MKT927: INTRO TO QUANTITATIVE MARKETING

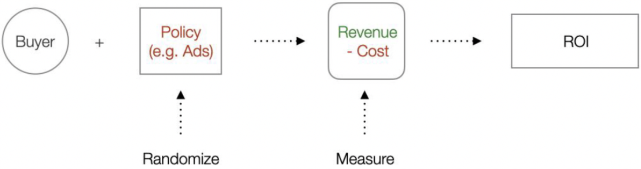
Prof. Andrey Fradkin

Lecture 9: Experiments on Platforms

TODAY'S AGENDA

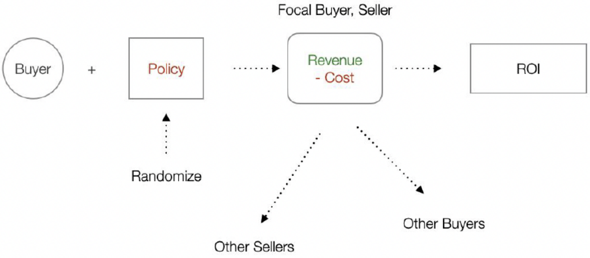
- Types of experiments on platforms, advantages and disadvantages.
- Holtz et al. (2024)
- Rajkumar et al. (2022), (For next time).

A/B TESTS IN SIMPLE ENVIRONMENTS



- A/B tests work well in simple environments.
- These environments have no spillovers.

SPILLOVERS IN MARKETPLACES



- In marketplaces, when two parties transact, they affect others.
- Violation of SUTVA.
- A lot of the literature thinks about dealing with various types of spillovers.

TYPES OF SUTVA VIOLATIONS

- **Capacity constraints:**
  - Treated buyers use a different algorithm.
  - Algorithms redirect demand from some sellers to others.
  - Sellers with more demand may lack the capacity to handle it.

TYPES OF SUTVA VIOLATIONS: INFORMATIONAL SPILLOVERS

- **Informational spillovers:**
  - Information from one user's actions affects others.
  - Example: online reviews.
  - This impacts future buyers and algorithms.

TYPES OF SUTVA VIOLATIONS: EQUILIBRIUM ADJUSTMENTS

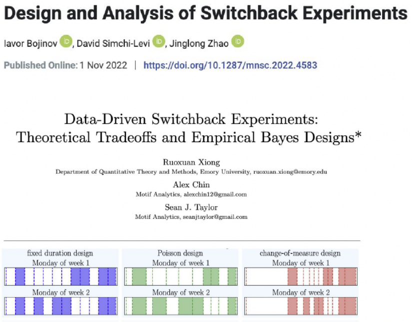
- **Equilibrium adjustments:**
  - Sellers change prices, bidding, and listing information.
  - Buyers change platform usage.
  - Examples: number of applications sent, use of the platform.

WHICH TREATMENTS TO TRY?

- **Azevedo et al. (2020, 2023)**
  - Argue that companies often try very small tweaks, so that tweaks have tiny effects. It requires huge sample sizes to detect these tiny effects, even if the cost of making the wrong decision is low.
  - Show that companies should run experiments that make large changes, and for these changes, statistical power is less of an issue.

SWITCHBACK EXPERIMENTS

- Basic idea: turn the treatment on and off periodically within a market/cluster.
- Works well when spillovers are contained within the time unit.
- Very frequently used by ride-sharing platforms.



THE PROBLEM OF STATISTICAL POWER

- Experiments are pointless if they don't yield informative estimates.
- Power calculations are critical ex-ante.
- Methods for power calculations:
  - Simulating data generating process.
  - Using historical experimental estimates.
  - Use exposure and counterfactual policy logging.
  - See Johnson's Inferno paper for some ideas here.

GETTING AWAY WITHOUT CLUSTER RANDOMIZATION

- The ultimate goal is to make the right decision.
- Under certain conditions, the sign of the global average treatment effect is the same as the sign of the individual average treatment effect.
- If these conditions hold, simpler designs may suffice and offer better statistical power.

When Does Interference Matter?
Decision-Making in Platform Experiments

Ramesh Johari[1], Hannah Li[2], Anushka Murthy[1], and Gabriel Y. Weintraub[3]

[1]Management Science and Engineering, Stanford University
[2]Columbia Business School
[3]Stanford Graduate School of Business

October 11, 2024

Home > Management Science > Vol. 68, No. 10 >

Experimental Design in Two-Sided Platforms: An Analysis of Bias

Ramesh Johari, Hannah Li, Inessa Liskovich, Gabriel Y. Weintraub

Published Online: 25 Jan 2022 | https://doi.org/10.1287/mnsc.2021.4247

SYNTHETIC CONTROL EXPERIMENTS

- Use when experiments are costly or risky.
- Example: TV advertising campaign.
- Careful selection of treated and control units ex-ante can be very beneficial.

Synthetic Controls for Experimental Design

Alberto Abadie — MIT
Jinglong Zhao — Boston University

September 2024

Abstract

This article studies experimental design in settings where the experimental units are large aggregate entities (e.g., markets), and only one or a small number of units can be exposed to the treatment. In such settings, randomization of the treatment may result in treated and control groups with very different characteristics at baseline, inducing biases. We propose a variety of experimental non-randomized synthetic control designs (Abadie, Diamond and Hainmueller, 2010, Abadie and Gardeazabal, 2003) that select the units to be treated, as well as the untreated units to be used as a control group. Average potential outcomes are estimated as weighted averages of the outcomes of treated units for potential outcomes with treatment, and weighted averages the outcomes of control units for potential outcomes without treatment. We analyze the properties of estimators based on synthetic control designs and propose new inferential techniques. We show that in experimental settings with aggregate units, synthetic control designs can substantially reduce estimation biases in comparison to randomization of the treatment.

BUDGET-SPLIT EXPERIMENTS

- Idea, create two mini-marketplaces, splitting the budget between them.
- Useful for understanding short-run marketplace dynamics.
- Less useful for understanding spillovers over time.
- Reason: One side is in both treated and control groups.

Trustworthy Online Marketplace Experimentation with Budget-split Design

Min Liu mliu@linkedin.com
Jialiang Mao jimao@linkedin.com
Kang Kang kakang@linkedin.com

LinkedIn Corporation, Sunnyvale, CA

December 17, 2020

TAKEAWAYS

- No single experimental design answers all important questions.
- It is important to pick the appropriate design for the question at hand.
- One of the key ways academia influences practice.

HOLTZ ET AL. (2024)

- Paper: “Platform Pricing with Spillovers: Evidence from Airbnb”
- Key Question: How do we estimate price elasticities in two-sided markets?
- Setting: Airbnb’s service fees

THE (FEE) PRICING PROBLEM

- Prices affect everything in a market:
  - Who participates.
  - Who matches with whom.
  - Buyer search effort.
  - Seller quality investment.
  - How sellers set prices.
  - Optimal pricing depends on network effects and competing platforms.

WHAT WE LEARN FROM A/B TESTS

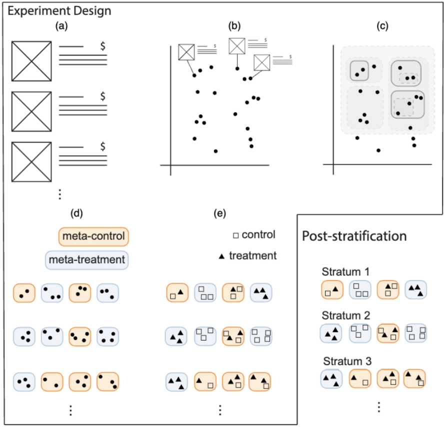
- Assign half the listings one fee, and the other half a different fee.
- Compare bookings (revenue, profits) between treated and control listings.
- We want market-level price elasticity of demand.
- But SUTVA violations occur.

HOLTZ ET AL. (2024) IS AN EXAMINATION OF THIS BIAS.

- Step 1: Need to create clusters of listings.
- Step 2: Need some part of the data to be listing level randomization.
- Step 3: Need another part to be cluster level randomization.
- “Meta-experiment”

VISUALIZING THE STRATEGY

Figure 4. (Color online) The Experiment Design Process



Notes. In panel (a), we use listing-level co-occurrence in search in order to learn "demand embeddings" (panel (b)). A hierarchical clustering algorithm is then applied to those embeddings in order to generate clusters (panel (c)). Clusters are randomly assigned to meta-treatment or meta-control (panel (d)); within meta-control, treatment is assigned at the individual-listing level, whereas in meta-treatment, treatment is assigned at the cluster level (panel (e)). We arrange clusters into strata after treatment assignment to facilitate post-stratification (Miratrix et al. 2013).

Figure 5. (Color online) These Maps Illustrate Clusters Generated Using the Hierarchical Clustering Scheme Described in This Paper

PRACTICAL DETAILS

- Airbnb was very hesitant to do these experiments and to publish them.
- Lost of details missing from the results.
- Very short-run experiment (5 days!).

BALANCE CHECK

Table 1. Confirming Balance Between Conditions

	Individual-level randomized			Cluster-randomized			Meta-experiment		
	Control	Treatment	*p*-value	Control	Treatment	*p*-value	Meta-control	Meta-treatment	*p*-value
Pretreatment statistics									
Bookings	11.864	11.882	0.78	11.760	11.572	0.49	11.790	11.666	0.65
	(26.275)	(26.174)		(10.559)	(10.256)		(10.664)	(10.408)	
Nights Booked	44.984	44.953	0.90	43.288	42.497	0.37	43.195	42.893	0.73
	(101.570)	(102.677)		(34.339)	(33.646)		(34.517)	(33.994)	
Gross Guest Spend	5,920.370	5,934.694	0.72	5,554.392	5,399.833	0.37	5,587.642	5,477.087	0.53
	(15,751.420)	(15,824.250)		(6,764.090)	(6,412.172)		(6,953.921)	(6,590.321)	
$N_{\text{individuals}}$	323,734	323,643							
N_{clusters}				2,979	2,981		1,987	5,960	

Notes. This table tests for statistically significant differences in pretreatment outcomes between treatment and control in the individual-level randomized meta-treatment arm, treatment and control in the cluster-randomized meta-treatment arm, and meta-treatment and meta-control. Each comparison uses a two-sided *t*-test. Analysis is conducted at the individual level within the meta-control arm and at the cluster level within the meta-treatment arm and when comparing the two meta-treatment arms.

HOLTZ META-REGRESSION

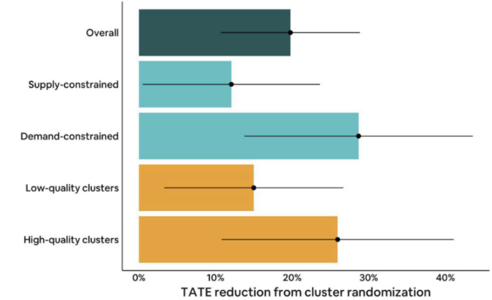
Table 3. Summary of the Meta-experiment Results for the Number of Bookings

	Dependent variable: *Bookings*				
	Overall (1)	Supply constrained (2)	Demand constrained (3)	Low-quality clusters (4)	High-quality clusters (5)
Treatment	−0.277***	−0.433***	−0.140***	−0.360***	−0.196***
	(0.012)	(0.022)	(0.011)	(0.019)	(0.016)
Individual-level Randomized	0.021	0.019	0.013	0.021	0.015
	(0.014)	(0.025)	(0.014)	(0.022)	(0.018)
Individual-level Randomized × Treatment	−0.068***	−0.059*	−0.056***	−0.063**	−0.069***
	(0.018)	(0.031)	(0.018)	(0.027)	(0.023)
Pretreatment bookings	0.175***	0.174***	0.175***	0.172***	0.178***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Pretreatment nights booked	−0.003***	−0.003***	−0.003***	−0.003***	−0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pretreatment gross guest spend	−0.000***	−0.000***	−0.000***	−0.000***	−0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pretreatment nights available	0.001***	0.003***	0.000***	0.002***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Pretreatment searches/night	0.050**	0.021**	0.775***	0.203***	0.028**
	(0.020)	(0.010)	(0.062)	(0.024)	(0.013)
Interference bias estimate, %	19.76	12.05	28.65	14.98	25.92
	(±9.06)	(±11.55)	(±14.91)	(±11.69)	(±15.14)
Stratum F.E.	Yes	Yes	Yes	Yes	Yes
Robust s.e.	Yes	Yes	Yes	Yes	Yes
Semiclustered s.e.	Yes	Yes	Yes	Yes	Yes
R^2	0.405	0.404	0.365	0.408	0.402
Adjusted R^2	0.405	0.404	0.364	0.407	0.402

Notes. Column (1) presents the overall results. Columns (2) and (3) explore heterogeneity with respect to supply/demand constrainedness. Columns (4) and (5) explore heterogeneity with respect to cluster quality. F.E., fixed effect; s.e., standard error.
*$p < 0.1$; **$p < 0.05$; ***$p < 0.01$.

HOLTZ HETEROGENEITY ANALYSIS

Figure 7. (Color online) Reduction in Bias from Cluster Randomization



Note. This graph visualizes the reduction in interference bias from cluster randomization that we estimate across different samples: overall, listings in supply-constrained geographies, listings in demand-constrained geographies, listings in geographies with low-quality clusters, and listings in high-quality clusters.

DISCUSSION

- Does this experiment tell us the treatment effect of interest to Airbnb?
- What would you do differently if you had no constraints in running the experiment?
- What would you do differently if you had no constraints in conducting the analysis?

NEXT TIME

- Brief discussion of Rajkumar et al. (2022).
- Assignment 3 is due.
- Read “Vertical Integration and Consumer Choice: Evidence from a Field Experiment.” (Will post on BU Learn).
- Two Discussions: Please read corresponding papers (Levy 2021, Braghieri et al. 2022).
- End of class time, 5 - 10 minute meetings with students I haven’t spoken to yet about the final project.