Diff in Diff: Applications

Lachlan Deer

Social Media and Web Analytics, Spring 2024

Learning Goals

1. Interpret Difference in Difference results found in the literature
2. Explain advantages and shortcomings of choices made in existing research designs that leverage natural experiments
3. Interpret analysis in search engine advertising markets and on social media platforms

Where are we now?

In the previous class:

- Difference in differences as a research design to analyse data from natural experiments

This class

- Applications of Difference in Differences research design in digital markets

Today's Topics

1. **Search engine advertising effectiveness**

- Randomised Control Trial, but with imperfect randomization into Treatment and Control
- Shuts down search engine ads by eBay in geographic regions of the US
- Examines impact on sales of eBay products
- Discussion below from Blake, Nosko and Tadelis

2. **Effect of social media on product demand**

- Natural experiment leveraging a shutdown of social media in mainland China but not in Hong Kong * To study how social media impacts TV viewership
- Discussion below from Seiler, Yao and Wang

1/ Search Engine Ad Effectiveness

The Business Problem

Seeking answers to the following strategic questions:

- Are Brand based SEM ads effective at bringing traffic to my site?
- Are non-Brand based SEM ads effective at generating sales?
- Are the effects heterogenous across consumers?
- Are the effects heterogenous across companies?

Are Paid Search Ads Effective?

Motivation: Is Search Engine Marketing Effective?

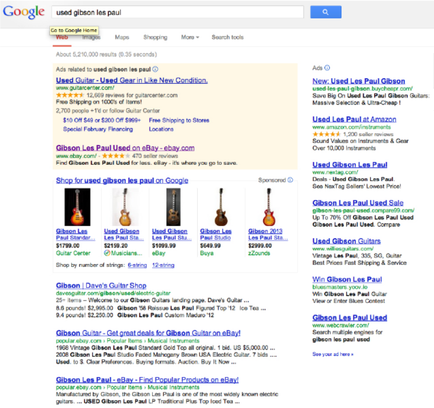
Specific Questions:

- Does SEM generate a positive Return on Investment?
- Is SEM an informative or persuasive form of advertsing?

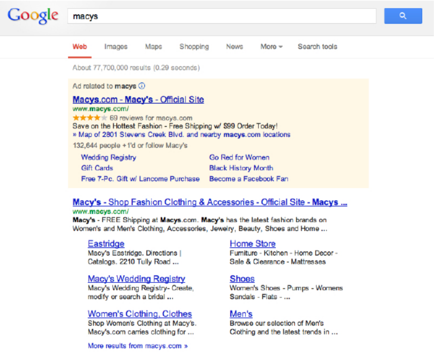
How?

- A series of controlled experiments at eBay
- First, a “proof of concept”
- Then a larger scale experiment

Paid Search in 2012



(a) Used Gibson Les Paul



(b) Macys

Brand Search Terms Experiment

Brand Terms: any queries that include the name of the brand

- Examples: 'ebay shoes', 'de bijenkorf dress',

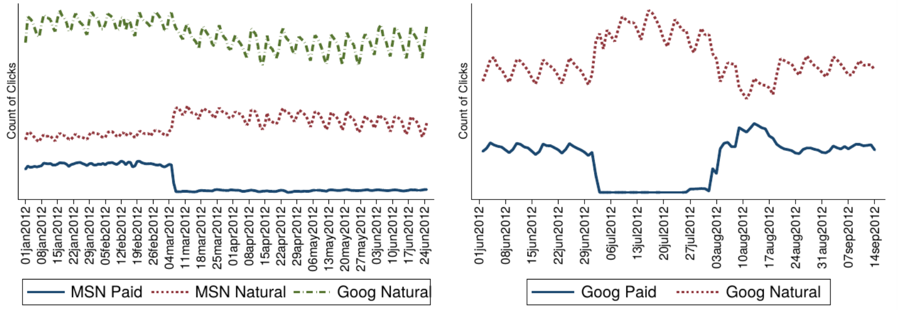
Hypothesis: Users who type the brand name intend to go to that site anyway

$\implies$ brand ads are intercepting what would otherwise be organic clicks

Experiments:

- Experiment 1 (March to June 2012): Shutdown brand ads on MSN and Yahoo!
- Experiment 2 (July 2012): Shutdown brand ads on Google

Brand Search Terms Eyeconometrics



(a) MSN Test (b) Google Test

- 99.5 % of click traffic is retained!

Non-Brand Search Terms

Non-Brand Search terms: queries that do not include the name of the brand

- Examples: ‘shoes’, ‘long dress’

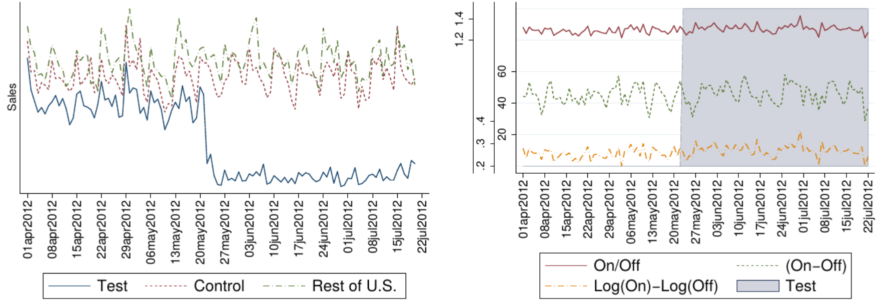
Key difference: Users might not know product is available at a advertiser’s website

Hypothesis: Non brand ads steer consumers to advertiser’s site

Experiment: Large scale Randomized Control Trial

- Suspend non-brand ads in 30% of all DMAs in USA
- Control vs Test Split chosen via an algorithm
- DMA: region of the US, roughly equivalent to a metro area

Non-Brand Search Terms Eyeconometrics



(a) Attributed Sales by Region

(b) Differences in Total Sales

Non-Brand Search Terms Econometrics

Method: Difference in Differences

$$\ln(\text{Sales}_{it}) = \beta_0 + \beta_1 \text{Treatment Group}_i + \beta_2 Post_t \\ + \delta \text{Treatment Group}_i \times Post_t + \text{Fixed Effects} + \varepsilon_{it}$$

- i is a DMA (region) of the US
- t is time (calendar date)

Coefficient of Interest: δ

Non Brand Search Terms Results

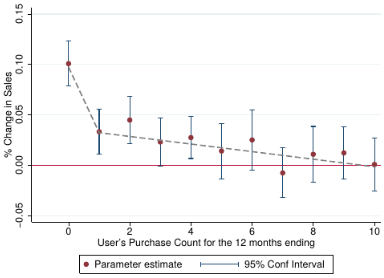
Table 1: Return on Investment

	OLS		IV		DnD	
	(1)	(2)	(3)	(4)	(5)	
Estimated Coefficient	0.88500	0.12600	0.00401	0.00188	0.00659	A
(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	10500	10500	23730	23730	23730	
$\Delta\ln$(Spend) Adjustment	3.51	3.51	3.51	3.51	1	B
$\Delta\ln$(Rev) (β)	3.10635	0.44226	0.01408	0.00660	0.00659	C=A*B
Spend (Millions of $)	$ 51.00	$ 51.00	$ 51.00	$ 51.00	$ 51.00	D
Gross Revenue (R')	2,880.64	2,880.64	2,880.64	2,880.64	2,880.64	E
ROI	4173%	1632%	-22%	-63%	-63%	F=A/(1+A)*(E/D)-1
ROI Lower Bound	4139%	697%	-2168%	-124%	-124%	
ROI Upper Bound	4205%	2265%	1191%	-3%	-3%	

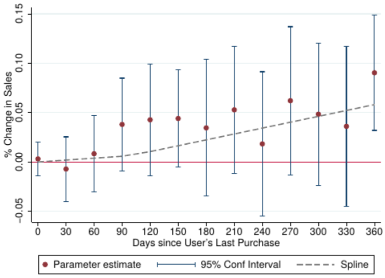
The upper panel presents regression estimates of SEM's effect on sales. Columns (1) and (2) naively regress sales on spending in the pre-experiment period. Columns (3) and (4) show estimates of spending's effect on revenue using the difference-in-differences indicators as excluded instruments. Column (5) shows the reduced form difference-in-differences interaction coefficient. The lower panel translates these estimates into a return on investment (ROI) as discussed in Section 4 and shows its 95% confidence interval.

Consumer Heterogeneity

Figure 4: Paid Search Effect by User Segment



(a) User Frequency



(b) User Recency

Panel (a) shows difference-in-differences estimates and 95% confidence bands of paid search effects on sales for different user segments as defined by how many purchases were made in the previous 12 months. Panel (b) shows similar estimates where users were segmented by the time since last purchase.

Main Takeaways

- Ads served via **Brand Search terms** are, **on average**, **ineffective** at bringing clicks to site
- Ads served via **Non-Brand Search terms** are, **on average**, **do not generate sales**
- **Non-Brand Search terms might be** effective for:
  - Consumers who **do not purchase frequently** on site
  - Consumers who **haven't purchased in a long time**

Results are suggestive of Search Engine Ads being **informative**

Discussion Q:

- Are the consumers for whom ads might be effective usually the type of consumers a firm would advertise to?

Generalizability of Results?

How **generalizable** are the eBay results across different companies?

- Coviello, Gneezy and Goette (2017) run the same experiments for a 'more representative company'
  - Company: Edmunds - a large auto insurer in the US
  - Experiment: Shutdown branded keyword ads on Yahoo and Bing
    - Split markets into 'Treatment' and 'Control'
  - Analysis: Difference in Differences

Generalizability of Results?

Table 2: Difference-in-differences estimates of the treatment effects
WLS Regressions

Dependent variable: change in web-traffic category, normalized by average total web traffic in market during the baseline phase.

Dependent variable:	paid traffic		organic traffic		total traffic	
Treatment Market (=1)	-0.098***	-0.102***	0.042***	0.040***	-0.056***	-0.062***
	(0.008)	(0.003)	(0.012)	(0.011)	(0.017)	(0.012)
Fraction of paid sessions in BL		-0.756***		-0.435*		-1.191***
		(0.100)		(0.255)		(0.321)
Constant	-0.020***	0.092***	-0.077***	-0.012	-0.097***	0.080
	(0.002)	(0.015)	(0.009)	(0.040)	(0.009)	(0.050)
R^2	0.746	0.918	0.163	0.232	0.173	0.473
Obs	210	210	210	210	210	210

Notes: Heteroskedasticity-robust WLS standard errors are in parentheses. Estimates are weighted by the average total web traffic in a market during the baseline (the normalizing variable). *, **, and *** denote significance at the 10, 5, and 1 percent levels, respectively.

Result: **5.6 percentage point reduction in total traffic**

$\Longrightarrow$ search engine ads are not a "zero" effect for all firms

2/ Does Online Word of Mouth Matter?

What is Word of Mouth Marketing?

Consumer's interest in a company's product or service is reflected in their "daily dialogues"

- Why is this *new* in "social media"?
    - It isn't a new idea ...
    - The "social web" with it's increasing connectivity makes it more salient
    - ... and measurable

Types of Word of Mouth

Organic word of mouth:

- People become advocates for a product and have a desire to share their views.
- **This is our focus this week**

Amplified word of mouth:

- Marketers launch campaigns designed to encourage or accelerate WoM in existing or new communities.
- We'll come back to this later in the course – "Social Advertising"

Online versus Offline

- Distinction is always lurking in the background

Social Media Word of Mouth Matters

- Consumers now spend more than 135 mins per day on social media
    - Social media sites contain a treasure-trove of decision relevant information
    - Twitter is the main platform for opinion exchange
- Social Media fostered growing importance of WoM marketing
- Chief Marketing Officers think online WoM matters
    - … Rationalized by consumer's trust in online info from peers (Nielsen, 2013)
    - 64% of marketing executives believe word of mouth is the most effective form of marketing
    - Only 6% say they have mastered it.

Why Word of Mouth Might Matter

Four potential mechanisms at play:

1. Awareness
2. Buzz
3. Social learning
4. Consumption complementarities

Most often we see:

- Awareness & Buzz $\rightarrow$ volume of tweets
- Social learning $\rightarrow$ sentiment in tweet's text
    - Sentiment often called valence

Online WoM & Causality

Motivation: Causal inference is particularly difficult in the realm of online WOM due to the fact that firms are not directly in control of the amount of WOM.

Specific Business Questions:

- What is the demand elasticity of demand wrt volume of posts?
- What is the mechanism through which online WoM influences choice?

How?: Natural experiment – shutdown if Sina Weibo due to political events in mainland China but not HK

- Sina Weibo $\approx$ Chinese Twitter

Empirical Approach

Industry: TV show viewership – soapies

- Not really **new** products

Data:

- TV ratings (i.e. viewership) at episode/city level in mainland China and HK
- Microblogging activity about each show

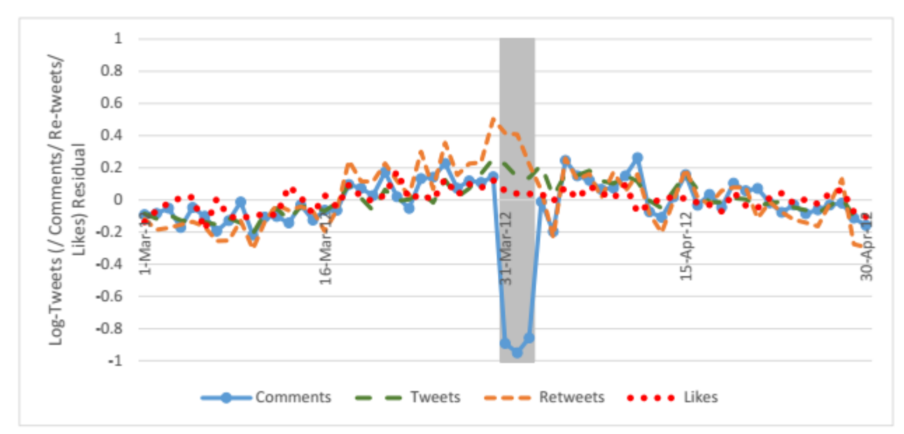
The Natural Experiment: Censorship block on Sina Weibo

- Large, random shock, unrelated to TV
- Block in mainland China, **but not** HK

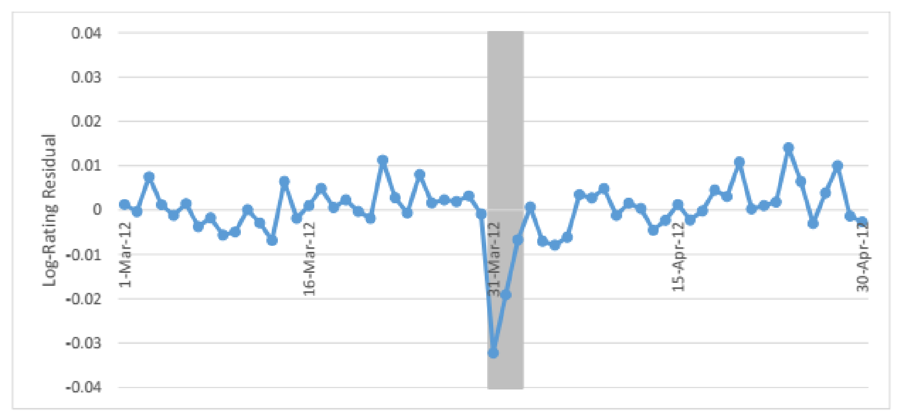
Difference in Differences Regression

$$\begin{aligned} LogRating_{jt} &= \alpha Block_t + \beta Mainland_j + \delta_j Block_t \times Mainland_j \\ &\quad + Weekday_t'\gamma + \varepsilon_{jt} \end{aligned}$$

Graphical Evidence I



Graphical Evidence II



Diff in Diff Results

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log Rating	Log Rating	Log Rating	Log Rating	Log Rating	Log Rating
Sample	Mainland China	HK and Mainland China	HK and Shenzhen (respective shows)	HK and Shenzhen (mainland shows)	24 Cities in Mainl. China	24 Cities in Mainl. China
Censor Dummy	-0.017***	0.005	0.002	-0.008***	-0.010	-0.008
	(0.005)	(0.010)	(0.010)	(0.002)	(0.006)	(0.006)
Mainland × Censor Dummy		-0.026**				
		(0.012)				
Shenzhen × Censor Dummy			-0.035**	-0.017*		
			(0.014)	(0.010)		
Sina Weibo Penetration × Censor Dummy					-0.027*	
					(0.014)	
Above Median Penet. × Censor Dummy						-0.016***
						(0.006)
Show FEs	Yes	Yes	Yes	Yes	Yes	Yes
Weekday Dummies	Yes	Yes	Yes	Yes	Yes	Yes
City FEs	n/a	n/a	n/a	n/a	Yes	Yes
Observations	7,899	11,427	11,427	15,798	189,576	189,576
Shows	193	325	325	193	193	193
R^2	0.881	0.964	0.951	0.774	0.479	0.479

Table 2: **Difference-in-Differences Regressions: Geographical Differences.** The unit of observation is an episode in columns (1) to (4) and an episode/city combination in columns (5) and (6). Standard errors are clustered at the show level.

What is the Mechanism?

Dependent Variable	(1) Log Rating	(2) Log Rating	(3) Log Rating	(4) Log Rating
Censor Dummy	-0.005	-0.001	-0.002	-0.002
	(0.005)	(0.007)	(0.007)	(0.007)
Medium **Daily** Activity	-0.008			
× Censor Dummy	(0.011)			
High **Daily** Activity	-0.026**			
× Censor Dummy	(0.012)			
Medium **Pre-Show** Activity		-0.007	-0.007	-0.007
× Censor Dummy		(0.010)	(0.012)	(0.012)
High **Pre-Show** Activity		0.011	0.024	0.028
× Censor Dummy		(0.020)	(0.019)	(0.019)
Medium **Post-Show** Activity		-0.007	-0.007	-0.008
× Censor Dummy		(0.009)	(0.009)	(0.009)
High **Post-Show** Activity		-0.041**	0.001	0.005
× Censor Dummy		(0.020)	(0.018)	(0.019)
Medium **Post-Show (Any) Sentiment** Comments			0.007	
× Censor Dummy			(0.014)	
High **Post-Show (Any) Sentiment** Comments			-0.060***	
× Censor Dummy			(0.016)	
Medium **Post-Show Positive Sentiment** Comments				0.017
× Censor Dummy				(0.014)
High **Post-Show Positive Sentiment** Comments				-0.039**
× Censor Dummy				(0.017)
Medium **Post-Show Negative Sentiment** Comments				-0.017
× Censor Dummy				(0.014)
High **Post-Show Negative Sentiment** Comments				-0.041**
× Censor Dummy				(0.018)
Show FEs	Yes	Yes	Yes	Yes
Day of the Week Dummies	Yes	Yes	Yes	Yes
Observations	7,899	7,899	7,899	7,899
Shows	193	193	193	193
R^2	0.881	0.881	0.881	0.881

Table 5: **Timing and Content: The Differential Impact of Weibo Activity.** The unit of observation is an episode. Standard errors are clustered at the show level.

Takeaways

- Estimated Volume elasticity: between 0.016 and 0.026
- WoM influnces demand via consumption complementarities
    - Can chat about it later online
- Managerial Implications:
    - Fostering post-show discussion
    - Doesn't appear to be sentiment effects
        - (maybe because quality is known?)
        - Does sentiment matter is a big conversation in the literature

An Alternative Approach?

How far to ”believable” numbers can get get without experimental variation?

Can we reconcile the volume vs sentiment debate?

- If we can **control** for (almost) all the omitted variables
- And impose **structure** on the consumer decision making problem
  - Substitute: Clean variation (experiment) for more mathematical modelling and assumptions
- Studied by Deer, Crawford, Chintagunta (2022)

Setting: US Movie Industry & Twitter WoM

Important Distinction for new products:

- Pre- vs Post- release volume and sentiment

Main Result - Demand Elasticities

Table 1: Demand Responses to Twitter WoM and Advertising

	Own Demand Elasticities		Δs_{jt} w.r.t. to 0.5 Std Devn Change	
	Opening Weekend	Post Opening	Opening Weekend	Post Opening
Volume_{pre}	0.06	-0.04	5.44	-3.73
Volume_{post}	-	0.08	-	8.10
Sentiment_{pre}	-0.02	-0.17	-1.29	-10.41
Sentiment_{post}	-	0.27	-	12.01
Ad Spend_{pre}	0.04	0.18	1.12	4.67
Ad Spend_{post}	-	0.12	-	6.58

Notes: Own demand elasticities are computed using $\eta_{jj} = \frac{\beta_{j,t-r_j}}{(1-\rho)} x_{jt}(1 - \rho \log(s_{j|gt}) - (1-\rho)s_{jt})$ and averaged over movies within relevant time frame. Responses to quarter standard deviations computed from $\Delta s_{jt} = \eta_{jj} \frac{0.5sd(x_{jt})}{\bar{x_{jt}}}$. The magnitude of the elasticities and demand responses differ across the opening weekend and post opening periods due to the different estimated parameters for each of these phases, $\beta_{j,open}$ and $\beta_{j,post}$.

Expected Performance Heterogeneity

	Own Demand Elasticities	
	Opening Weekend	Post Opening
Exp. Performance Tier = Large		
$Volume_{pre}$	0.34	0.19
$Volume_{post}$	-	0.11
$Sentiment_{pre}$	0.09	0.93
$Sentiment_{post}$	-	-0.81
Exp. Performance Tier = Medium		
$Volume_{pre}$	0.01	-0.17
$Volume_{post}$	-	0.23
$Sentiment_{pre}$	0.00	0.14
$Sentiment_{post}$	-	-0.20
Exp. Performance Tier = Small		
$Volume_{pre}$	0.10	0.10
$Volume_{post}$	-	0.07
$Sentiment_{pre}$	-0.06	-0.52
$Sentiment_{post}$	-	0.70

License and Citation

Suggested Citation:

```
@misc{smwa2024_didapps,
      title={"Social Media and Web Analytics: Diff in Diff: Applications
      author={Lachlan Deer},
      year={2024},
      url = "https://tisem-digital-marketing.github.io/2024-smwa"
}
```

This course adheres to the principles of the Open Science Community of Tilburg University. This initiative advocates for transparency and accessibility in research and teaching to all levels of society and thus creating more accountability and impact.