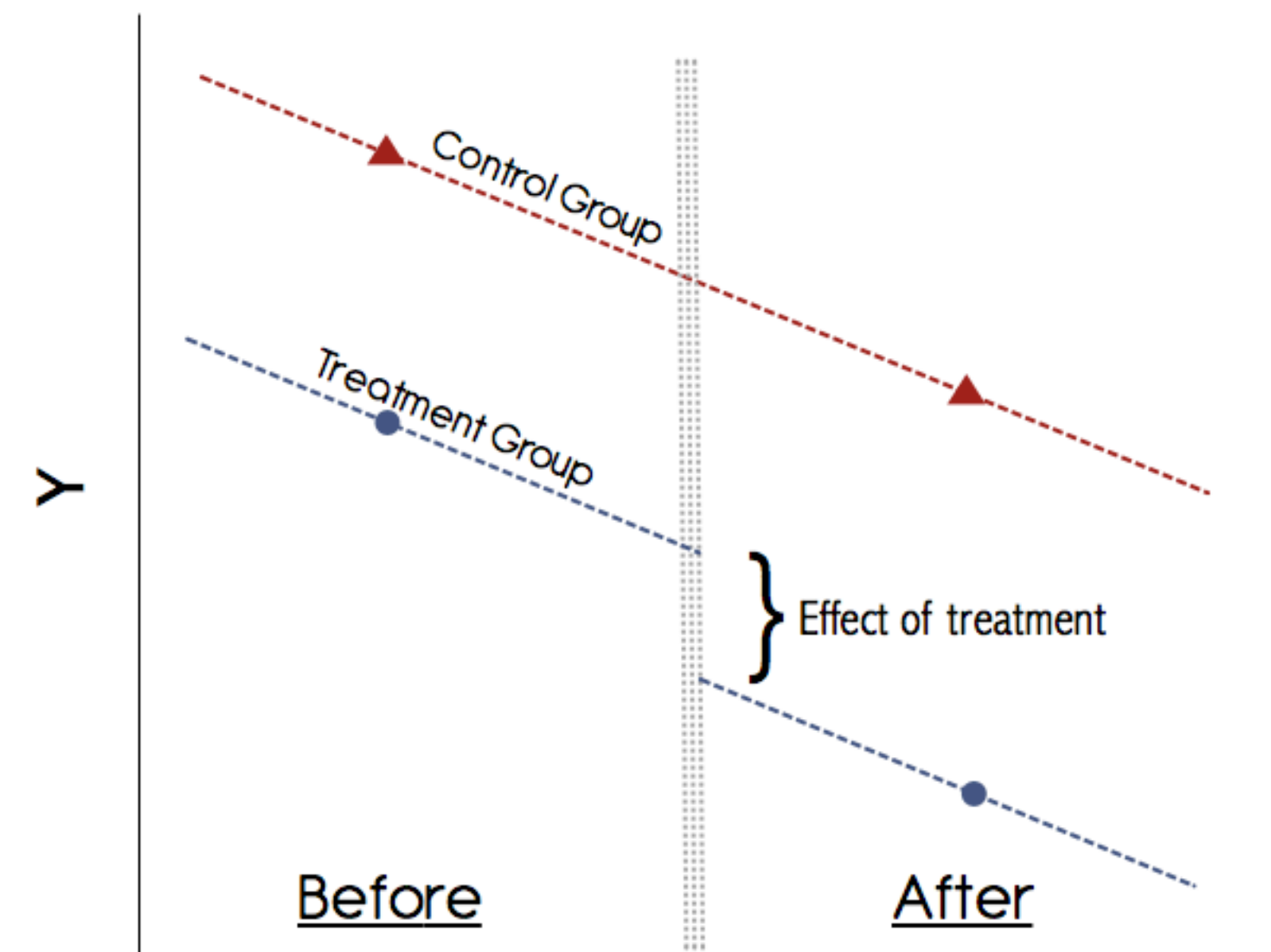


Business Experimentation and Causal Methods

Prof. Fradkin
Topic: Difference-in-Differences



This Time

1. **Fixed Effects**
2. Difference in Difference - with an experiment.
3. Interpreting logarithms in regressions.

A fixed effect is a special type of covariate.

- Suppose we had a column: personid - where each person gets a different id.
- If we have 4 unique persons, we create 4 new columns:
 - is it person 1? (1 or 0)
 - is it person 2? (1 or 0)
 - is it person 3? (1 or 0)
 - is it person 4? (1 or 0)
- We then add 3 of the 4 new columns to the regression as covariates.
- Note, we must always remove one fixed effect column if we're including a constant term in the regression.

Intuition of Fixed Effects

- There are persistent differences across units (individuals, cities, stores, etc...).
- Some workers tend to be more productive than others.
- Some markets have more sales than others. E.g. New York City vs Boston.
- Some people have higher cholesterol rates than others.
- Fixed effects allow us to ‘control’ for this.

Very Important For: Repeated Observations

- TV Advertising: We observe sales in each market before and during the ad campaign.
- Medicine: We observe cholesterol before and after taking statins (a medicine to reduce cholesterol).
- Employee performance, we observe each worker in week both before and during the experiment.

Setting: Categorical Variables With Many Values

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

- Have a variable with many values, such as passenger class in the above table of Titanic passengers.
- One option, convert to categorical and add to regression.
- Another option, use fixed effect notation.

Package: PyFixest

```
from pyfixest.estimation import feols  
from pyfixest.utils import get_data  
from pyfixest.summarize import etable
```

- Port of a very powerful R package called fixest.
- Has great features for fixed effects, robust standard errors, outputting regressions, difference in differences estimation.

Note, this gives the same answer!

```
# Convert to categorical:
df_titanic['Pclass_cat'] = df_titanic['Pclass'].astype('category')

model_nofe = feols("Survived ~ Age + Pclass_cat", data = df_titanic).vcov("hetero")
model = feols("Survived ~ Age | Pclass", data = df_titanic).vcov("hetero")
etable([model_nofe, model])
```

✓ 0.0s

	est1	est2
depvar	Survived	Survived
Intercept	0.969*** (0.055)	
Age	-0.008*** (0.001)	-0.008*** (0.001)
Pclass_cat[T.2]	-0.245*** (0.051)	
Pclass_cat[T.3]	-0.524*** (0.041)	
Pclass	-	x
R2	0.181	0.181
S.E. type	hetero	hetero
Observations	714	714

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001

Can put in many fixed effects

```
# Can put in many fixed effects:
```

```
model = feols("Survived ~ Age | Pclass + Embarked", data = df_titanic).vcov("hetero")  
etable(model)
```

✓ 0.0s

	est1
depvar	Survived
Age	-0.008*** (0.001)
Embarked	x
Pclass	x
R2	0.188
S.E. type	hetero
Observations	712

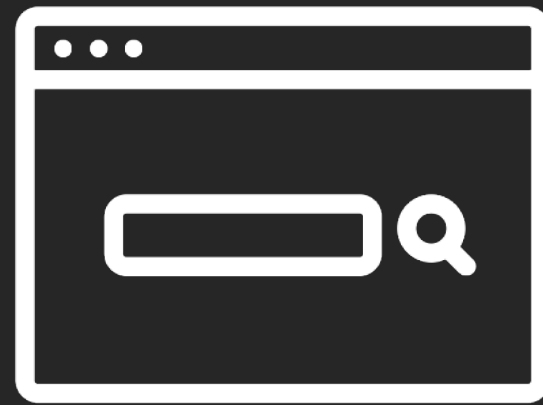
Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001

Fixed Effects Regression: Warnings

- We can't include fixed effects and variables that don't vary within the unit of the fixed effect.
- Example: we can't put in a variable for 'born in the usa' and person fixed effects.
The reason is that the variable 'born in the usa' is constant within a person, so it is 'absorbed' by the fixed effect.
- Example: We can't put in year of birth for the same reasons.
- Example: We can put in something that changes in our dataset for that person, such as whether you work from home or hours worked.

This Time

1. Fixed Effects
2. **Difference in Difference - with an experiment.**
3. Interpreting logarithms in regressions.

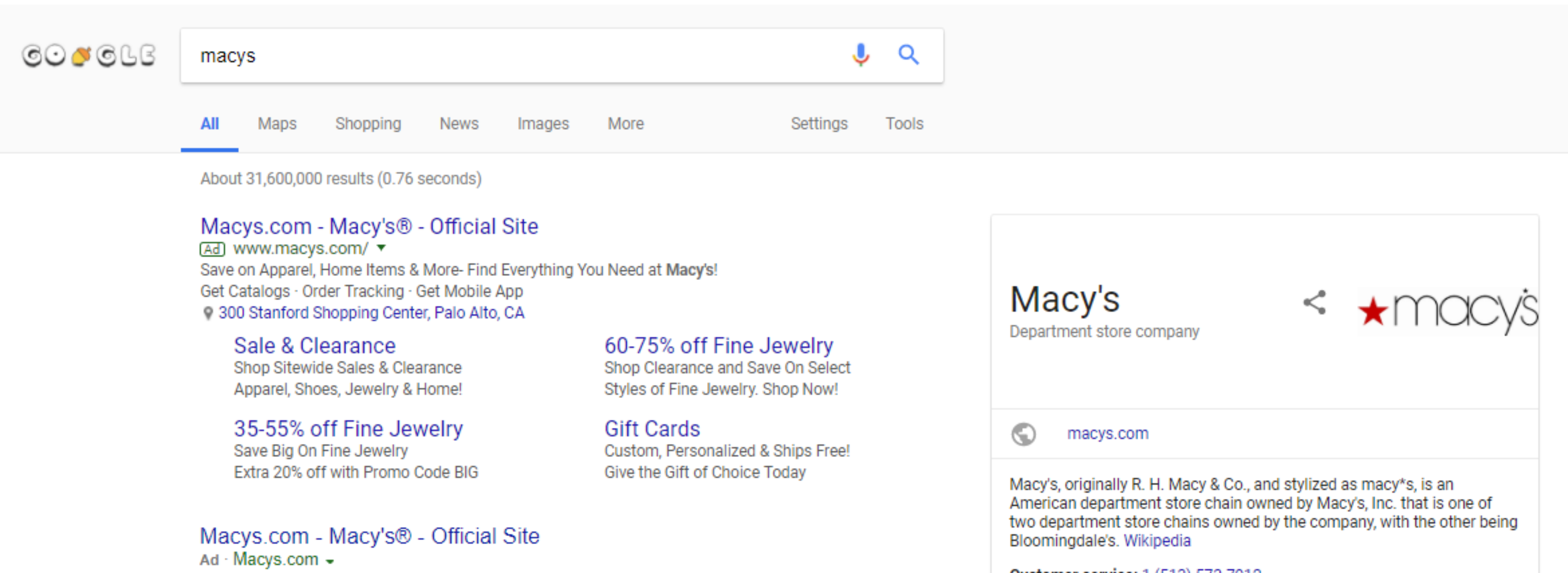


BRAND SEARCH ADS “INTERVENTION” EXAMPLE

Blake, Nosko, and Tadelis (2015)

Causal Problem

- What is the effect of Search Advertising?



The image is a screenshot of a Google search interface. At the top, the Google logo is on the left, and a search bar contains the text 'macys'. To the right of the search bar are icons for voice search and a magnifying glass. Below the search bar, there are tabs for 'All', 'Maps', 'Shopping', 'News', 'Images', 'More', 'Settings', and 'Tools'. The 'All' tab is selected. Below the tabs, it says 'About 31,600,000 results (0.76 seconds)'. The main search results area shows a top result for 'Macys.com - Macy's® - Official Site' with a green 'Ad' label. Below this, there are four promotional tiles: 'Sale & Clearance', '60-75% off Fine Jewelry', '35-55% off Fine Jewelry', and 'Gift Cards'. To the right of the search results is a knowledge panel for 'Macy's', which includes the company logo, a share icon, the website 'macys.com', and a brief description of the company.

GOOGLE

macys

All Maps Shopping News Images More Settings Tools

About 31,600,000 results (0.76 seconds)

Macys.com - Macy's® - Official Site
Ad www.macys.com/ ▼
Save on Apparel, Home Items & More- Find Everything You Need at **Macy's!**
Get Catalogs · Order Tracking · Get Mobile App
📍 300 Stanford Shopping Center, Palo Alto, CA

Sale & Clearance
Shop Sitewide Sales & Clearance
Apparel, Shoes, Jewelry & Home!

60-75% off Fine Jewelry
Shop Clearance and Save On Select
Styles of Fine Jewelry. Shop Now!

35-55% off Fine Jewelry
Save Big On Fine Jewelry
Extra 20% off with Promo Code BIG

Gift Cards
Custom, Personalized & Ships Free!
Give the Gift of Choice Today

Macys.com - Macy's® - Official Site
Ad · Macys.com ▼

Macy's
Department store company

🌐 macys.com

Macy's, originally R. H. Macy & Co., and stylized as macy's, is an American department store chain owned by Macy's, Inc. that is one of two department store chains owned by the company, with the other being Bloomingdale's. [Wikipedia](#)

Customer service: 1 (512) 572-7012

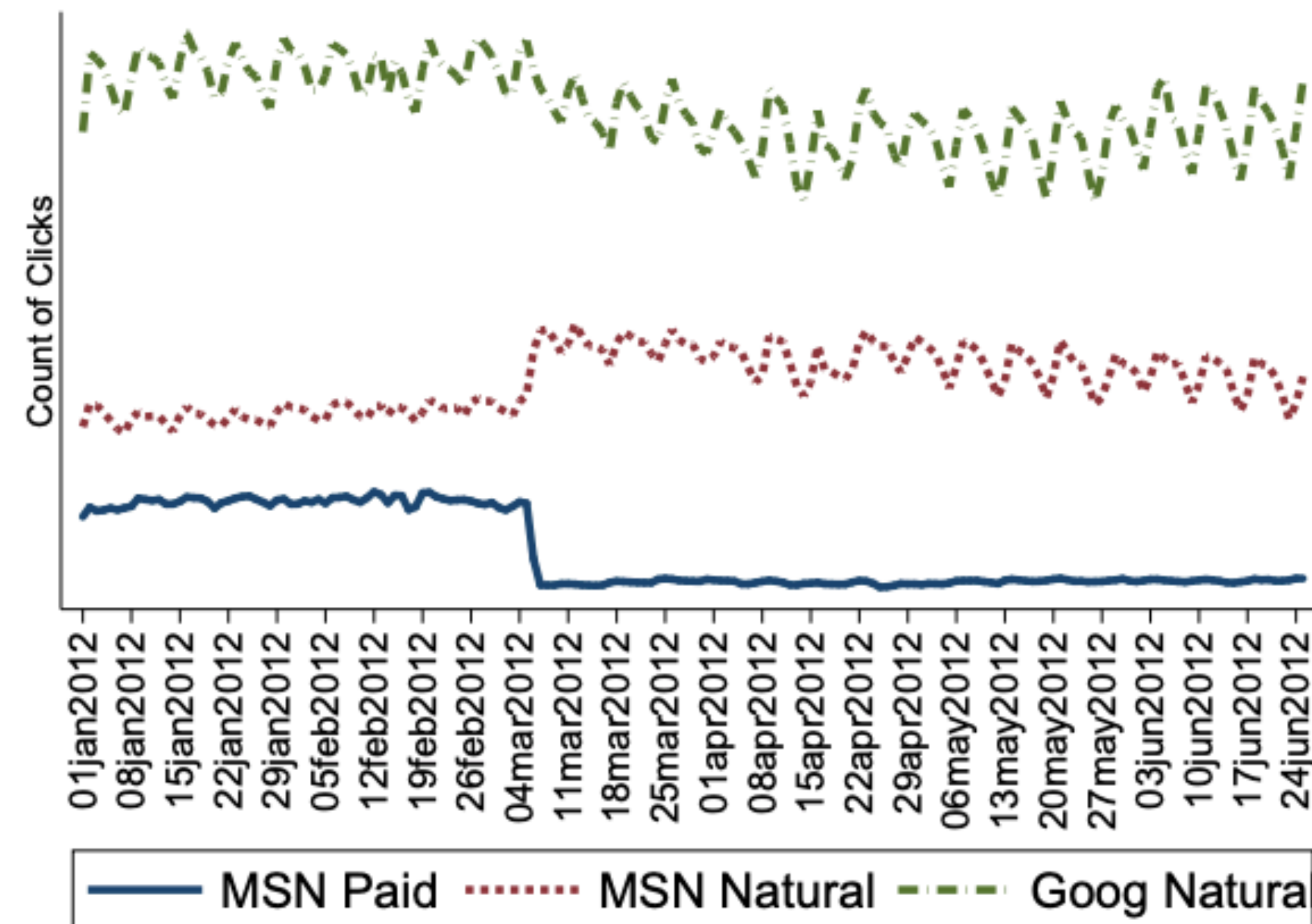
Bad way of doing this.

- Attribute the sale to the last click.
 - Makes the assumption that if you click on an ad, that's what causes the sale.
- But, this may just be driven by searchers for your product.

eBay stopped spending money on Bing (MSN).

Evidence suggestive that paid clicks steal clicks from organic search results.

Note, this is not an experiment, since no randomization or control group



eBay ran a follow-up experiment to test this.

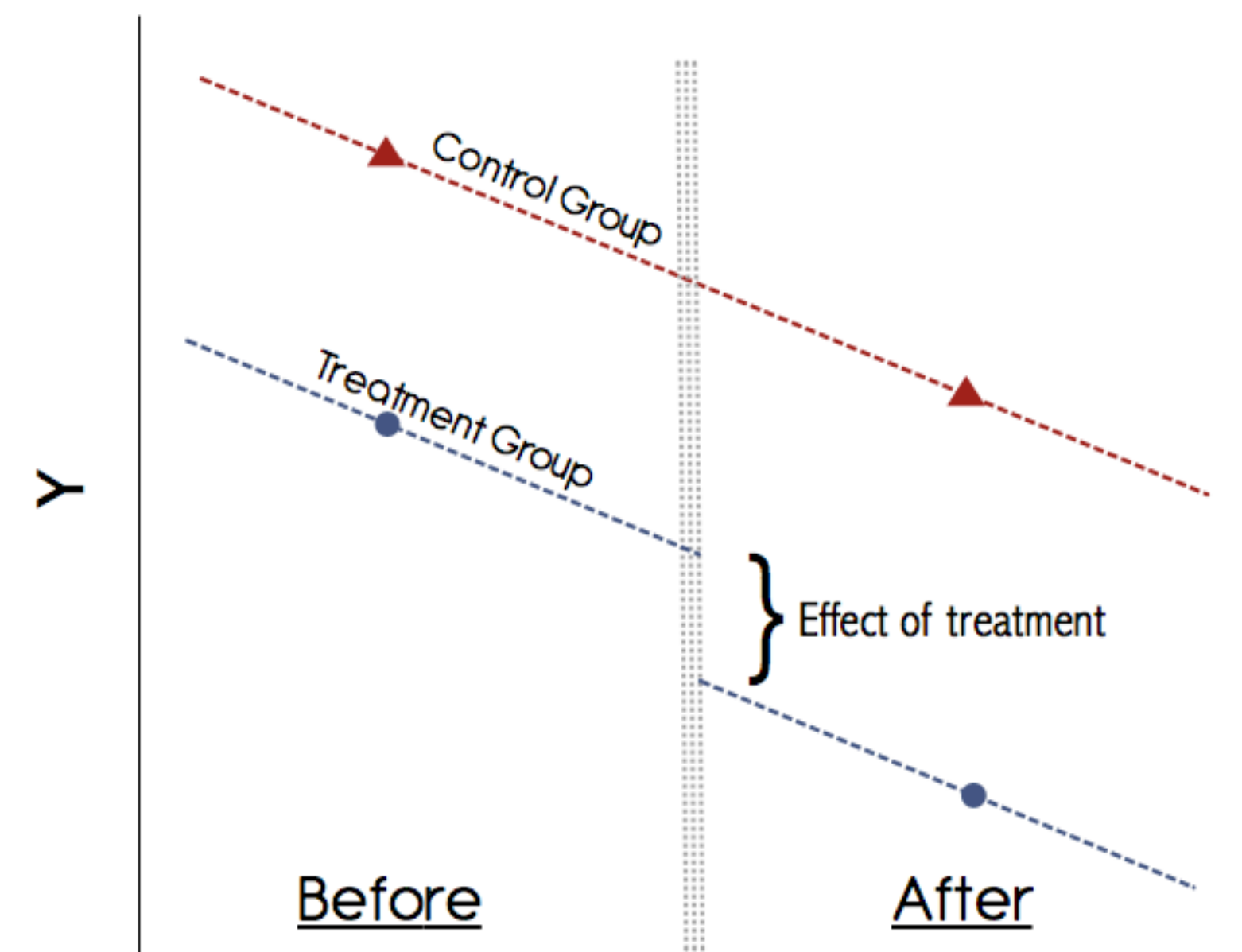
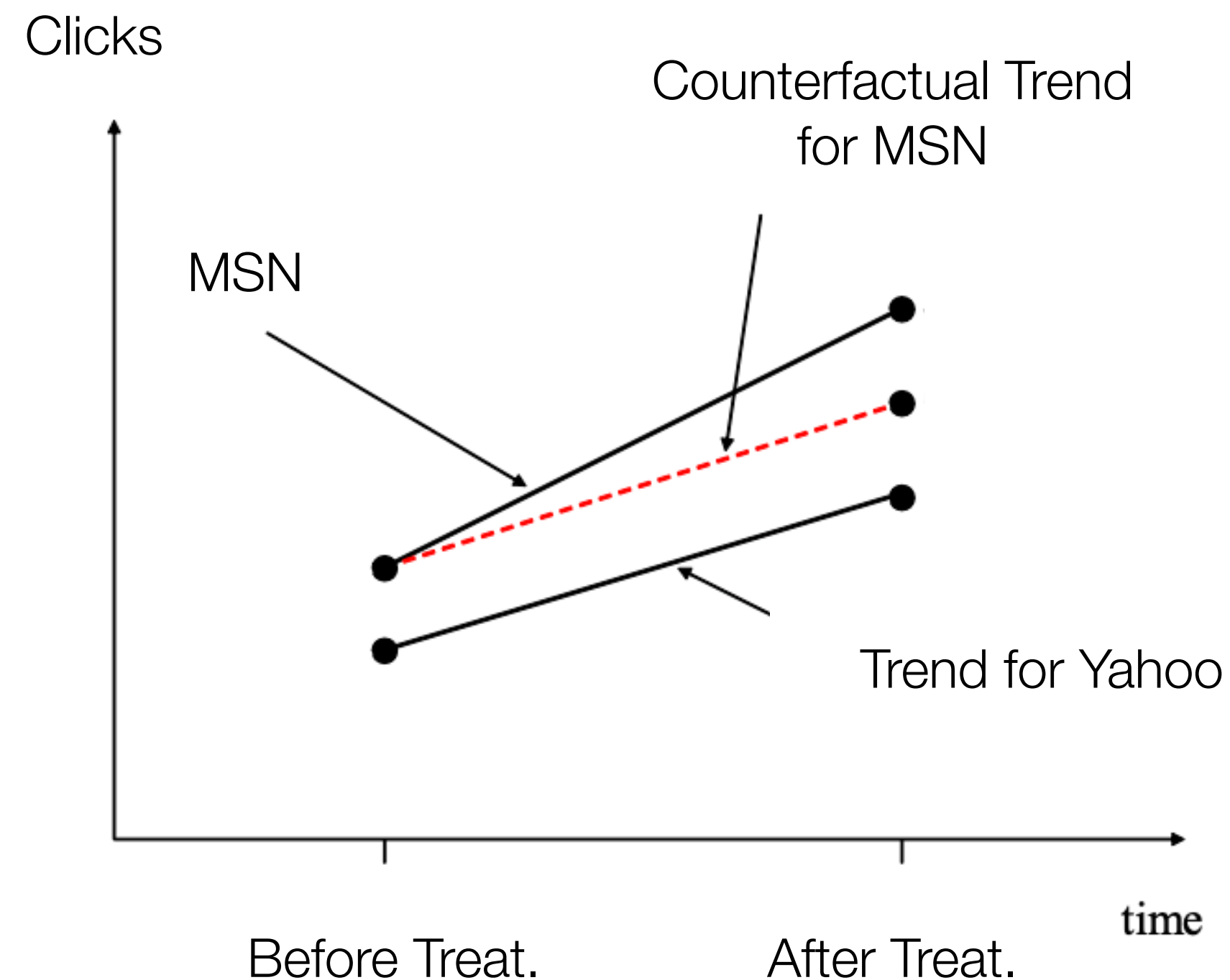
- Analyzed using a technique called difference-in-differences in order to make the estimates more precise.

Difference-in-Differences

- Also called, Diff-in-Diff, or DiD
- It is a technique where we regress an outcome on a treatment that gets turned on for one group and not for another at a particular time.
- We analyze it with a regression containing unit and time fixed effects, and the treatment as a covariate.

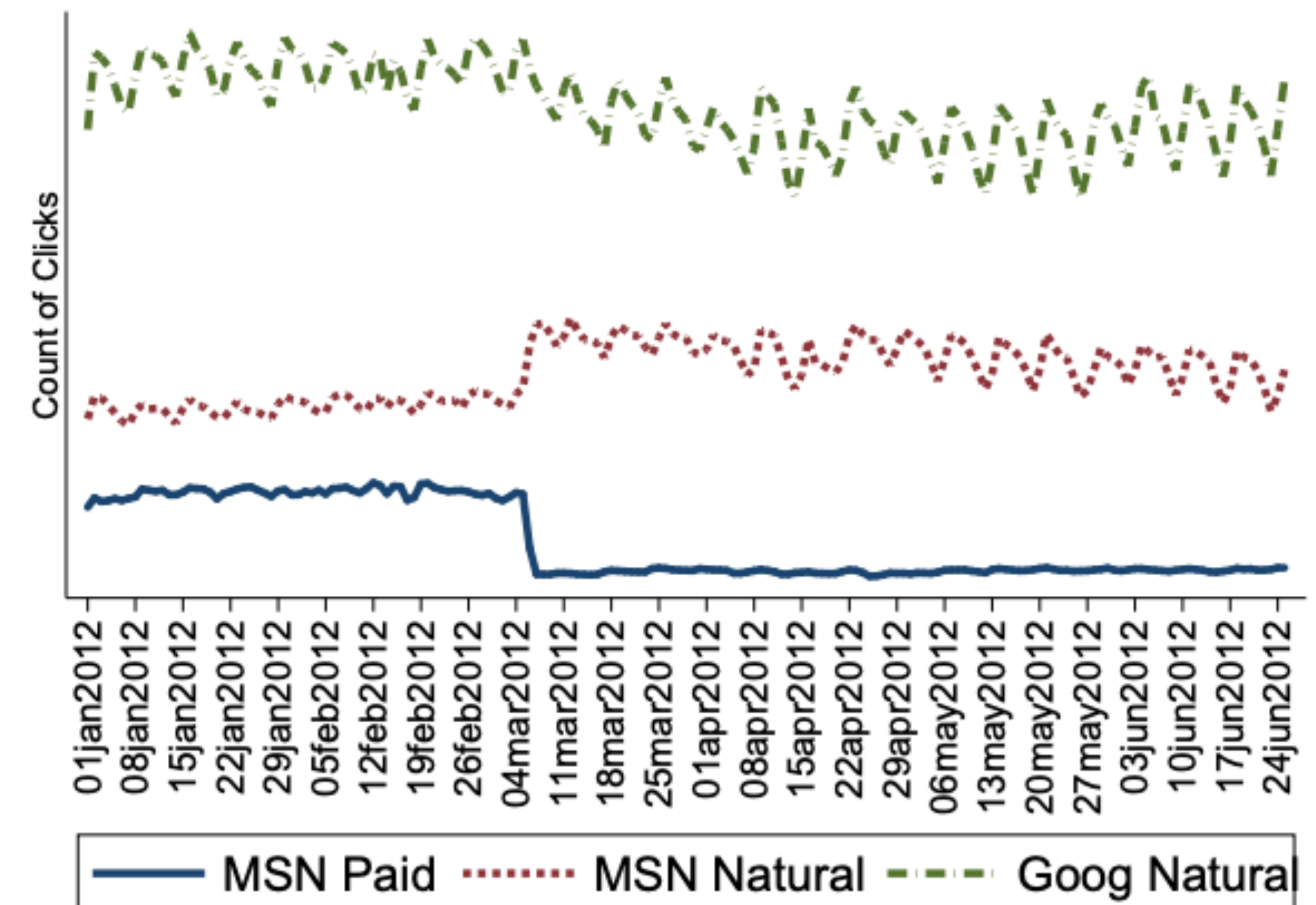
Assumptions for Difference-in-Differences

- Parallel trends: The group that got 'treated' would have had the same trends as the control, absent the treatment.



Parallel Trends in Practice:

- If we have an experiment, then randomization should ensure parallel trends.
- If we don't have an experiment, this is hard to show.
- Having similar trends in the pretreatment period helps.



Some Algebra

$$Y_{it}$$

Total clicks on platform at time t.

$$E[Y_{it}] = \gamma_i + \lambda_t + \delta T_{it}$$

The expected value is function of a time fixed effect, a platform fixed effect, and the treatment T.

If we have two periods (Pre and Post) and use Google as a “control” group:

$$(Y_{MSN,Post} - Y_{Google,Post}) - (Y_{MSN,Pre} - Y_{Google,Pre})$$

$$(\gamma_{MSN} + \lambda_{Post} + \delta - \gamma_{Goog} - \lambda_{Post}) - (\gamma_{MSN} + \lambda_{Pre} - \gamma_{Goog} - \lambda_{Pre}) = \delta$$

δ = **Average Treatment Effect on the Treated**

Simple Example

Log of Clicks

“Treatment Effect”

Common Shocks
(e.g. Seasonality)

$$\log(Clicks_{platform,time}) = \delta * MSN * Post + \beta * MSN + \gamma * Post + Constant + Error$$

MSN Mean
Before Treatment

Constant, mean on Google
before Treatment

Simple Example

Log of Clicks

“Treatment Effect”

Common Shocks
(e.g. Seasonality)

$$\log(Clicks_{p,d}) = \delta * MSN * Post_d + \beta_1 * Google + \beta_2 * Yahoo + \gamma * Post_d + Constant + Error$$

MSN Mean
Before Treatment

Constant, mean on Google
before Treatment

		Log Clicks
Interaction		-0.00529 (0.0177)
Google		5.088 (10.06)
Yahoo		1.375 (5.660)
Constant		11.33* (5.664)
Date FE		Yes
Platform Trends		Yes
N	118	180

Platform Fixed Effects

The eBay experiment

- Ebay bids on over 100 million keywords.
- Experimental design:
 - Test: Randomly pick 30% of DMAs (Designated Market Areas) and turn off all keywords.
 - Why cluster at location and not keyword level?
 - Same user may search multiple keywords leading to spillovers / interference (violation of Stable Unit Value Assumption).

Design Considerations

- Only 210 total DMAs so the number of observations is limited.
- Worry: statistical power.
- They block based on pre-experiment trends in markets to form a treatment and control group.
- Procedure: Divide potential markets by characteristics. Randomize within characteristics. This reduces variance.

Analyze using Difference-in-Differences

- Basic Idea:
- Outcomes in a market are correlated over time. For example, there will be more sales in New York than in Boston.
 - **Control for market fixed effects.**
- Similarly sales across markets are correlated at a given time. For example due to seasonality.
 - **Control for month fixed effects.**
- **In an experiment, this can lead to a reduction in variance.**

Regression of log(sales) on treatment

Before Experiment:
Regress Sales on Ad Spend

Experimental Treatment
Effect

	OLS		DnD
	(1)	(2)	(5)
Estimated Coefficient	0.88500	0.12600	0.00659
(Std Err)	(0.0143)	(0.0404)	(0.0056)
DMA Fixed Effects		Yes	Yes
Date Fixed Effects		Yes	Yes
N	10,500	10,500	23,730

Fixed Effect Controls

Experimental Treatment Effect

Before Experiment:
Regress Sales on Ad Spend

Experimental Treatment
Effect

	OLS		DnD
	(1)	(2)	(5)
Estimated Coefficient (Std Err)	0.88500 (0.0143)	0.12600 (0.0404)	0.00659 (0.0056)
DMA Fixed Effects		Yes	Yes
Date Fixed Effects		Yes	Yes
<i>N</i>	10,500	10,500	23,730

Authors calculated ROI: -63%

This Time

1. Fixed Effects
2. Difference in Difference - with an experiment.
3. **Interpreting logarithms in regressions.**

Logarithms

- They convert a multiplicative scale into an additive scale.
- This makes it possible for us to use linear models to study multiplicative effects.
- That means that log scales look at percentage change.
- Changes the data from 'revenue increased by \$10,000' to 'revenue increased by 8%'

When to use logarithms

- Think about how you expect the independent variable to change the dependent variable.
- Blake, Nosko, and Tadelis analyzed *log* clicks, rather than revenue itself.
 - Where would you expect advertising to have a larger absolute effect? New York City or Rochester?
 - What if, instead of measuring the presence of ads, we were looking at the dollars spent in advertising?
- The stock market, and in particular stock prices are often analyzed using logs.

How to interpret effects on logs.

- To convert the resulting coefficient to a percentage, just put the coefficient to the power of your base
- If the coefficient is 1.2, then increasing the independent variable by 1 increases y by $e^{1.2}$
- For small values, e^x is approximately x , so no conversion is necessary.
- e^{-2} is roughly equal to .2

Total effect of ads is a .66% increase in sales.

Before Experiment:
Regress Sales on Ad Spend

Experimental Treatment
Effect

	OLS		DnD
	(1)	(2)	(5)
Estimated Coefficient (Std Err)	0.88500 (0.0143)	0.12600 (0.0404)	0.00659 (0.0056)
DMA Fixed Effects		Yes	Yes
Date Fixed Effects		Yes	Yes
<i>N</i>	10,500	10,500	23,730

When X and Y are logarithms:

- Suppose $\log(\text{Sales}) = \alpha * \log(\text{Spend})$
- 1% increase in spend results in an $\alpha\%$ increase in sales.

- Here .88% increase.

	(1)
Estimated Coefficient	0.88500
(Std Err)	(0.0143)
DMA Fixed Effects	
Date Fixed Effects	
N	10,500

Practical considerations

- Logarithms are only defined on values > 0 , $\log(0)$ is negative infinity.
 - If you have zeros in your data, simply adding 1 to all records is often a reasonable hack.
 - Omitting the zeros is usually not ok because it results in sampling bias. But can be justified sometimes.
- Logs are often used in data with outliers. Reduces variance.
- Often used when you suspect a diminishing relationship.
- NOT more flexible than a linear relationship, just different

Recap

1. Fixed Effects
2. Difference in Difference - with an experiment.
3. Interpreting logarithms in regressions.