MKT927: INTRO TO QUANTITATIVE MARKETING

Prof. Andrey Fradkin

Lecture 3: Observational Causal Inference and Advertising

OBSERVATIONAL CAUSAL INFERENCE

WHY EXPERIMENTS ARE NOT ENOUGH

- Not always feasible to run experiments or ones that are powerful enough to detect the effect we want to study.
- Want to learn about things that happened without an experiment.
- Equilibrium effects are not directly observable with:
  - Individual level randomization.
  - Short interventions.
  - Lack of data on competitors' actions and outcomes.

THE CREDIBILITY OF OBSERVATIONAL STUDIES

- Whether something is credible or not is a judgement call.
- It is critical to be transparent about the assumptions that justify the credibility of the study.
- Generally, something about the intervention you're studying needs to be arbitrary ('exogenous'). Very few things in the real world are 'random'.
- Requires a lot of 'shoe-leather' to do this well, as in the Shapiro (2018) paper.

OUR EXAMPLE

- An industry with two companies: A and B.
- Company A and B both advertise to different extents across DMAs and over time.
- Company B hired an economist and decided to stop advertising in year X.
- We want to understand the effect of advertising on industry outcomes.
  - Sales of A and B.
  - Total sales.

2X2 DIFFERENCE-IN-DIFFERENCES DESIGN

- DD compares treatment and untreated groups over time (pre & post periods).
- **DD Estimator Formula:**

$$DD = (\bar{Y}_{1,post} - \bar{Y}_{1,pre}) - (\bar{Y}_{0,post} - \bar{Y}_{0,pre})$$

  Where:

  - $Y_{1,post}$ = Mean outcome, treatment group, post-period
  - $Y_{1,pre}$ = Mean outcome, treatment group, pre-period
  - $Y_{0,post}$ = Mean outcome, untreated group, post-period
  - $Y_{0,pre}$ = Mean outcome, untreated group, pre-period
- **Interpretation:** DD estimates the treatment effect by differencing the changes in outcomes between the treated and untreated groups.

THE TWO-WAY FIXED EFFECTS ESTIMATOR

$$Y_{it} = \alpha_i + \lambda_t + \beta \text{Post}_t \times \text{Treatment}_i + \epsilon_{it}$$

- α_i is the firm-specific effect.
- λ_t is the time-specific effect.
- β is the treatment effect.
- Notice: ignores heterogeneity in the treatment effect over time.

2X2 DID DECOMPOSITION

$$\begin{aligned}
\hat{\delta}^{2\times 2}_{kU} &= \underbrace{\left(\mathbb{E}[Y_k|\text{Post}] - \mathbb{E}[Y_k|\text{Pre}]\right) - \left(\mathbb{E}[Y_U|\text{Post}] - \mathbb{E}[Y_U|\text{Pre}]\right)}_{\text{Basic 2x2 DiD}} \\
&= \underbrace{\left(\mathbb{E}[Y^1_k|\text{Post}] - \mathbb{E}[Y^0_k|\text{Pre}]\right) - \left(\mathbb{E}[Y^0_U|\text{Post}] - \mathbb{E}[Y^0_U|\text{Pre}]\right)}_{\text{Switching equation}} \\
&\quad + \underbrace{\mathbb{E}[Y^0_k|\text{Post}] - \mathbb{E}[Y^0_k|\text{Post}]}_{\text{Adding zero}} \\
&= \underbrace{\mathbb{E}[Y^1_k|\text{Post}] - \mathbb{E}[Y^0_k|\text{Post}]}_{\text{ATT}} \\
&\quad + \underbrace{\left[\mathbb{E}[Y^0_k|\text{Post}] - \mathbb{E}[Y^0_k|\text{Pre}]\right] - \left[\mathbb{E}[Y^0_U|\text{Post}] - \mathbb{E}[Y^0_U|\text{Pre}]\right]}_{\text{Non-parallel trends bias in } 2\times 2 \text{ case}}
\end{aligned}$$

PARALLEL TRENDS ASSUMPTION

$$(\mathbb{E}[Y_k^0|\text{Post}] - \mathbb{E}[Y_k^0|\text{Pre}]) = (\mathbb{E}[Y_U^0|\text{Post}] - \mathbb{E}[Y_U^0|\text{Pre}])$$

- This is a very strong assumption!
- Some things to note:
  - This assumption is sensitive to the functional form of the trends. For example, if it holds for Y it probably doesn't hold for $\log Y$.
  - Roth & Sant'Anna (2023) show that parallel trends assumption is not sensitive to monotonic transformations only if treatment is randomly assigned (paraphrasing here).
  - In our example, is parallel trends likely to hold?
  - Parallel trends is not directly testable.
  - Note, this also rules out anticipation effects.

OTHER KEY ISSUES

- Inference
- Staggered Roll-out
- Covariates
- Continuous or multi-valued treatments
- Placebo tests, robustness checks

INFERENCE

- Standard approach is to ‘cluster’ standard errors at the unit of treatment assignment. Requires a large number of clusters. Does not work well when there are two states and two periods, for example. (Why? Cluster level shocks). Some approaches try to model cluster level shocks, but this is rarely done in practice.
- ‘Design-based inference’ is probably the best option. General recommendation is to cluster at the unit of treatment assignment, although this may be conservative.

STAGGERED ROLL-OUT

- When the treatment happens at different times for different units, we need to be careful.
- Large literature shows that the standard (two-way fixed effects) DiD estimator can be biased when treatment effects differ by time of treatment or over time (dynamic treatment effects). Can even be the wrong sign!
- For best practice, read “What’s trending in difference-in-differences? A synthesis of the recent econometrics literature.”
- Will discuss how to deal with this as we talk about papers using staggered roll-out.

COVARIATES

- Maybe parallel trends assumption is only satisfied conditional on covariates.
- Provides motivation for doing a ‘matching’ style DiD. Key requirement is ‘strong overlap’, which means that for a given level of covariate, it is in both the treatment and control groups.
- Controling for covariates by including them in the regression is often done, but can work poorly if there are heterogenous treatment effects or if there is a lack of overlap in the covariates.

DOUBLY ROBUST ESTIMATORS

- We can adjust for covariates in two ways:
  - In the outcome equation (modeling the outcome)
  - In the treatment equation (modeling who gets treated and when).
  - Doubly robust estimators combine these two approaches, so that if either one is correct, the estimator is consistent.
  - For DiD, see Sant'Anna and Zhao (2020), Journal of Econometrics.

MULTI-VALUED TREATMENTS

- Suppose company B was advertising to different extents across DMAs and over time.
- What if we put in a continuous variable in a DiD instead of a binary one?
- Callaway, Goodman-Bacon, and Sant'Anna (2024) discuss this type of DiD in detail. Show that a standard two-way fixed effects estimator can be very wrong and propose solutions.

TESTING FOR PARALLEL TRENDS

- It's an untestable assumption.
- What people do is check for parallel pre-trends and say that parallel trends are likely to hold if there are no pre-trends.
- Need to normalize the pre-trends to a particular period (usually '-1', the period before the treatment).
- But be careful: Newer methods default to different normalization periods (Jonathan Roth has a note on this).

EXAMPLE OF 'MODERN' PRE-TRENDS TESTS

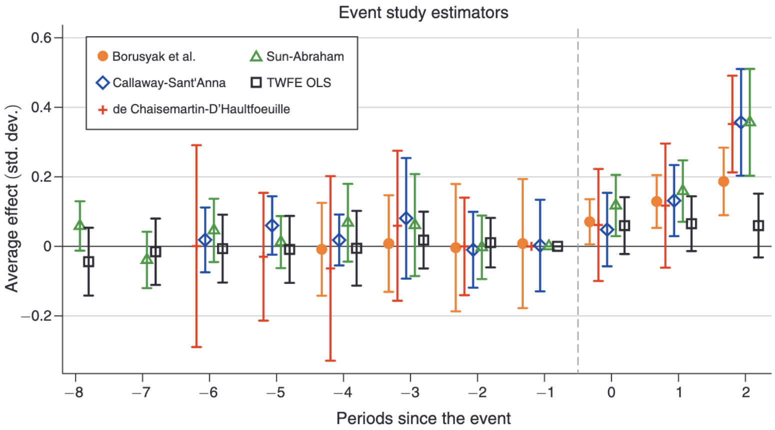
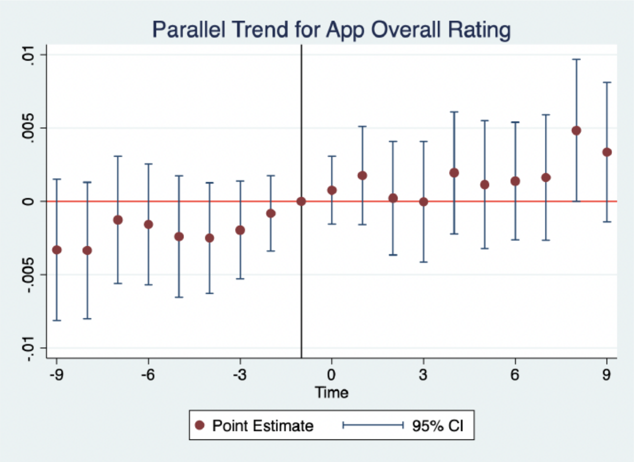


FIGURE 2. EFFECTS OF FACEBOOK ON THE INDEX OF POOR MENTAL HEALTH BASED ON DISTANCE TO/FROM FACEBOOK INTRODUCTION

BUT WHAT IF PRE-TRENDS ARE HIGHLY UNCERTAIN?

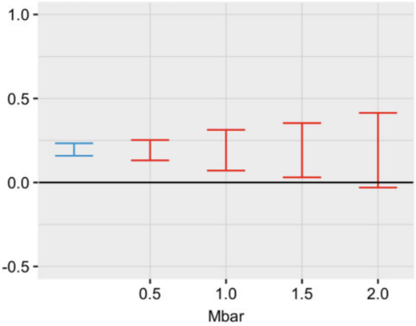


Even though the pre-trends are not statistically significant, they are wide.

Also note that if we draw a sloping line upward, it is consistent with the data.

CHECKING FOR ROBUSTNESS TO VIOLATIONS OF PARALLEL TRENDS

- Rambachan and Roth (2022) propose a test for violations of parallel trends.
- Assume that there is a violation of parallel trends, and then check if the data is consistent with this assumption.
- $\bar{M}$ is the degree of violation allowed.



BACK TO ADVERTISING

SHAPIRO (2018) JPE

- Advertising is something firms do to compete with rivals.
- But advertising can also lead customers to consider the product category, rather than just the product of the advertiser.
- Important issue especially in the context of new products.
- Shapiro (2018) studies this spillover effect of advertising for antidepressants.

BROADER CONTEXT

- Pharmaceutical advertising to consumers is very controversial. (Banned in the EU, highly regulated in the US).
- In certain cases, for example stigmatized conditions such as depression, it may be societally beneficial for more people to take the drug.
- TV advertising is particularly important for this category.

TV ADVERTISING IS NOT RANDOM

- So how do we study the effect of TV advertising on sales?
- Need to find something ‘arbitrary’.
- Television market borders may be arbitrary. Two people living on one side or another may see different ads but would otherwise have similar preferences.

DATA

- Prescription data: 5% random sample of physicians (IMS Health)
- Advertising data: National and DMA-level TV advertising (Kantar Media)
- Geography: 210 Designated Market Areas (DMAs)
- Time period: 1997-2004
- Key advertisers: Prozac, Paxil, Zoloft

SHOWING THE VARIATION

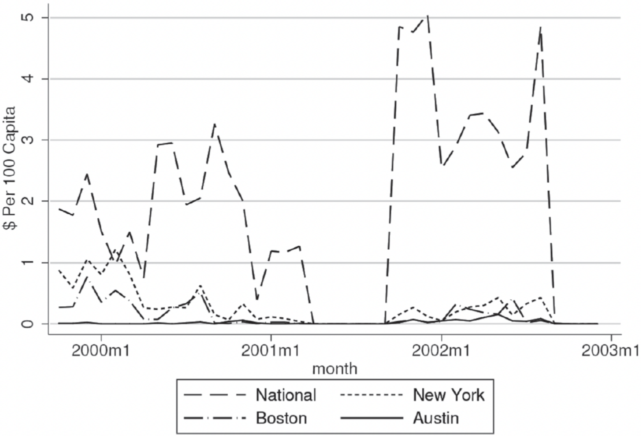


FIG. 3.—Variation across three markets in advertising

SHOWING THE VARIATION

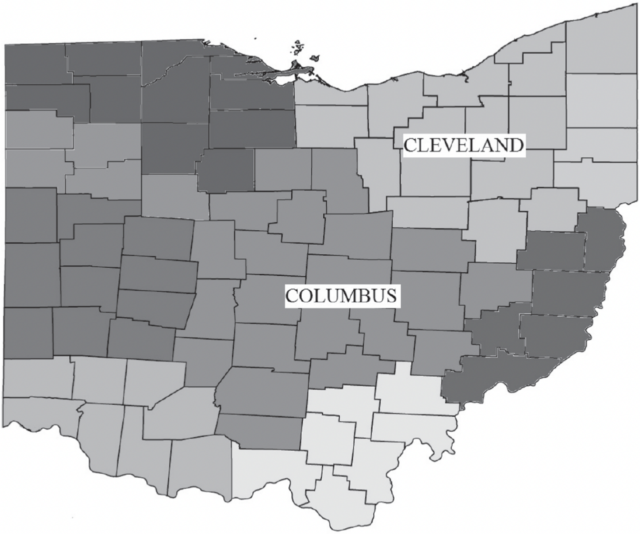


Fig. 5.—Ohio and its DMAs

IDENTIFICATION STRATEGY

- Border discontinuity design
- Compares counties on opposite sides of DMA borders
  - See different ads
  - Otherwise similar characteristics
- Controls for:
  - Market fixed effects
  - Time fixed effects (month)
  - Border-specific trends

REGRESSION MODEL

$$\log(Q_{jmt}) = \lambda \log(Q_{jm,t-1}) + \gamma_1 a_{jmt}^{own} + \gamma_2 a_{jmt}^{cross} + \gamma_3 (a_{jmt}^{own})^2 + \gamma_4 (a_{jmt}^{cross})^2 + \gamma_5 a_{jmt}^{own} a_{jmt}^{cross} + \varepsilon_{jmt}$$

- Quantity of product j in time t for market m as a function of own and rival advertising.
- Problem: advertising is endogenous.

PICKING OUT THE GOOD VARIATION

- Identify 153 county pairs on opposite sides of DMA borders.
- Observation: Product-border-DMA-month.
- Border-time fixed effects make this a ‘Diff-in-Diff’ setup.

ACTUAL SPECIFICATION

$$\log(Q_{jbmt}) = \lambda \log(Q_{jbm,t-1}) + g(a_{jmt}) + \alpha_{jbq} + \alpha_{jbm} + \varepsilon_{jbmt}$$

- **Components:**
  - j, b, m, t: product, border, market, time indices
  - α_{jbq}: quarter fixed effects (time control)
  - α_{jbm}: treatment group fixed effects
  - $g(a_{jmt})$: function of advertising (treatment). Note continuous.
  - $\lambda \log(Q_{jbm,t-1})$: persistence in demand. Note, this is not standard in DiD.

VERY LONG DISCUSSION ABOUT WHY BORDERING COUNTIES ARE GOOD FOR THIS TYPE OF COMPARISON

- Weather
- Medical conferences
- Advertising as a function of prior-DMA outcomes.
- Contrast to many business DiD papers, which have no discussion of specific threats to identification and whether they are likely to be a problem.
- Enough variation even conditional on fixed effects.

KEY TABLE

TABLE 2
THE EFFECT OF OWN AND RIVAL ADVERTISEMENTS ON SALES

Variable	Log(Q)
Lagged log(Q)	.334***
	(.00746)
DTC	.0240***
	(.00621)
DTC^2	−.00216*
	(.00113)
$\text{DTC}_{\text{rival}}$	.0164***
	(.00266)
$\text{DTC}^2_{\text{rival}}$	−.000938***
	(.000252)
$\text{DTC} \times \text{DTC}_{\text{rival}}$	−.00134**
	(.000631)
Product-border-time	Yes
Product-border-DMA	Yes
Observations	316,428
R^2	.955

NOTE.—Product-DMA clustered standard errors are in parentheses.

* $p < .1$

OTHER PARTS OF THE PAPER

- Demand model as a function of advertising.
- Advertising elasticity.
- Supply model → think about how advertising would be different if firms cooperated.
- Motivation for industry-wide advertising (e.g. “Got Milk?”).

SHAPIRO, HITCH, AND TUCHMAN (2021) ECMA

MOTIVATION

- Is TV advertising effective?
- How does effectiveness vary by product?
- What is the ROI?

POWERFUL IDEA IN SOCIAL SCIENCE: META-ANALYSIS

- Estimate TV ad effects for 288 brands.
- Do it using several identification strategies (including borders of Shapiro (2018) and others), find robust results.
- Negative ROI for 80% of brands.
- Interesting and surprising results.

INTERACTIVE WEB APP

https://advertising-effects.chicagobooth.edu/

SUMMARIZING THE ADVERTISING LITERATURE

- Measuring effects precisely is hard.
- When are able to do it, we often find negative ROI.
- Why do firms make such bad decisions? ‘Agency issues’, bad data analysts, something else?
- Surely advertising does have positive ROI sometimes, can we say something productive about when and how it works?

NEXT TIME

- Estimating demand.
- Discuss homework.
- Pages 1 - 35 of Dube Handbook Chapter.
- Conlon, Christopher, and Julie Holland Mortimer. "Empirical properties of diversion ratios".
- Moshary, Sarah, Bradley Shapiro, and Sara Drango. Preferences for firearms and their implications for regulation. No. w30934. National Bureau of Economic Research, 2023.