in assignment is:
  - 1) totally random, 2) uncorrelated with outcome.
- The counterfactual response of the user who gets treated (not treated) is same as a similar user who is not treated (treated).
- The treatment assignment is "ignorable"

Matching Estimator: Mechanics

$$E(Y^1 \mid d=1,s) = E(Y^1 \mid d=0,s)$$

$$E(Y^0 \mid d=1,s) = E(Y^0 \mid d=0,s)$$

$$(ATET) = \mathrm{E}(Y^1 - Y^0 | d=1,s)$$

$$= \mathrm{E}(Y^1 \mid d=1,s) \text{ -}\mathrm{E}(Y^0 | d=1,s)$$

$$= \mathrm{E}(Y^1 \mid d=1,s) \text{ -}\mathrm{E}(Y^0 | d=0,s)$$

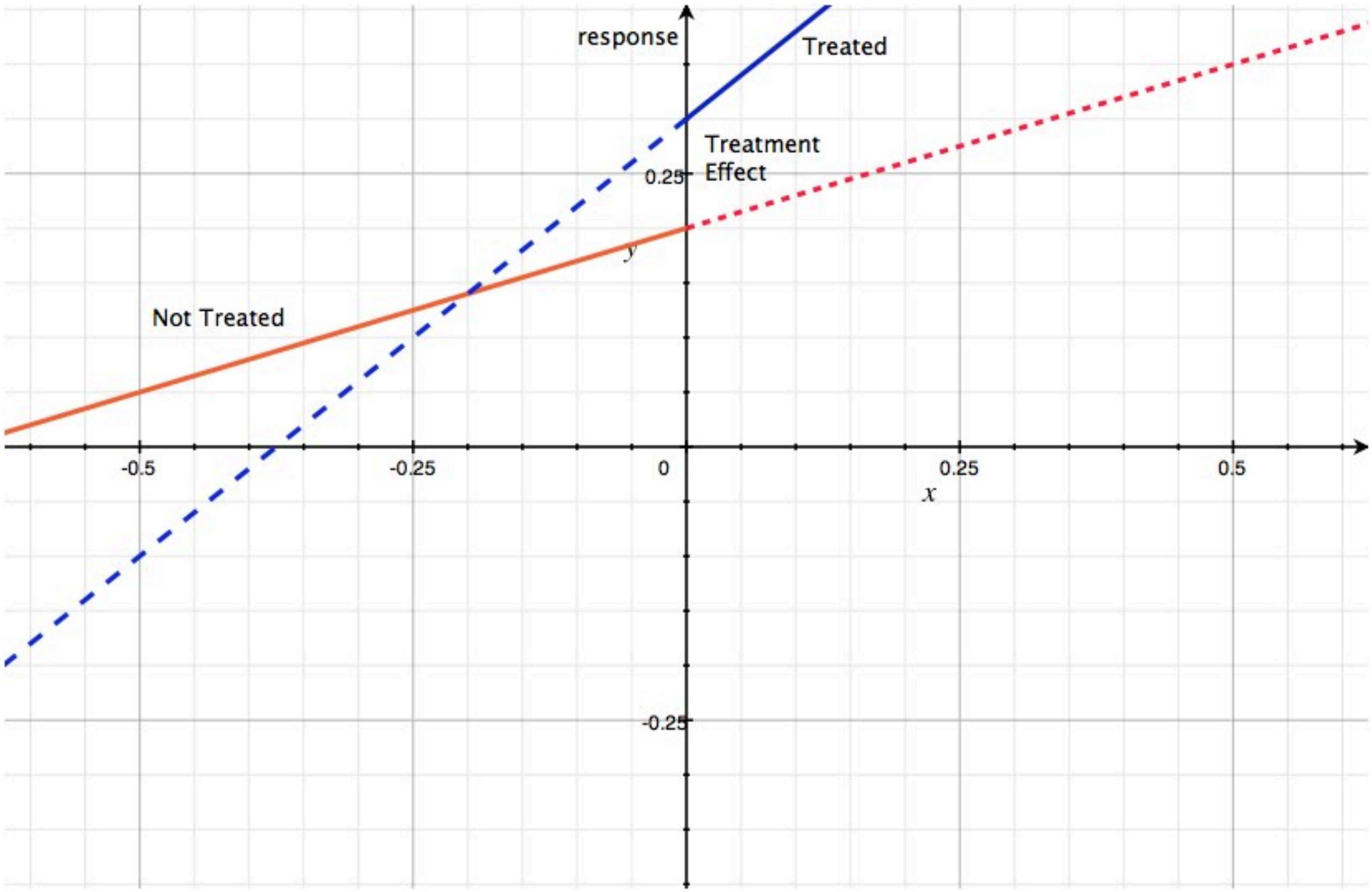
Regression Discontinuity: Overview

- Regression discontinuity estimates the causal effect of treatment by exploiting a given exogenous threshold determining assignment to treatment.
- Subjects above (below) the threshold get treated (not treated).
- Subjects right above (below) threshold can serve as test (control).
- Examples include evaluating look-a-like models.

Regression Discontinuity: Example

- Conjecture : Users who look like users who converted are also likely to convert.
- Build a model to score users based on their propensity to convert.
- Users above (below) a certain threshold are shown (not) Ads.
- Users who have a score of 0.5001 are treated.
- Users who have a score of 0.4999 are not treated.
- Assignment to test and control around score of 0.5 is random.
- Users with score of 0.5001 (0.4999) are test (control) group.

Regression Discontinuity: Mechanics



Regression: Overview

- In a number of cases, the treatment is not dichotomous.
  - Example: number of Ad impressions.
- Other confounding but observable factors are believed to influence response.
  - Examples: Age, gender, and income influence sales.
- The goal is to estimate the impact of advertising.
  - Estimate advertising elasticity.
- Some examples come from Marketing Mix Models (MMM).

Regression: Example

- Sales is a function of advertising effort.
- Advertising effort is a function of sales.

$$s^{t=T} = \alpha_0 + \beta_1 a_1^{t=T}$$

$$a_t^{t=T} = \gamma_0 + \phi_1 s_1^{t=T}$$

$$s^{t=T-1} = \alpha_0 + \beta_1 a_1^{t=T-1}$$

$$a_t^{t=T-1} = \gamma_0 + \phi_1 s_1^{t=T-1}$$

$$s^{t=T-2} = \alpha_0 + \beta_1 a_1^{t=T-2}$$

$$a_t^{t=T-2} = \gamma_0 + \phi_1 s_1^{t=T-2}$$

Regression: Mechanics

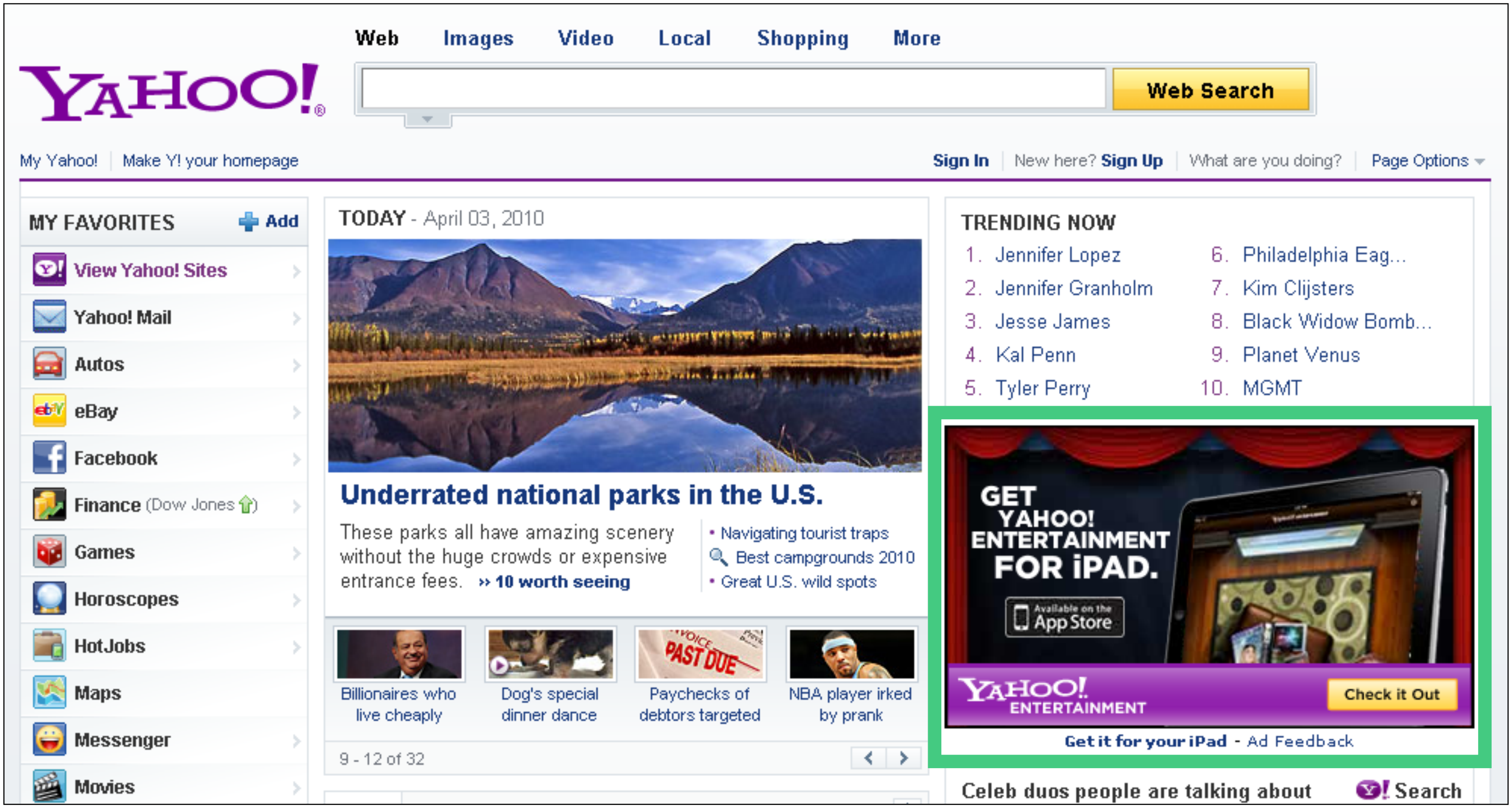
- The Advertising effort is not exogenous.
- Failure to take into account the dependence of advertising on sales leads to biased estimates.
- Solved using 2SLS or any of its variants.
- All MMM models are variants of the Bass 68 model.

CASE STUDY 1:

Results from a Large Scale Randomized Experiment

The Y! Front Page is an excellent website for a targeting field study

www.yahoo.com



Front page split campaigns provide a natural experiment to examine targeting for many users

Hypothetical example: Verizon and Home Depot

- Share day's traffic based on impression's timestamp:
  - Verizon gets "even second" (0,2,4,…) traffic
  - Home Depot gets "odd second" (1,3,5,…) traffic

Pseudo-random ad delivery

- Users choose how many times to visit Y! Front Page
- Each visit is like a coin toss
- # of heads? Verizon
- # of tails? Home Depot

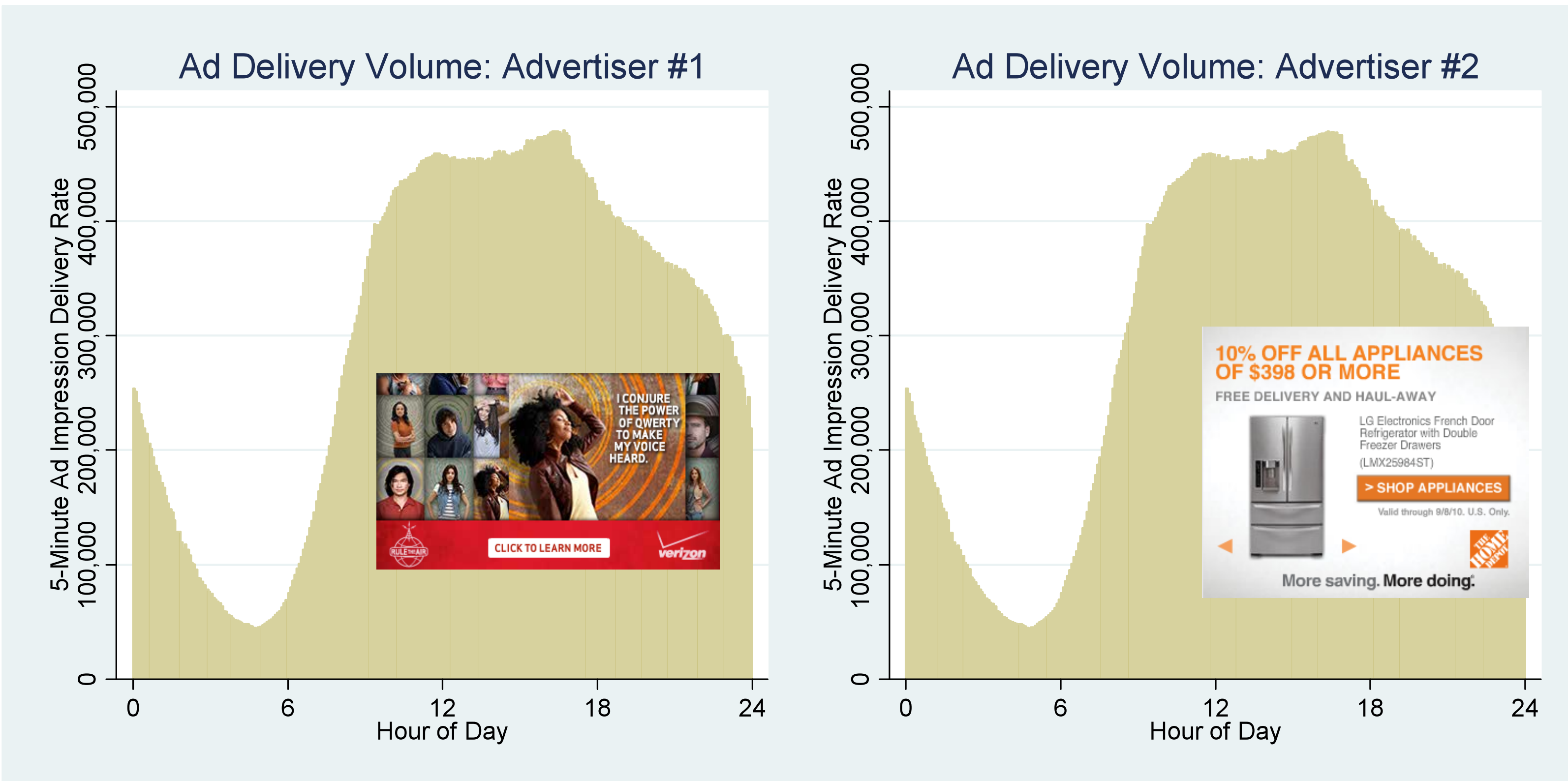
"Natural experiment"

- "Exogenous variation" ☺
- "Endogenous variation" ☹

The split advertisers' delivery on even and odd seconds was statistically identical.

5-minute ad-delivery rates look identical over 24 hours.

- Peak at 5pm Eastern and trough at 5am Eastern.



Randomized Experiment for Estimating Impact of Targeting

- Randomly split the users who belong to the targeting category into test and control.
- Account for unobserved heterogeneity: compare similar users.
- In our case we restricted the analysis to users with one Ad view.
- Measure the impact of advertising on search through rate (STR).

	Seen Ad	No Ad	
Targeted	STR_{11}	STR_{10}	$\frac{(STR_{11} - STR_{10})}{STR_{10}}$
No Target	STR_{01}	STR_{00}	$\frac{(STR_{01} - STR_{00})}{STR_{00}}$
Lift	$\frac{(STR_{11} - STR_{01})}{STR_{01}}$	$\frac{(STR_{10} - STR_{00})}{STR_{00}}$	

The data has three parts: daily searches, raw ad views and clicks, and targeting membership.

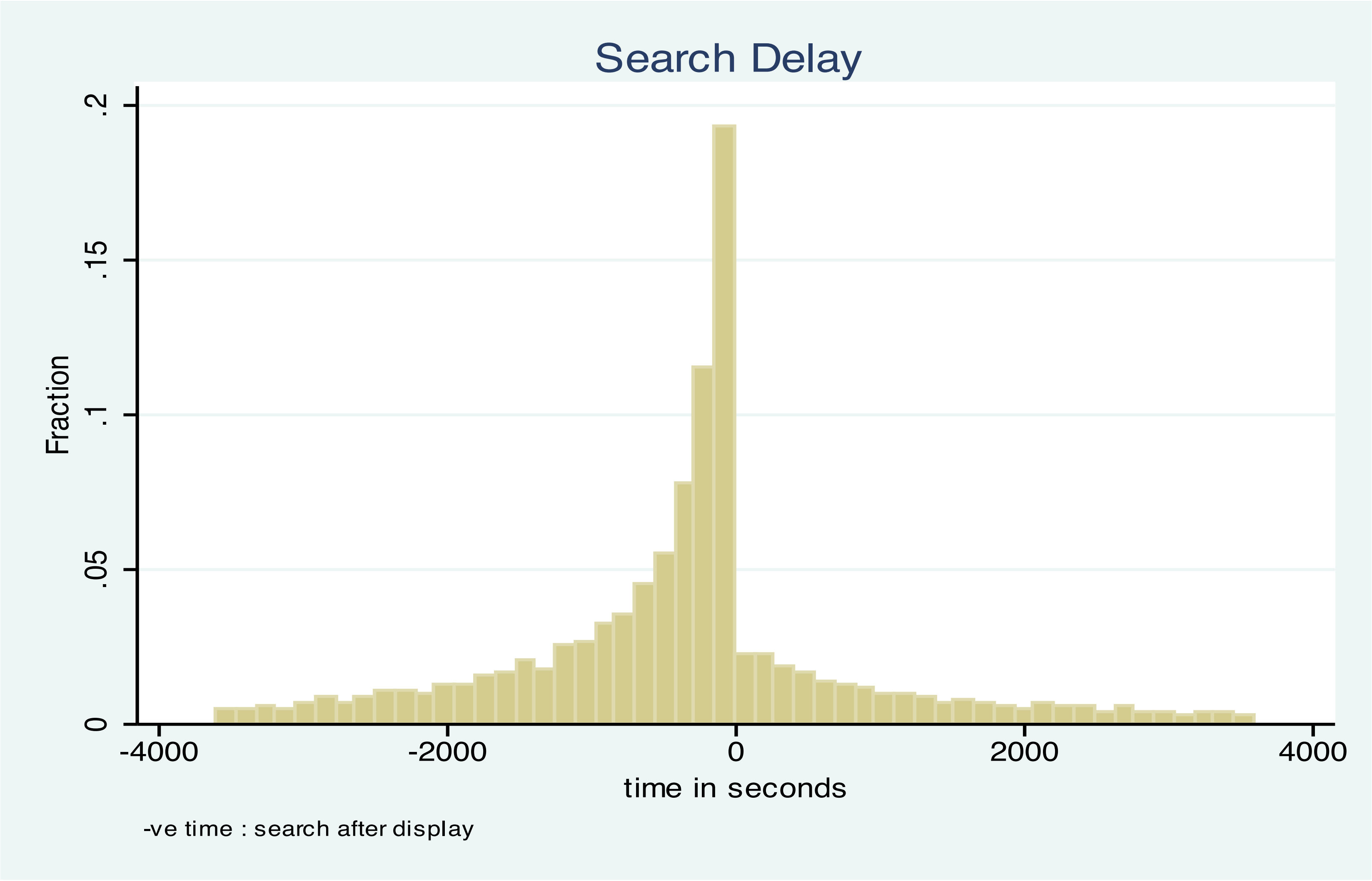
22 “split” campaigns: Feb – July 2011

- 2 campaigns per day
- Advertiser from different verticals : Telco, insurance, finance, retail, technology, pharmaceuticals, etc.
- 609 million users*days x 2 campaigns/users =
  1.2 billion campaign*user*days

Different creative technology

- Large range of creative campus: rich media, expanding Ads… etc.

Search Results: When do users search?



Search Results

AVERAGE	Seen Ad	Not Seen Ad	Lift
Target	0.949%	0.923%	3.856%
No Target	0.095%	0.093%	4.195%
Lift	896.737%	891.981%	

Search Results

Finance	Seen Ad	Not Seen Ad	Lift
Target	2.121%	2.322%	-8.620%
No Target	0.100%	0.097%	2.805%
Lift	2030.460%	2296.836%	

Insurance	Seen Ad	Not Seen Ad	Lift
Target	0.255%	0.254%	0.271%
No Target	0.034%	0.032%	5.125%
Lift	647.023%	683.185%	

Credit Card	Seen Ad	Not Seen Ad	Lift
Target	0.204%	0.190%	7.439%
No Target	0.068%	0.071%	-3.987%
Lift	198.572%	166.819%	

Insurance	Seen Ad	Not Seen Ad	Lift
Target	0.324%	0.307%	5.499%
No Target	0.041%	0.038%	10.445%
Lift	681.156%	717.777%	

Retail	Seen Ad	Not Seen Ad	Lift
Target	2.270%	2.276%	-0.256%
No Target	0.966%	0.966%	-0.041%
Lift	135.061%	135.568%	

Finance	Seen Ad	Not Seen Ad	Lift
Target	0.173%	0.180%	-3.781%
No Target	0.077%	0.078%	-1.483%
Lift	125.999%	131.397%	

Search Results

Finance	Seen Ad	Not Seen Ad	Lift
Target	0.250%	0.228%	9.873%
No Target	0.033%	0.033%	2.395%
Lift	648.992%	598.011%	

Insurance	Seen Ad	Not Seen Ad	Lift
Target	0.308%	0.312%	-1.332%
No Target	0.035%	0.036%	-1.163%
Lift	767.387%	768.873%	

Finance	Seen Ad	Not Seen Ad	Lift
Target	2.808%	2.569%	9.324%
No Target	0.113%	0.105%	7.235%
Lift	2392.614%	2344.979%	

Insurance	Seen Ad	Not Seen Ad	Lift
Target	0.288%	0.307%	-6.166%
No Target	0.036%	0.036%	-0.881%
Lift	698.863%	743.857%	

Credit Card	Seen Ad	Not Seen Ad	Lift
Target	2.762%	2.627%	5.119%
No Target	0.117%	0.115%	1.360%
Lift	2261.499%	2177.053%	

Insurance	Seen Ad	Not Seen Ad	Lift
Target	0.340%	0.339%	0.494%
No Target	0.041%	0.042%	-1.680%
Lift	724.450%	706.616%	

Search Results

Education	Seen Ad	Not Seen Ad	Lift
Target	0.174%	0.234%	-25.589%
No Target	0.037%	0.037%	0.613%
Lift	370.093%	535.625%	

Insurance	Seen Ad	Not Seen Ad	Lift
Target	0.310%	0.308%	0.931%
No Target	0.039%	0.040%	-2.847%
Lift	704.614%	674.495%	

Technology	Seen Ad	Not Seen Ad	Lift
Target	0.349%	0.469%	-25.621%
No Target	0.026%	0.023%	13.368%
Lift	1224.940%	1919.461%	

Pharma	Seen Ad	Not Seen Ad	Lift
Target	2.912%	2.846%	2.320%
No Target	0.018%	0.017%	1.296%
Lift	16424.400%	16259.044%	

Telco	Seen Ad	Not Seen Ad	Lift
Target	0.482%	0.373%	29.109%
No Target	0.026%	0.023%	14.412%
Lift	1741.627%	1531.999%	

Insurance	Seen Ad	Not Seen Ad	Lift
Target	0.347%	0.316%	9.888%
No Target	0.039%	0.041%	-3.825%
Lift	790.082%	679.008%	

Search Results

Entertainment	Seen Ad	Not Seen Ad	Lift
Target	0.414%	0.319%	29.858%
No Target	0.033%	0.032%	2.586%
Lift	1156.646%	892.733%	

Credit Card	Seen Ad	Not Seen Ad	Lift
Target	2.452%	2.482%	-1.210%
No Target	0.108%	0.101%	7.193%
Lift	2168.080%	2361.011%	

Technology	Seen Ad	Not Seen Ad	Lift
Target	0.455%	0.390%	16.541%
No Target	0.025%	0.023%	7.386%
Lift	1753.349%	1607.762%	

Entertainment	Seen Ad	Not Seen Ad	Lift
Target	0.872%	0.667%	30.747%
No Target	0.082%	0.062%	31.984%
Lift	967.537%	977.638%	

Results and Conclusions

- Advertising does work.
- Naïve estimate of search lift is 891%.
- When we take bias into account, lift drops to 4.79%.
- 3.85% lift (0.026% absolute) on targeted.
- 4.19% lift (0.002% absolute) on untargeted.
- 0.46% correlation between targeted and untargeted lifts.
- Ad creative plays a significant role.

THANK YOU!

Measuring effectiveness : Econometric Methods

Data	Observational data	Natural Experiment
Problem	Selection bias Omitted variable bias Simultaneous causality	We cannot observe the counterfactual; what if the treatment group had not received the treatment?
Method	•Matching Estimator •Propensity Score and matching •Heckman correction •Regression Discontinuity • Regression •Instrumental variable (IV) •2SLS	Difference in Differences

Matching Estimator :Propensity Score

If the treatment assignment D is completely random then we can compare test and control.
In case of observational data, some of the users
In some case, the assignment to treatment is not random ,
 for example might have more access to a subset of population
Within a strata of S, the reaming variation in assignment is 1)totally random , 2) uncorrelated
 with outcome.
The treatment assignment is “ignorable”

$$Selection : d_j = f(s_j) + e_j^1$$

$$output : y_j^2 = f(x_j, s_j) + e_j^2$$

$$E(\mathrm{cov}(e_j^1, e_j^2)) = 0$$

$$y_j^2 \ \mathrm{P} d_j \ | \ s$$

Matching Estimator

$$E(Y^1 \mid d=1,s) = E(Y^1 \mid d=0,s)$$

$$E(Y^0 \mid d=1,s) = E(Y^0 \mid d=0,s)$$

If someone who was treated responded in a certain way,
then the same person will respond exactly the same way like
someone who was not treated and had the same set of variables S.
We can estimate the treatment effect ATET

$$\mathrm{E}(Y^1 - Y^0 | d=1,s) = \mathrm{E}(Y^1 \mid d=1,s) \text{ -}\mathrm{E}(Y^0 | d=1,s)$$

$$=\mathrm{E}(Y^1 \mid d=1,s) \text{ -}\mathrm{E}(Y^0 | d=0,s)$$

The ATET can be directly estimated since all the quantities are directly observable

Matching Estimator

If someone who was treated responded in a certain way,
then the same person will respond exactly the same way like
someone who was not treated and had the same set of variables S.
We can estimate the treatment effect ATET

$$\mathrm{E}(Y^1 - Y^0 | d = 1, s) = \mathrm{E}(Y^1 \mid d = 1, s) - \mathrm{E}(Y^0 | d = 1, s)$$

$$= \mathrm{E}(Y^1 \mid d = 1, s) - \mathrm{E}(Y^0 | d = 0, s)$$

The ATET can be directly estimated since all the quantities are directly observable

Matching Estimator :Heckman

Assignment are not random, for example survey response or health insurance applicants.
Unlike the propensity score, not all variables are observable, for example we
cannot observe whether someone is risk averseness or altruism.
If the unobserved variable is correlated with the response, for example risk averseness impacts medial cost
Altruism impacts survey response.
Regression estimation is biased due to omitted variable bias.
Heckman approach leverages the correlation between the model errors to correct for bias.

$$Selection : d_j = f(s) + e_j^1$$

$$output : y_j^2 = f(x_j^2) + e_j^2$$

$$E(\mathrm{cov}(e_j^1, e_j^2)) \neq 0$$

Regression Discontinuity : Mechanics

RD estimates the causal effect of treatment by exploiting a given exogenous threshold determining assignment to treatment.
In the case of scholarships for example, students who have a grade higher than a certain threshold are awarded threshold.
Students below the threshold are not awarded scholarship.
What is the casual impact of scholarship on future earning?
In case of advertising, sophisticated models are used to score users.
Users above the threshold are shown the ads (treated) while users below the threshold are not shown Ads
What is the casual impact of the Ad on conversions?

Regression

•Estimate the impact of the treatment D on outcome y.

•If the assignment is totally random D i

$$y = \alpha + \lambda D + \varepsilon$$

$$\hat{\lambda} = \frac{\mathrm{cov}(y_i, d_i)}{Var(d_i)} = E[y_i \mid d_i = 1] - E[y_i \mid d_i = 0]$$

•If the assignment is not totally random, we need to take a deeper look at the error term.
•If the error term includes variables that are correlated with the treatment, for example, in
•case of advertising, we only target users who have visited a site.
•If the site impacts the outcome, then our estimate of the treatment effect is biased.
•Expand the model to account for additional variables.

$$y = \alpha + \beta x + \lambda D + \varepsilon$$

IV: basic idea

Suppose we want to estimate a treatment effect using observational data

The OLS estimator is biased and inconsistent (due to correlation between regressor and error term) if there is

- omitted variable bias
- selection bias
- simultaneous causality

If a direct solution (e.g. including the omitted variable) is not available, instrumental variables regression offers an alternative way to obtain a consistent estimator

IV: basic idea

Consider the following regression model:

$$y_i = \beta_0 + \beta_1 X_i + e_i$$

Variation in the endogenous regressor X_i has two parts

- the part that is uncorrelated with the error ("good" variation)
- the part that is correlated with the error ("bad" variation)

The basic idea behind instrumental variables regression is to isolate the "good" variation and disregard the "bad" variation

IV: conditions for a valid instrument

The first step is to identify a valid instrument

A variable Z_i is a valid instrument for the endogenous regressor

X_i if it satisfies two conditions:

1. **Relevance: corr $(Z_i , X_i) \neq 0$**

2. **Exogeneity: corr $(Z_i , e_i) = 0$**