ROBUST CAUSAL INFERENCE FOR INCREMENTAL RETURN ON AD SPEND WITH RANDOMIZED GEO EXPERIMENTS

Aiyou Chen
Tim Au

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Chen, Aiyou and Au, Tim, "ROBUST CAUSAL INFERENCE FOR INCREMENTAL RETURN ON AD SPEND WITH RANDOMIZED GEO EXPERIMENTS", Technical Disclosure Commons, (August 05, 2019)
https://www.tdcommons.org/dpubs_series/2381
ROBUST CAUSAL INFERENCE FOR INCREMENTAL RETURN ON AD SPEND WITH RANDOMIZED GEO EXPERIMENTS

BACKGROUND

For the better part of the last decade, the annual growth rate of online advertising has been significantly outpacing that of all other advertising media. The success of online advertising has been attributed to its measurability and targetability. Incremental return on ad spend (iROAS) of a progressive marketing strategy—that is, the ratio of the strategy’s causal effect on some response metric of interest (e.g., incremental sales caused by advertising on new keywords) relative to its causal effect on the ad spend (incremental ad spend required to advertise on the new keywords)—has become progressively more important as advertisers increasingly seek to better understand the impact of their marketing decisions.

To measure the impact of their marketing campaigns, advertisers frequently employ randomized paired “geo experiment” designs which partition a geographic region of interest into a set of smaller non-overlapping “geos” that are regarded as the units of experimentation rather than the individual users themselves. Indeed, since their introduction, geo experiments have gone on to become a standard tool for the causal measurement of online advertising. However, geo experiments also introduce some additional complexity which makes the iROAS estimation problem difficult. Often only a small number of heterogeneous experimental units are available for experimentation, which makes it challenging to obtain reliable estimates of the iROAS with existing methods.
SUMMARY

This publication considers randomized paired geo designs, wherein there are \( n \) pairs of geos, and within each pair, one of them is randomly selected as the treatment geo and the other as the control. The main contributions of this publication are, 1) the formulation of a novel statistical framework for inferring the population iROAS, 2) the development of a robust model-free estimator called Trimmed Match, which adaptively trims poorly matched pairs. Randomized paired geo designs rely on pairs that match very closely. To achieve a more accurate estimate of iROAS, the Trimmed Match approach trims automatically removes poorly matched pairs from the set of paired geos based on a fixed trim rate. The following analysis describes a statistical framework for inferring the population iROAS.

Let \( G \) be the set of geos for a target population. Given a geo \( g \in G \), let \((S_g, R_g) \in \mathbb{R}^2\) denote its observed bivariate outcome, where \( S_g \) is ad spend and \( R_g \) is the response variable. We denote geo \( g \)’s potential outcome under the control and treatment ad serving conditions as \((S_g^{(C)}, R_g^{(C)})\) and \((S_g^{(T)}, R_g^{(T)})\) respectively, where we can only ever observe one of these two bivariate potential outcomes for each geo \( g \). For each geo \( g \), there are two unit-level causal effects caused by the new marketing strategy: incremental ad spend and incremental response, which are defined by \( S_g^{(T)} - S_g^{(C)} \) and \( R_g^{(T)} - R_g^{(C)} \) respectively. The incremental return on ad spend (iROAS) w.r.t. geo \( g \), denoted as \( \theta_g \), is the ratio of incremental response to incremental ad spend:

\[
\theta_g = \frac{R_g^{(T)} - R_g^{(C)}}{S_g^{(T)} - S_g^{(C)}}.
\]
and the iROAS w.r.t. the population $G$ is defined similarly:

\[
\theta^* = \frac{\frac{1}{|T|} \sum_{g \in G} R_g^{(T)} - R_g^{(C)}}{\frac{1}{|T|} \sum_{g \in G} S_g^{(T)} - S_g^{(C)}}.
\]

Advertisers frequently find $\theta^*$ to be a more informative causal estimate of advertising performance, which is the parameter used in this paper.

In a randomized experiment, where a subset of $G$ are randomly selected for treatment and another subset for control, one may obtain unbiased estimates of average incremental response and average incremental ad spend, whose ratio then gives a natural estimate of $\theta^*$, referred to as the empirical estimator later, i.e.

\[
\theta^{(emp)} = \frac{\frac{1}{|T|} \sum_{g \in T} R_g - \frac{1}{|C|} \sum_{g \in C} R_g}{\frac{1}{|T|} \sum_{g \in T} S_g - \frac{1}{|C|} \sum_{g \in C} S_g},
\]

where $T$ and $C$ denote the set of geos in treatment and in control, respectively. Rearranging Equation (1), we can obtain:

\[
R_g^{(T)} - \theta_g S_g^{(T)} = R_g^{(T)} - \theta_g S_g^{(T)}.
\]

Based on this analysis, we can create estimators to solve for the value of $\theta^*$, which gives us an estimate for the population iROAS. The following table describes the notation as it shall be used in the rest of this publication.
Table 1: Description of the notation used for the $i$th pair of geos

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{ic}$, $S_{ic}$</td>
<td>Response and ad spend for the control geo</td>
</tr>
<tr>
<td>$R_{it}$, $S_{it}$</td>
<td>Response and ad spend for the treatment geo</td>
</tr>
<tr>
<td>$Y_i = R_{it} - R_{ic}$</td>
<td>Difference in the responses</td>
</tr>
<tr>
<td>$X_i = S_{it} - S_{ic}$</td>
<td>Difference in the ad spends</td>
</tr>
<tr>
<td>$\epsilon_i(\theta) = Y_i - X_i\theta$</td>
<td>Difference in the response background noise with respect to $\theta$</td>
</tr>
</tbody>
</table>

With a randomized paired design of geo experiments, the distribution of $\epsilon_i(\theta^*)$ is symmetric about zero for $i = 1, \ldots, n$. Therefore, the expected value of $\epsilon_i(\theta^*)$ is zero. To calculate the incremental return on ad spend, it is the goal of this publication to accurately estimate the value of $\theta^*$ based on $\epsilon_i(\theta^*)$. One way to estimate the value of $\epsilon_i(\theta^*)$ is from a binomial sign test. For any $\theta \in \mathbb{R}$, let $B_n(\theta)$ be the number of positive pairs defined by:

$$B_n(\theta) = \sum_{i=1}^{n} I(\epsilon(\theta) > 0),$$

where $I$ is the indicator function.

Therefore, $B_n(\theta^*)$ describes a binomial distribution of $n$ geos with a probability of $\frac{1}{2}$. Because $B_n(\theta^*)$ is symmetric about its expected value of $n/2$, a point estimate for $\theta^*$ can be obtained by first finding the values of $\theta$ that satisfy $B_n(\theta) = n/2$ if $n$ is an even number. If $n$ is an odd number, then the point estimate for $\theta^*$ can be obtained by finding all values of $\theta$ that satisfy the condition that $B_n(\theta^*) = (n-1)/2$ or $B_n(\theta^*) = (n+1)/2$. After finding the values of $\theta$ that satisfy one of the two conditions, an estimate can be calculated by averaging over the minimal and maximal values using the following equation:

$$\hat{\theta}(binom) = \frac{\inf \Theta_B + \sup \Theta_B}{2},$$

where $\Theta_B = \{\theta \in \mathbb{R} : |B_n(\theta) - \frac{n}{2}| \leq \frac{1}{2}\}$.

A more efficient estimator for $\theta^*$ can be found by trimming certain heterogeneous pairs in the set of geos. That is, by removing certain geos that may disproportionally affect the results...
of a causal experiment, one can arrive at a more accurate estimate for the \( \text{iROAS} \). This is called
the Trimmed Match method. The following derivation of the Trimmed Match method assumes
\( \epsilon_1(\theta) \leq \epsilon_2(\theta) \leq \epsilon_3(\theta) \leq \ldots \leq \epsilon_n(\theta) \) to be the corresponding order statistics. This method uses a fixed value, \( \lambda \), to be a fixed trim rate, where \( 0 \leq \lambda < \frac{1}{2} \). The trimmed mean statistic is defined as the following equation:

\[
\bar{\epsilon}_{n\lambda}(\theta) \equiv \frac{1}{n-2m} \sum_{i=m+1}^{n-m} \epsilon_i(\theta),
\]

where \( m \) is the minimal integer greater or equal to \( n\lambda \). It should be noted that \( \lambda \) must satisfy \( n - 2m \geq 1 \), otherwise all members of the set of geos would be trimmed away. Following the derivations above, the trimmed mean statistic has an expected value of zero. Therefore, we can estimate the value of \( \theta^* \) by solving the equation below:

\[
\bar{\epsilon}_{n\lambda}(\theta^*) = 0.
\]

When multiple roots exist, we choose the root which minimizes the statistic, in part using the equation below:

\[
D_{n\lambda}(\theta) \equiv \frac{1}{n-2m} \sum_{i=m+1}^{n-m} |\epsilon_i(\theta) + \epsilon_{n-i+1}(\theta)|,
\]

which measures the symmetric deviation from 0. The Trimmed Match estimator can be formally defined as:

\[
\theta^{(\text{trim})}_{\lambda} = \text{argmin} \{D_{n\lambda}(\theta); \bar{\epsilon}_{n\lambda}(\theta) = 0\}
\]
When two geos in the $i$th pair are perfectly matched, one expects $\epsilon_i(\theta^*) = 0$. Therefore, $\theta_{\lambda}^{(\text{trim})}$ has a nice interpretation: trim the poorly matched pairs in the sense of $\epsilon_i(\theta^*)$, and estimate iROAS based on the un-trimmed pairs. The algorithm to solve for the trimmed match estimate is included in Figure 1 below.

---

**Input:** $\{(x_i, y_i) : 1 \leq i \leq n\}$ and trim rate $\lambda > 0$;

**Output:** roots of Eq. (5.2).

i) Reorder the pairs $\{(x_i, y_i) : 1 \leq i \leq n\}$ such that $x_1 < \cdots < x_n$; Calculate $\{\theta_{ij} : 1 \leq i < j \leq n\}$ and order them such that $\theta_{i_1,j_1} < \theta_{i_2,j_2} < \cdots < \theta_{i_N,j_N}$. (Break ties with negligible random perturbation if needed)

ii) Start with $\theta = -\infty$ and initialize the set of untrimmed indices with

$$\mathcal{I} \leftarrow \{i : \lfloor n\lambda \rfloor < i \leq n - \lfloor n\lambda \rfloor\}$$

Calculate

$$a \leftarrow \sum_{i \in \mathcal{I}} y_i \text{ and } b \leftarrow \sum_{i \in \mathcal{I}} x_i.$$  

Initialize two ordered sets $\Theta_1 = \{}$ and $\Theta_2 = \{}$.

iii) For $k = 1, \cdots, N$:

(a) If $i_k \in \mathcal{I}$ and $j_k \notin \mathcal{I}$, then update

$$\mathcal{I} \leftarrow \mathcal{I} + \{j_k\} - \{i_k\},$$  

$$a \leftarrow a + y_{j_k} - y_{i_k}$$  

$$b \leftarrow b + x_{j_k} - x_{i_k}$$

and append $a/b$ to $\Theta_1$ and $\theta_{i_k,j_k}$ to $\Theta_2$, i.e.

$$\Theta_1 \leftarrow \Theta_1 + \left\{ \frac{a}{b} \right\}$$  

$$\Theta_2 \leftarrow \Theta_2 + \{\theta_{i_k,j_k}\}.$$  

(b) If $i_k \notin \mathcal{I}$ and $j_k \in \mathcal{I}$, then update

$$\mathcal{I} \leftarrow \mathcal{I} + \{i_k\} - \{j_k\}$$

and repeat the similar procedure as in (a).

(c) Otherwise, continue.

iv) Output a subset of $\Theta_1$:

(a) Append $\infty$ to $\Theta_2$;

(b) For $k = 1, \cdots, |\Theta_1|$, 

   (i) output $\Theta_1[k]$ iff $\Theta_2[k] \leq \Theta_1[k] \leq \Theta_2[k + 1]$.

---

**Figure 1:** The Algorithm to Implement the Trimmed Match Estimator
To correctly solve the trimmed match equation above, the algorithm in Figure 1 uses a fixed trim rate to determine which pairs of geos in the causal experiment to exclude based on how well they match. The geo pairs that are matched the most poorly are trimmed from the set, while maintaining the pairs that are matched very well for the experimental analysis. The idea of the algorithm is to look at all candidate values of $\theta$ as it grows from $-\infty$ to $\infty$, and identify the set of thresholds where the ordering of $\epsilon_i(\theta)$ changes whenever $\theta$ passes those thresholds.

In order for the Algorithm included in Figure 1 to work properly, a proper trim rate must be chosen. The trim rate is chosen by minimizing the asymptotic variance of $\theta_{\lambda}^{(trim)}$. The equation for an estimate of asymptotic variance can be found in the equation below.

\[
\sigma^2 = \frac{E(\epsilon^2 \Lambda q^2)}{[E(X \cdot I(|\epsilon| \leq q))]^2},
\]

In Equation (9), the value of $E(\epsilon^2 \Lambda q^2)$ is defined as:

\[
\hat{E}(\epsilon^2 \Lambda q^2) \equiv \frac{1}{n} \left( m(\hat{\epsilon}_{m+1}^2 + \hat{\epsilon}_{m+1}^2) + \sum_{i=m+1}^{n-m} \hat{\epsilon}_i^2 \right),
\]

and $E(X \cdot I(|\epsilon| \leq q))$ is defined as:

\[
\hat{E}(X \cdot I(|\epsilon| \leq q)) = \frac{1}{n} \sum_{i=1}^{n} X \cdot 1(\hat{\epsilon}_{m+1} \leq \hat{\epsilon}_i \leq \hat{\epsilon}_{n-m}),
\]

where $\hat{\epsilon} = Y_i - \theta_{\lambda}^{(trim)} X_i$. The value for the trim rate can be found by minimizing Equation (9) with respect to $\lambda$. 

ABSTRACT

A method for determining estimates for the incremental return on ad spend in randomized causal geo experiments are described. Randomized causal geo experiments are performed on pairs of matched geos, wherein one is randomly selected to be the control geo and the other is randomly selected to be the treatment geo. However, the estimates for the incremental return on ad spend depend on well matched geos to guarantee an accurate estimate. The algorithm described in this publication describes a method to automatically trim the most unmatched geo pairs from a causal geo experiment based on a fixed trim factor to increase the accuracy of incremental return on ad spend estimations.