# Marketing Campaign Evaluation in Targeted Display Advertising

Joel Barajas*
UC, Santa Cruz
Santa Cruz CA, USA
jbarajas@soe.ucsc.edu

Ram Akella
UC, Santa Cruz
Santa Cruz CA, USA
akella@soe.ucsc.edu

Marius Holtan
AOL Research
Palo Alto CA, USA
marius.holtan@teamaol.com

Jaimie Kwon
AOL Research
Palo Alto CA, USA
jaimie.kwon@teamaol.com

Aaron Flores
AOL Research
Palo Alto CA, USA
aaron.flores@teamaol.com

Victor Andrei
AOL Research
Palo Alto CA, USA
victor.andrei@teamaol.com

## ABSTRACT

In this paper, we develop an experimental analysis to estimate the causal effect of online marketing campaigns as a whole, and not just the media ad design. We analyze the causal effects on user conversion probability. We run experiments based on A/B testing to perform this evaluation. We also estimate the causal effect of the media ad design given this randomization approach. We discuss the framework of a marketing campaign in the context of targeted display advertising, and incorporate the main elements of this framework in the evaluation. We consider budget constraints, the auction process, and the targeting engine in the analysis and the experimental set up. For the effects of this evaluation, we assume the targeting engine to be a black box that incorporates the impression delivery policy, the budget constraints, and the bidding process. Our method to disaggregate the campaign causal analysis is inspired on randomized experiments with imperfect compliance and the intention-to-treat (ITT) analysis. In this framework, individuals assigned randomly to the study group might refuse to take the treatment. For estimation, we present a Bayesian approach and provide credible intervals for the causal estimates. We analyze the effects of 2 independent campaigns for different products from the *Advertising.com* ad network for 20M+ users each campaign.

## Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Economics*; G.3 [**Mathematics of Computing**]: Probability and Statistics—*Experimental Design*

*Main contact.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 20XX ACM X-XXXXX-XX-X/XX/XX ...$10.00.

## General Terms

Algorithms, Economics, Management, Measurement

## Keywords

Causal Attribution, Marketing, Online Targeted Advertising

## 1. INTRODUCTION

Evaluating the effectiveness of marketing campaigns is an important current problem in online display advertising. In this context, running a randomized experiment with study and control groups (A/B testing) has been suggested recently to evaluate marketing campaigns assuming the ad impression as the treatment to evaluate [10]. In this framework, users are assigned to each group randomly when they visit a specific publisher website, and the mean change of probability for the measure of interest is attributed to the campaign. However, this framework does not incorporate the real scenario in targeted advertising where several decisions are made before the ad impression is shown to a given user. Implementation of A/B testing to estimate the effect of the media ad suggests randomizing the users after the final decision of showing the ad impression to the user has been made. This idea does not take into account the implications of the other campaign components in the estimation and experimental set up. These components include: the targeting algorithm, budget constraints, and the auction process, among others.

We propose to evaluate the marketing campaign as a whole, not just the media ad design, using A/B testing. We randomize the users before any decision has been made to model the presence of the main campaign components in the evaluation. This provides the average causal effect of the whole campaign which is useful for campaign conversion attribution. Then, we disaggregate the analysis and estimate the effect of the media ad design, under the presence of a targeting algorithm given this randomization framework. We consider this algorithm to be a black box which dictates which users in the study group are exposed to the media ad. This is based on budget constraints and the user targeting policy. This measures the average causal effect of the advertising message as a campaign component.

In causal analysis, the user targeting selection introduces a bias in the evaluation as only the optimal subset of the

users in the study group is actually exposed. To eliminate this bias, a statistically equivalent set of users in the control group should be selected. In general, this selection is not observed for the control group since the bidding process is not performed by the analyzed campaign in this group. Our approach is inspired by the analysis of randomized experiments with imperfect compliance [7, 2], where individuals in the study group might refuse to take the treatment introducing some bias in the analysis.

This paper is organized as follows. In section 2, we discuss related work on online campaign evaluation. We also contrast our approach and problem framework with previous literature. In section 3, we discuss the general framework for targeted display advertising. We address the implications of this framework in the experimental set up and the technical challenge. In section 4, we define and develop the model to evaluate the campaign and the media ad design. Section 5 shows the results for two campaigns and discuss the highlights in terms of user conversion probabilities. Finally, in section 6, we provide a discussion of the method presented with future directions of improvement.

## 2. RELATED WORK

Recent research on campaign evaluation has focused primarily on two approaches: running randomized experiments, and bias correction based on user features in observational studies. In [9], a detailed method to estimate the impact of user exposure to ad impressions, based on A/B testing, is provided. Here, the authors verify the impact of ad exposure to users on their commercial actions. This is also recommended in [10], where the authors address potential over-estimation issues due to user activity bias. The authors in [3] and [14] propose methods to correct the bias in the user selection for ad exposure in observational studies. However, running experiments is generally preferable when advertisers assume the opportunity cost of avoiding advertising to a fraction of the user audience [10].

Previous work on evaluation using A/B testing shows the effect of user visitations to a web portal on search activity [10]. Here, all visitors are randomized and no targeting algorithm is taken into account. Generally, this framework is not the case for online display advertising where the delivery of ad impressions is dictated by a targeting algorithm. This algorithm is often based on budget constraints [8], a real time auction process in a market place [4], and user conversion probabilities [1]. The idea of a market target segment in a experimental framework has been recently addressed in campaign evaluation [6]. However, the targeting algorithm as a required condition to show an ad is not incorporated. In this framework, only the ad causal effect is analyzed for users inside and outside the target segments allowing the ad exposure to all the users visiting one publisher website.

Modeling the causal effect of marketing campaigns on the probability of user conversion has been suggested in a white paper by Collective Audience Engine [5]. Although the user randomization is performed to evaluate the campaign as a whole, they assume *all* the registered users are part of the control or study groups to guarantee statistical significance. Thus, users who never arrive at any publisher website are considered in this analysis. This incorrectly introduces bias, despite the apparent benefit, by over-smoothing the estimation.

In the current paper, we follow a experimental attribution approach for campaign evaluation. We design the experiment to measure the causal effect of the campaign as a whole on user conversion probability. For this design, we incorporate the main components of the real targeted advertising process: multiple publishers, budget constraints, the auction process, and the targeting engine. Given this design, we estimate the effects of the media ad design on conversion probability. We condition the analysis on the users who can potentially be exposed to the campaign, namely those that constitute the advertising demand.

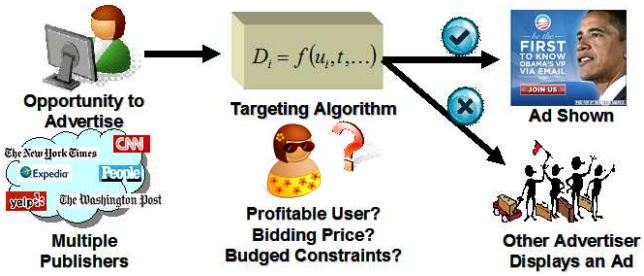



**Figure 1: Online targeted display advertising framework.**

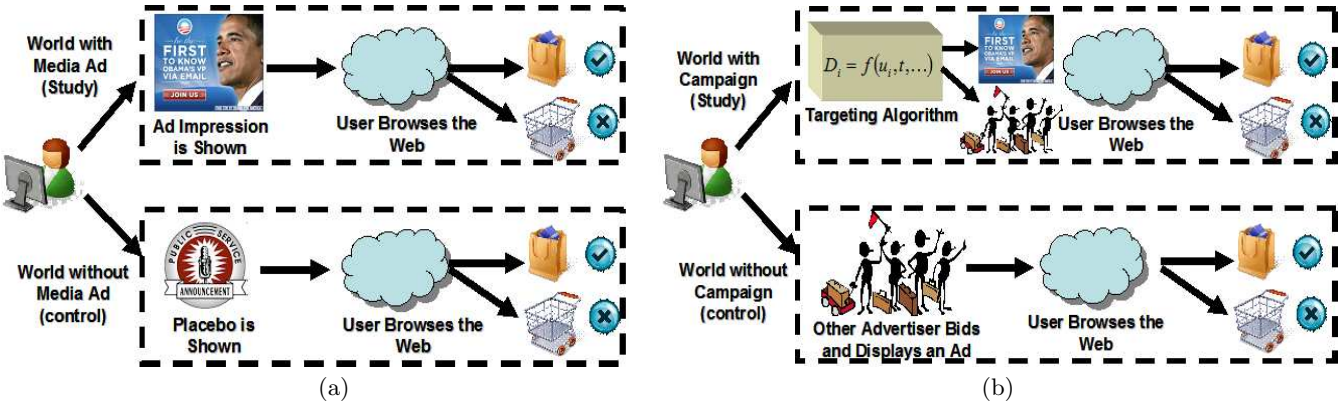
## 3. PROBLEM DESCRIPTION

### 3.1 Targeted Display Advertising Framework

Targeted advertising has become the standard practice in online display advertising. In this framework, marketing campaigns are run by advertisers through a close interaction with a given ad network. Figure 1 depicts the main components of this process. Here, to display an ad to a user, three main elements are usually present: a targeting algorithm, a bidding process, and a budget constraint. To target users, advertisers have a profile of the ideal user based on demographics. In practice, the ad network employs a more sophisticated algorithm to determine if a user is exposed to an ad based on how likely the user is to convert [1], user behavior and history, among other features. Often, an auction process takes place to win the advertising slot in a market place. Real time bidding is frequently performed. Similarly, this influences when an ad impression is shown to a user [4]. Moreover, marketing campaigns are run with budget constraints which also influence if the ad is shown to the user [8]. All these decisions are performed in a highly uncertain environment.

### 3.2 Implications in Evaluation

The traditional approach to evaluate a marketing campaign is to use A/B testing [10, 6] as illustrated in figure 2(a). Here, randomization is performed for every user that can be potentially exposed to the ad impression. Those users assigned randomly to receive the treatment are exposed to the ad. For the users assigned to be in the control group, a placebo, in the form of a public service announcement (PSA) or a completely unrelated ad, is shown to them. The underlying assumption is that the ad media design is the *treatment* and the use of a placebo completely identifies the control set of users. However, this procedure measures the effect of the campaign under very controlled conditions which are not the




(a) (b)

**Figure 2: User randomization framework for: (a) Campaign evaluation as proposed in [10], (b) Proposed campaign evaluation.**

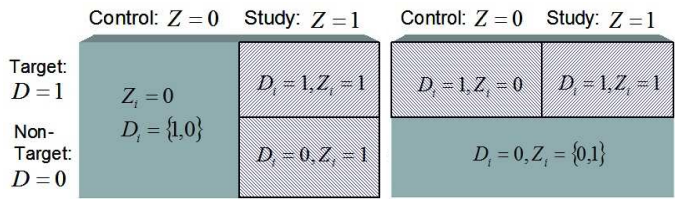
true conditions of online targeted advertising.

There are three fundamental differences between the evaluation of targeted advertising and the standard A/B testing evaluation. The first one is the presence of a highly sophisticated targeting algorithm. This is a required condition for ad exposure as discussed above. Defining ad exposure as treatment excludes the ability to select profitable users from the evaluation. This is an important limitation when the objective of the evaluation is to optimize the campaign.

The second difference concerns the use of a placebo. In medical treatment analysis, the main objective is to control as much as possible the effects of the treatment by simulating the administration to every individual in the experiment. In this framework, the impact on the response variable due to patient perception is minimized. The use of PSA in the evaluation of display advertising is convenient to clearly identify the users in the control group. However, this is not the actual situation encountered when multiple advertisers bid to gain the opportunity to advertise [4, 8]. In this context, we need to model what would happen if we decided (randomly) not to expose a user to the campaign given the advertising opportunity. This is a required condition to employ A/B testing effectively as a control group without the campaign, and consequently without campaign bidding process, needs to be modeled. Thus, in the absence of a marketing campaign and given the opportunity to advertise, other advertisers will compete to win this opportunity.

The third difference is campaign advertising budgets. Even when a user is optimum for ad exposure based on the targeting algorithm, the campaign might have consumed all the budget assigned. Ideally, all the desired audience should be exposed to the advertising message, as the experiments developed by the authors of [10, 6]. However, there is a market place which prices each opportunity to advertise, which suggests a limited budget allocation for each campaign.

To evaluate the campaign for attribution under real conditions, we randomize the users before they are *exposed* to any component of the campaign. Figure 2(b) illustrates this framework. We consider the whole campaign as the treatment to evaluate, where a campaign is composed of: the ad design and message, the targeting algorithm, and the specific impression budget. In this framework, users in the control group are not exposed placebos. In addition, there is no bidding process by the campaign in the control group thus the targeting algorithm is not run for these users. Given this randomization framework, we evaluate the media ad design




(a) (b)

**Figure 3: User segments based on control/study ($Z_i$) and non-targeted/targeted ($D_i$) groups. (a) Observed segments. (b) Idealized segments to evaluate the causal effects of the media ad design.**

and the user selection made by the targeting algorithm.

## 3.3 Technical Challenge

To evaluate the analyzed campaign, we estimate the average effect of the campaign as in standard A/B testing. This estimates the campaign effects on the exposed population under the conditions it is run. For attribution, this is the proper performance measure given the user targeting process, the campaign budget, and the bidding performed. However, this a very aggregated campaign performance measure. For campaign optimization, a more disaggregated evaluation is desired to attribute any success to the right components. Moreover, the final user is influenced by the campaign only when the ad impression is shown to him/her. Thus, in addition to the evaluation of the campaign as a whole, we estimate the effect of the media ad on the users whom the ad is shown to. This framework poses a technical challenge due to the selection bias introduced by the targeting algorithm in the study group (those users that are exposed to the campaign).

We approach the problem in the framework of randomized experiments with imperfect compliance and the intention-to-treat (ITT) analysis [7, 2, 12]. Here, individuals are assigned randomly to a study and control groups. However, individuals in the study group might refuse to take the treatment. In the online advertising problem, the user is randomly assigned to be exposed to the campaign as a whole or not. Then, the targeting algorithm, together with the bidding process, ultimately decides if the ad is shown to the user. The main challenge is that we do not observe if the users in the control group would have been assigned to see the ad impression. This is because no bidding process is per-

formed by the campaign in the control group. Therefore, to evaluate the effect of the media ad on the exposed users to the ad, we infer this assignment variable for the users in the control group. For this evaluation, we assume the targeting engine to be a black box that incorporates the impression delivery policy, the budget constraints, and the bidding process. This engine manages the advertising budget and the bidding process. Therefore, a targeted user is exposed to the ad impression when the user is part of the study group. Figure 3 shows the observed data, and the idealized segments to evaluate the effect of the media ad. These segments are explained in detailed in the following section.

# 4. METHODOLOGY

In this section we define the variables of interest and the model we use. We discuss details of the assumptions and their feasibility. Our goal is to measure the campaign and media ad effects on the probability that the users perform commercial actions. We denote these users as converting users. As discussed above, targeted users are those who are selected by the targeting algorithm. Thus, this a required condition to show the ad. We emphasize the difference with the target segments as defined by advertisers. Here, ad impressions might be shown to users who are not part of these segments [6].

## 4.1 Model Description

According to figure 3, which describes the study at hand, we define for user $i$ the following binary variables:

$\mathbf{Z}_i$ for control/study group user assignments, $\{0, 1\}$

$\mathbf{D}_i$ for non-targeted/targeted users, $\{0, 1\}$

$\mathbf{Y}_i$ for non-converting/converting users, $\{0, 1\}$

We model the above variables to be random binary variables. Thus, the probability of user $i$ being targeted, $D_i$, is Bernoulli distributed with parameter $\pi$. The probability of user $i$ performing one or more commercial actions, $Y_i$, is Bernoulli distributed with parameters $\theta_{dz}$ for the four combinations $D_i = d$, $Z_i = z$. Following a Bayesian approach in the parameter estimation, we define the set $\Theta = \{\theta_{dz}, \pi\}$ as random variables. Thus, we have the joint distribution:

$$P(Y, Z, D, \Theta) = P(\Theta) \prod_{\forall i} P(D_i|\pi) P(Z_i) P(Y_i|D_i, Z_i, \theta_{dz}) \quad (1)$$

One key assumption in this model is the stable unit treatment value assumption (SUTVA) [7]. This allows us to write the conversion probability for the $i^{th}$ user, $P(Y_i)$ as a function of $D_i$ and $Z_i$ conditionally independent on the distribution parameters $\theta_{dz}$. Thus no interactions among users is assumed. The probability of user $i$ being assigned to study or control groups, $P(Z_i)$ is independent of any variables which expresses the randomization of users. In practice, $Z_i$ is Bernoulli distributed with a predefined probability which is set when the experiment is designed. Due to the advertising opportunity cost of no running the campaign for the control group, this probability is rarely symmetric.

As illustrated in figure 3(a), in reality $D_i$ is not observed for users in the control group, $Z_i$=0. That is, the users in the control group are not exposed to the targeting algorithm. Thus we do not observe which users in this group would have been targeted had they been in the study group. We define $D_i^c$ and $D_i^s$ as the targeting indicator for users in the control and study group leading to the following joint distribution:

**Table 1: Observed Data Description.** $d$ **represents unobserved targeting assignment.**

| **Control Study,** $Z_i$ | 0 | 0 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|
| **Target** $D_i$ | $d$ | $d$ | 0 | 0 | 1 | 1 |
| **Convert** $Y_i$ | 0 | 1 | 0 | 1 | 0 | 1 |
| **Observed Counters** | $N_{d0}^0$ | $N_{d0}^1$ | $N_{01}^0$ | $N_{01}^1$ | $N_{11}^0$ | $N_{11}^1$ |

$$P(Y, Z, D, \Theta) = P(\Theta) \prod_{\forall i, Z_i=1} P(D_i^s|\pi) P(Z_i = 1) P(Y_i|D_i^s, Z_i = 1, \theta_{dz}) \prod_{\forall i, Z_i=0} P(D_i^c|\pi) P(Z_i = 0) P(Y_i|D_i^c, Z_i = 0, \theta_{dz}) \quad (2)$$

This suggests estimating the value of the targeting algorithm for those users in the control group, $D_i^c$, based on the latent potential outcome for those users. Note that $P(D_i|\pi)$ is the same for the control and the study group. This guarantees that the control group for the targeted users, those that are selected to be exposed to the media ad, is statistically equivalent. However, solving this problem directly introduces identifiability issues [11, 2]. That is, the targeting assignment is not identifiable based solely on the conversion probability.

In the context of ITT analysis, a commonly invoked assumption is the *weak exclusion restriction* to simplify and make the problem identifiable for inference [7]. In the targeted advertising context, the main requirement for this assumption is that the control/study assignment is unrelated to potential outcomes for non-targeted users. In terms of random variables:

$$Y_i \perp Z_i | D_i = 0$$

Thus, as the users are assigned randomly to control/study group and the ad impression is shown only to targeted users, $D_i = 1$, we have:

$$P(Y_i|D_i = 0, Z_i = 0, \theta_{00}) = P(Y_i|D_i = 0, Z_i = 1, \theta_{01}) \quad (3)$$

In other words, if user $i$ is not targeted, $D_i = 0$, no ad impression is shown to the user regardless of the control/treatment assignment, $Z_i$. This is depicted in figure 3(b) for the idealized segments. In contrast, this is not the case for the targeted group. Given that the user is chosen optimally by the targeting algorithm, $D_i = 1$, the randomized control/study indicator determines if the ad is shown to the user or not. Therefore, we define $\theta_{01} = \theta_{00} = \theta_0$ leading to $\Theta = \{\theta_0, \theta_{1z}, \pi\}$.

## 4.2 Model Estimation

We estimate the posterior distribution for $\Theta$ given the observed data. Table 1 illustrates the observed data counters. In addition, we need to estimate the posterior distribution for the unobservable targeting assignments, $D_i^c$, as defined by equation 2. To solve this inference problem, we follow a sampling approach based on Gibbs sampling to find the

posterior distribution for the parameters of interest $\Theta$. We assume standard Beta prior distribution for the Bernoulli parameters. To state full prior ignorance about the parameters $\Theta$, we use the Beta distribution with parameter values representing a flat uniform distribution, Beta$(1, 1)$. Therefore, we obtain the following conditional posterior distributions:

$$\begin{aligned} &P(\theta_{1z}|\cdot) = \text{Beta}(1+N^1_{1z}, 1+N^0_{1z}), \qquad z = \{0,1\} \\ &P(\theta_0|\cdot) = \text{Beta}(1+N^1_{00}+N^1_{01}, 1+N^0_{00}+N^0_{01}) \\ &P(\pi|\cdot) = \text{Beta}(1+N_1, 1+N_0) \end{aligned} \tag{4}$$

where $N^1_{dz}$ are the number of users who perform one or more conversions and $N^0_{dz}$ are those who do not, given $D_i = d, Z_i = z$. $N_0$ and $N_1$ are the counters for targeted and non-targeted users summed over $Z_i = \{0, 1\}$, and $Y_i = \{0, 1\}$.

The unobservable targeting assignments, $D^c_i$, are sampled as follows:

$$P(D^c_i = 1|\cdot) \propto \frac{\pi P(Y_i|D_i = 1, Z_i = 0, \theta_{10})}{\pi P(Y_i|D_i = 1, Z_i = 0, \theta_{10}) + (1-\pi)P(Y_i|D_i = 0, Z_i = 0, \theta_0)} \tag{5}$$

Given an initial guess for $\Theta$ and $D^c_i$, we sample the parameters $\{\theta_0, \theta_{10}, \theta_{11}, \pi\}$ and the targeting labels for the control group $D^c_i$ iteratively. After discarding a set of burn-in samples, we use the samples as empirical posterior distribution for the parameters of interest.

## 4.3 Causal Effect Estimation

We estimate the average causal effect for the campaign as a whole and for the media ad. We define the average causal effect of the media ad (MCE) on the users who are exposed to the ad, and the average media causal lift (MCL) as follows:

$$\begin{aligned} \text{MCE} &= \text{E}(Y|D = 1, Z = 1) - \text{E}(Y|D = 1, Z = 0), \\ \text{MCL} &= \frac{\text{MCE}}{\text{E}(Y|D = 1, Z = 0)} \end{aligned} \tag{6}$$

As discussed above, our goal is to estimate the effect on the probability of a user becoming a converter. Based on the joint distribution from equation 1, $Y_i$ is Bernoulli distributed with expected value $\theta_{dz}$. Thus, MCE and MCL are estimated as:

$$\text{MCE} = \theta_{11} - \theta_{10}, \quad \text{MCL} = \frac{\theta_{11} - \theta_{10}}{\theta_{10}} \tag{7}$$

Since we follow a Bayesian approach in the estimation, MCE and MCL are random variables. As detailed in section 4.2, we follow a sampling based approach. This provides a random sample of the posterior distribution for $P(\Theta|Y, D, Z)$, which facilitates the confidence interval estimation. Thus, we compute the statistics MCE and MCL from the random samples leading to their empirical distribution.

To estimate the average campaign causal effect (CCE), and the average campaign causal lift (CCL), we follow the standard A/B testing approach. Thus, we define:

**Table 2: Data description of the analyzed campaigns.**

| Campaign | Number of Users | Duration in Days |
|---|---|---|
| Campaign 1 | 19,284,148 | 28 |
| Campaign 2 | 24,076,054 | 27 |

$$\begin{aligned} \text{CCE} &= \text{E}(Y|Z = 1) - \text{E}(Y|Z = 0) \\ \text{CCL} &= \frac{\text{CCE}}{\text{E}(Y|Z = 0)} \end{aligned} \tag{8}$$

In standard A/B testing, a non-Bayesian approach is used in the estimation [10]. Based on the Central Limit Theorem, the distribution of $\text{E}(Y|Z)$ is approximated to a Normal distribution, and thus a $t$-test is performed for $CCE$. This is reasonable when a large user population is analyzed. However, $CCL$ is typically estimated based on the point estimates for $\text{E}(Y|Z = 1)$ and $\text{E}(Y|Z = 0)$ without providing a confidence interval [10]. To find the distribution of the statistics $CCE$ and $CCL$, we follow a Bayesian approach. This framework accounts for the actual population size without the large sample assumption. Therefore, we provide confidence intervals for $CCL$ without relying on other approximations.

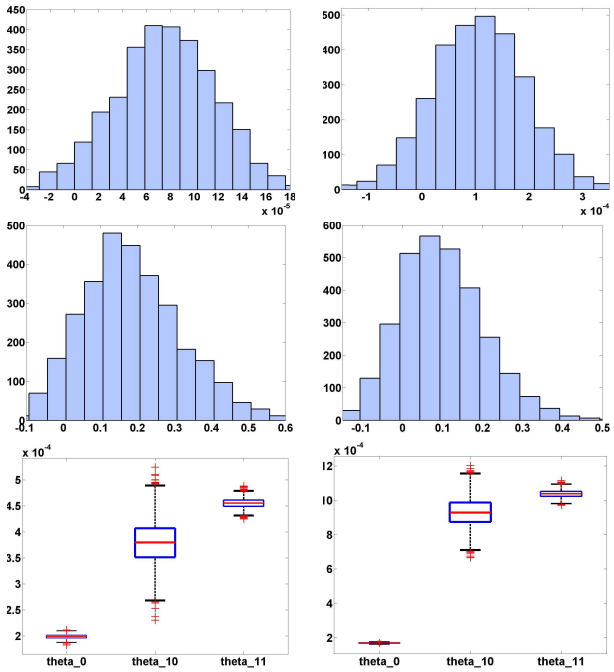
CCE and CCL estimate the causal effect of the campaign as a whole under the conditions it is run. These conditions include the budget constraints, and the objective function embedded in the targeting algorithm. For instance, under an infinitive budget, every advertising opportunity would be taken by the campaign, showing the media ad to every user. On the other hand, MCE and MCL estimate the effect of the advertising message, represented by the media ad, on the users exposed to the ad.

# 5. RESULTS

## 5.1 Experimental Settings

We consider the users who visit the websites where the media ad of the campaign is potentially displayed. We condition the analysis on the users arriving to these websites generating the advertising opportunity to the analyzed campaign. We randomize the users based on the last two digits of the time stamp their cookies are born. The percentage of control and study populations is set by these two digits randomly. This rule separates the users and leaves those in the control group without any contact with the campaign being evaluated. We notice that randomizing users in this fashion is simple to implement in an existing online targeting advertising engine. This is because the randomization rule is known before the user visits the publisher's website without the need of any adjustment at the ad serving time. A similar rule to randomization is suggested in [10] where the authors propose to randomize users based on the last digit of the time stamp of the user arrival to the website of the experiment. However, it requires changes at the time the ad is delivered. This is more problematic to implement in an existing online targeting advertising engine. For the experiments developed in this paper, we use 10% of users for the control group, $\text{P}(Z_i = 1) = 0.9$. We consider users whose cookie is active through the time window the campaign is evaluated to avoid cookie contamination.

We analyze the effects of two independent campaigns for




(a) (b)

**Figure 4: Model fitting results for: (a) campaign 1, (b) campaign 2. From top to bottom, posterior distribution for MCE, MCL, and the box plot for $\theta_0$, $\theta_{10}$, $\theta_{11}$. For distributions, $x$-axis and $y$-axis represent expected conversion probability and frequency respectively. For box plots, $y$-axis is the expected conversion probability.**

different products. Thus we conduct independent experiments for each case. Table 2 shows the details of these campaigns. For model fitting, we use 500 burning iterations and $3,000$ samples for the posterior distribution.

## 5.2 Campaign and Media Evaluation Results

Figure 4 shows the model fitting results for the campaigns analyzed. The posterior distribution of MCE is similar to a Normal distribution in shape. However, the posterior distribution for MCL shows a positive skew. Traditionally, lifts are estimated from the expected values once these have demonstrated to be statistically significant [3, 10]. However, as illustrated, the posterior distribution for MCL is not necessarily symmetric. This is a required condition to provide confidence intervals based on the Central Limit Theorem (Normal approximation).

The posterior distributions for $\{\theta_0, \theta_{10}, \theta_{11}\}$ are depicted at the bottom boxplot of figure 4. We observe that $\theta_{10}$ shows more variability as this distribution depends on the inference of the targeting indicator for the control group ($D_i^c$). The posterior distribution for $\theta_0$ shows the lowest variability of the set of random variables. This is because, in practice, just a fraction of the users in the study group are targeted. This selection translates into a larger remaining population to estimate $\theta_0$. We notice that the posterior distribution for $\theta_{11}$ is estimated directly as $D_i^s$ and $Z_i$ are fully observed.

A significant difference is evident between the conversion rates for the non-targeted ($\theta_0$) and the targeted ($\theta_{1z}$) groups.

**Table 3: MCE and MCL results for 90% confidence level.**

| | MCE | | | MCL (%) | | |
|---|---|---|---|---|---|---|
| | Low | Med | High | Low | Med | High |
| Camp 1 | 4.00E-5 | 7.55E-5 | 14.30E-5 | 0.88 | 19.88 | 45.83 |
| Camp 2 | -0.36E-4 | 1.07E-4 | 2.49E-4 | -3.37 | 11.53 | 31.06 |
| | **CCE** | | | **CCL (%)** | | |
| | Low | Med | High | Low | Med | High |
| Camp 1 | 0.13E-5 | 2.38E-5 | 4.55E-5 | 0.46 | 9.25 | 19.08 |
| Camp 2 | -0.43E-5 | 1.28E-5 | 2.95E-5 | -1.56 | 4.98 | 11.90 |

This demonstrates the selection bias introduced by the targeting algorithm to show ad impressions. When campaign evaluation is performed in an observational study (without experimentation), this is the selection bias that has been addressed and diminished previously in the literature [3, 14]. We emphasize that this selection is not made by nature as a highly sophisticated targeting algorithm select those users optimally. In addition, this difference depicts the ability of this algorithm to choose the optimal users against a random targeting. To isolate this shift of the media ad effects, we analyze the shift in the control group:

$$\mathrm{E}(Y|D,=1,Z=0) - \mathrm{E}(Y|D=0,Z=0) = \theta_{10} - \theta_0$$

Thus, under the same budget constraints, the targeting algorithm produces this shift compared to a random user selection. For these campaigns, that average difference is greater than 100%.

Table 3 shows the campaign and media attribution results for the campaigns analyzed. We achieve statistical significance for campaign 1 at 90% confidence level. Campaign 2 is leaning towards positive values but with a small negative range in the interval. We notice that both campaigns report non-tight confidence intervals as conversions are sparser when compared to clicks, surveys or search keywords.

We observe that the campaign evaluation estimates, CCE and CCL, are in general lower. This is because each campaign is run with a limited budget (as in a real scenario), thus just a few users are exposed to the media ad. This diminishes the effect of the campaign as a whole compared to showing the media ad to every user with a unlimited budget (assuming the exposure does not hurt). Conditional on constant MCL and MCE, CCL and CCE can be used to adjust the budget of a campaign mid-flight and the targeting algorithm. This would close the feedback loop between campaign evaluation and optimization at least in the design of the targeting policy. Performance estimation based solely on MCL and MCE is helpful just to adjust the the media ad design, not the other components of a campaign.

# 6. CONCLUSION AND FUTURE WORK

We have developed an effective method to estimate the causal effect of the media ad design and the marketing campaign as a whole. We assume experiments are feasible to run but only to evaluate the whole campaign. We assume a dynamic market place where advertisers bid to win the opportunity to advertise under a limited budget. We have introduced the presence of a targeting algorithm in the analysis as this is often the real scenario in online advertising. Budget constraints, targeting policy, and the bidding process have been embedded in the targeting decision variable.

Inspired by the analysis of experiments with imperfect compliance [7, 2], we estimate the latent conversion probability difference between targeted and non-targeted users in the control group. This is based on the observed conversion rates for targeted/non-targeted users, and the targeting user rates in the study group. This estimation is key to calculate the causal effect of the media ad on the targeted population. In addition, we develop a Bayesian approach which provides confidence intervals for the causal estimates. We model conversion probabilities as Bernoulli distributions assuming no observable features are available. One key advantage is that this estimation procedure can be incorporated into an aggregation framework naturally by the use of the prior distribution. This can potentially improve estimation for sparse signals, such as user conversions.

As on-going work, we incorporate features in the analysis, when available. In this framework, we can model the conversion probability conditional on those features, and the probability of the users in the control group of being targeted. Following the framework we have presented, this analysis is similar to the widely used propensity scores [13] but with the inclusion of the latent assignment in the analysis and the partial randomization. Although we can assume a fully observational study in the study group, the user randomization is highly useful in the causal estimation as discussed in [2, 12]. Other extensions of this work also include modeling the total number of conversions per user, instead of converting/non-converting users, as the variable to estimate the causal effect on.

Evaluation of the targeting algorithm is another direction of on-going work. We have presented a framework to evaluate the targeting algorithm by comparing the conversion probabilities of the targeted users against a random targeting. A different evaluation question is to model what would have happened had the media ad been shown to all the users. The fundamental difficulty is to model the *absence* of a targeting engine. Our ultimate goal is to evaluate all the decisions that take place during the online advertising process under the sparsity conditions of user conversions.

## 7. REFERENCES


[1] D. Agarwal, S. Pandey, and V. Josifovski. Targeting converters for new campaigns through factor models. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 101–110, New York, NY, USA, 2012. ACM.

[2] A. Balke and J. Pearl. Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92:1171–1176, 1997.

[3] D. Chan, R. Ge, O. Gershony, T. Hesterberg, and D. Lambert. Evaluating online ad campaigns in a pipeline: causal models at scale. In *Proceedings of the 16th ACM SIGKDD*, pages 7–16. ACM, 2010.

[4] Y. Chen, P. Berkhin, B. Anderson, and N. R. Devanur. Real-time bidding algorithms for performance-based display ad allocation. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '11, pages 1307–1315, New York, NY, USA, 2011. ACM.

[5] A. E. Collective. Causal attribution: Proposing a better industry standard measure of digital advertising effectiveness. Available Online: `http://collective.com/insight/causal-attribution`, 2012. White Paper.

[6] A. Farahat and M. C. Bailey. How effective is targeted advertising? In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 111–120, New York, NY, USA, 2012. ACM.

[7] G. W. Imbens and D. B. Rubin. Bayesian Inference for Causal Effects in Randomized Experiments with Noncompliance. *The Annals of Statistics*, 25(1):305–327, 1997.

[8] K. J. Lang, B. Moseley, and S. Vassilvitskii. Handling forecast errors while bidding for display advertising. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 371–380, New York, NY, USA, 2012. ACM.

[9] R. Lewis and D. Reiley. Retail advertising works! measuring the effects of advertising on sales via a controlled experiment on Yahoo!, 2009. White Paper.

[10] R. A. Lewis, J. M. Rao, and D. H. Reiley. Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of the 20th international conference on WWW2011*, pages 157–166. ACM, 2011.

[11] C. F. Manski. Nonparametric bounds on treatment effects. *American Economic Review*, 80(2):319–23, May 1990.

[12] J. Pearl. *Causality: Models, Reasning and Inference*. Cambridge University Press, 2000.

[13] P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[14] O. Stitelman, B. Dalessandro, C. Perlich, and F. Provost. Estimating the effect of online display advertising on browser conversion. In *Proceedings of the Fifth International Workshop on Data Mining and Audience Intelligence for Advertising (ADKDD 2011)*, pages 8–16, 2011.