Causal Impact of Campaign Strategies Analysis

Xiaoxue Chen and Anqi Sun

October 2024

1 Average Causal Effect (ATE)

Assumptions:
1) Consistency: $D_i = d \rightarrow Y_i(d) = Y_i$
2) Unfoundness: given X, if $\{Y_i(1), Y_i(0)\} \perp D_i|X_i$, D_i is conditionally ignorable.

In our question, $D_i = 1$: have the specific VSA; $D_i = 0$: don't have the specific VSA

$$ATE = E[E[Y_i|D_i = 1, X_i]] - E[E[Y_i|D_i = 0, X_i]]$$

$$A\hat{T}E = \frac{1}{n}\sum_{i=1}^{n}[\hat{E}[Y_i|D_i = 1, X_i] - \hat{E}[Y_i|D_i = 0, X_i]]$$

Consider $E[Y_i|D_i, X_i] = \tau D_i + X_i^\top \beta$, then $ATE = \tau \rightarrow \tau$ has a causal interpretation.

Now we need to do linear regression to fit $E[Y_i|D_i = 1, X_i]$

In this setting, Assumption 2 is relatively difficult to fulfill, because Regression is more dependent on having a comprehensive enough set of covariates to satisfy "conditional ignorability".

Randomized Controlled Trails (RCT)
RCT is a well-known method to calculate causal effect. In our settings, we cannot randomize D_i, and it's also hard to determine which variable should we include in regression so that conditional ignorability holds; what will happen when X_i is assumed to be linear in model but the CEF is not linear on X_i?

However, because our data collected before we change the campaign strategy in different age groups (16-40; 40+) are almost parallel, we consider to use difference in differences to analysis the causal effect.

2 Difference-in Differences

Assumptions:
1) SUTVA: Stable Unit Treatment Value Assumption $\rightarrow$ the potential outcomes of unit i are not relevant to the treatment of other units $\rightarrow$ we should choose the units of analysis to minimize interfernce across units.
2) No-Anticipation Assumption: $Y_{i,t}(g) = Y_{i,t}(\infty)$ for all $t < g$ $\rightarrow$ Before the start of treatment, potential outcomes for units were consistent with those that did not receive treatment
(Selection Bias might be introduced here)
3) Parallel Trends Assumption: $E[Y_{i,t=2}(\infty)|G_i = 2] - E[Y_{i,t=1}(\infty)|G_i = 2] = E[Y_{i,t=2}(\infty)|G_i = \infty] - E[Y_{i,t=1}(\infty)|G_i = \infty]$ $\rightarrow$ In the absence of treatment, the evolution of the outcome among treated units is the same as the evolution among the untreated units on average.

Under 1) 2) 3) assumptions,

$$ATT = (E[Y_{i,t=2}|G_i = 2] - E[Y_{i,t=1}|G_i = 2]) - (E[Y_{i,t=2}|G_i = \infty] - E[Y_{i,t=1}|G_i = \infty])$$

ATT represents the average treatment effect on the treated group at time t=2.

We have two methods to estimate ATT in population,
1) plug-in estimator: $\hat{\theta}^{DID} = (\overline{Y}_{g=2,t=2} - \overline{Y}_{g=2,t=1}) - (\overline{Y}_{g=\infty,t=2} - \overline{Y}_{g=\infty,t=1})$
2) two-way fixed-effects (TWFE) regression: $Y_{i,t} = \alpha_0 + \gamma_0 \mathbf{1}\{G_i = 2\} + \lambda_0 \mathbf{1}\{T_i = 2\} + \beta_0^{twfe}(\mathbf{1}\{G_i = 2\}\mathbf{1}\{T_i = 2\}) + \epsilon_{i,t}$, where β_0^{twfe} is the estimate of $\hat{\theta}^{DID}$

In our setting, we use Repeated Cross-section data $\{Y_i, G_i, T_i\}_{i=1}^n$ i.i.d. draw from the mixture distribution:

$$P(Y \leq y, G = g, T = t) = \mathbf{1}\{t = 2\}\lambda P(Y_{t=2} \leq y, G = g|T = 2) + \mathbf{1}\{t = 1\}(1 - \lambda)P(Y_{t=1} \leq y, G = g|T = 1)$$

which means select the pre-intervention time period ($T = 1$) with probability $(1 - \lambda)$ Select the post-intervention time period ($T = 2$) with probability λ. Collect data for $G = 1, G = 2$, respectively. (Mixed distribution formula lies help explain the data generation process: How the variations in time T and group G are combined and unified into a single distribution.)

3 Case: Product click-through rate on Tiktok

The dataset we are analyzing comes from TikTok ad performance logs, primarily capturing data from September and October. This dataset includes detailed information on various campaigns, such as impressions, clicks, click-through rates (CTR), and conversions. In early October, we changed the influencer for a major campaign A on age group 1(16-40), significantly increasing its exposure and impression volume, while other campaigns remained unchanged.

In our dataset, we have

- Time-related variables
  Year/Month: records the time when the ads were placed and is used to divide the pre-intervention (September) and post-intervention (October) time periods.
  By Day: A finer time granularity that records daily data on ad placements for calculating daily average metrics or observing daily trend changes.
- Variables related to advertising campaigns
  Campaign name: the name of the campaign, used to label each Campaign and help us distinguish between which Campaign was intervened on (the treatment group) and which Campaigns stayed the same (the control group).
  Ad group name: the name of the ad group, a subset of the Campaign that may need to be further disaggregated to analyze the effectiveness of specific ad tiers.
  Age Group: the age group of the audience (16-40, 40+), controlling for differences in advertising preferences among different age groups.
- Outcome variable
  CTR (destination): the click-through rate, the main variable analyzed for targeting, calculated as Clicks (destination) / Impression.
- Covariate
  Cost, CPC, CPM, Video views...

Using the difference-in-differences (DID) method, we aim to study the impact of this intervention on CTR by comparing the changes in CTR between September (pre-intervention) and October (post-intervention) for the affected ad group (treatment group) and the unaffected ad groups (control group). This will help us assess the causal effect of reducing exposure on ad click performance.

Check Assumptions

1) SUTVA: Our unit of analysis is each individual user who viewed the ads.
2) No-Anticipation Assumption: Before the intervention (October), the click-through behavior of users in the treatment group ($G = 1$) is consistent with that of users in the control group ($G = \infty$).
3) Parallel Trends Assumption: If no intervention had occurred, the click-through rate (CTR) trends for users in the treatment group ($G = 1$) and control group ($G = \infty$) would have followed the same trajectory between September (pre-treatment) and October (post-treatment).

Definition of Variables

$Y_{i,t}$: For a given user i who viewed influencer1 of our major campaign, $Y_{i,t}$ is a binary indicator of whether the user clicked on the ad or not (1 for a click, 0 for no click).
G_i: treatment group ($G_i = 2$): Users who in age group 1 (16-40); Control Group ($G_i = \infty$): Users who in age group 1 (40+)
T: $T = 1$: Pre-treatment period, September; $T = 2$: Post-treatment period, October.

Data Collection

Between September and October, we conducted 20 cross-sectional samplings across the two periods. Each cross-section represents a random sample of users who interacted with the ads during a specific period (either September or October). At each sampling point: A subset of users was randomly selected from the population who interacted with the campaigns. At each sampling point: A subset of users was randomly selected from the population who interacted with the campaigns. For each selected user, their ad interaction data (e.g. click, date, Video views, Campaign name, etc.)and demographic data (gender, age group, region, Device Type, language, etc.) were recorded. After data cleaning, we got

User ID	**Time (T)**	**Group (G)**	**Campaign Name**	**...**	**Click**
User001	1 (September)	2 (age group 1)	PartyHighlightPinkInfluencer1	...	0
User002	1 (September)	∞ (age group 2)	PartyHighlightPinkInfluencer3	...	1
User003	2 (October)	2 (age group 1)	PartyHighlightPinkInfluencer1	...	1
User004	2 (October)	∞ (age group 2)	PartyHighlightPinkInfluencer3	...	0
...	...	...	...	...	...

Table 1: Organized Data Structure