UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ONLINE DISPLAY ADVERTISING CAUSAL ATTRIBUTION AND EVALUATION

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

Joel Barajas Zamora

September 2015

The Dissertation of Joel Barajas Zamora
is approved:

Professor Ram Akella, Chair

Professor Raquel Prado

Professor Philip B. Stark

SVP NativeX, Dr James G. Shanahan

Tyrus Miller
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Online Display Advertising Causal Attribution and Evaluation

by

Joel Barajas Zamora

The allocation of a given budget to online display advertising as a marketing channel has motivated the development of statistical methods to measure its effectiveness. Recent studies show that display advertising often triggers online users to search for more information on products. Eventually, many of these users convert at the advertiser's website. A key challenge is to measure the effectiveness of display advertising when users are exposed to multiple unknown advertising channels.

We develop a time series approach based on Dynamic Linear Models (DLM), to estimate the impact of ad impressions on the daily number of commercial actions when no user tracking is possible. This method uses aggregate observational data post-campaign and does not require an experimental set-up. We incorporate persistence of campaign effects and account for outliers in the time series without pre-defined thresholds. We analyze user conversions for 2,885 campaigns and 1,251 products during six months for model selection.

The current industry practice measures the campaign causal effect on online conversions by running a randomized experiment focused on the ad exposures (using placebo ads). This method ignores other campaign components, including user targeting in marketplaces with competitor effects. We propose a novel randomized design to estimate campaign and ad attribution in marketplaces. We determine the effect on the targeted users using the

Potential Outcomes Causal Model and Principal Stratification. We analyze the impact of 2 performance-based (CPA) and 1 Cost-Per-Impression (CPM) campaigns with 20M+ users for each campaign. We estimate a non-zero CPM campaign presence in the marketplace effect (currently ignored by industry). Evidence suggests that CPA campaigns incentivize targeting of users who buy regardless of the ad (always-buyers).

We propose a user targeting simulator that leverages data from campaign randomized experiments. Based on the response of 37 million visiting users (targeted and non-targeted) and their demographic user features, we simulate different user targeting policies. We provide evidence that the standard conversion optimization policy shows similar effectiveness to that of uniform targeting, and is significantly inferior to other causally optimized targeting policies. These results challenge the standard practice of targeting users with the highest conversion probability.

To guide the user targeting to optimize causally generated conversions, we analyze the campaign on the conversion probability of the users who click on the ad as a behavioral feature. We show that designing a randomized experiment to evaluate this effect is infeasible, and propose a method to estimate the local effect on the clicker conversions. Based on two large-scale randomized experiments, performed for 7.16 and 22.7 million users, a pessimistic analysis shows a minimum increase of the effect on the clicker conversion probability of 75% with respect to the non-clickers. This evidence contradicts a recent belief that clicks are not indicative of campaign success.

To my wife and children,

Karla, Dannah, Minerva and Cesar,

For being there during the most difficult times.

Acknowledgments

I want to thank and acknowledge the support from many people and institutions who have contributed one way of the other to the completion of this dissertation.

I want to thank my advisor Ram Akella for the opportunity to collaborate with companies from the Silicon Valley. The freedom I have had to lead and propose solutions in these collaborations is remarkable. These interactions have shaped my professional skills without a doubt. I have learned to be independent, to ground ideas, to structure the unstructured, to be unafraid of saying “I am not familiar”. More importantly, I have learned to navigate the waters of multiple research communities and personalities.

As a company, I want to thank AOL enormously for giving me the opportunity to conduct my dissertation research over all these years. The access to real production systems and data has helped me build invaluable field experience. A big thank you to Marius Holtan, Aaron Flores, Jaimie Kwon, Victor Andrei, and Brad Null for the great moments have experienced while working with you on my research. I greatly appreciate the unconditional help and support from Aaron during the ups and downs of this journey. I feel highly grateful to Marius for the incredible wisdom and knowledge of advertising that he has been always willing to discuss. You all have made me feel part of a team.

I want to thank my thesis committee members, Prof. Raquel Prado, Prof. Philip B. Stark, and Dr. James G. Shanahan. You all have been remarkably flexible and helpful to accommodate the constraints of my defense and review my dissertation work. I really appreciate the detailed feedback that Raquel has provided to me for the thesis defense and the advancement to candidacy. I greatly enjoyed Raquel’s time series class to the point that it inspired my early thoughts and solutions presented in this Dissertation. I want to thank Philip for the incredible wisdom of causal inference and the invaluable discussions of my Dissertation in this context. Interacting with Philip as a professor and as a person has been a really pleasant experience. I am very thankful to Jimi for the incredible advertising knowledge and for being always helpful through the years. I want to thank Jimi for his impact on my communication and presentation skills.

I want to recognize and thank Emily Gregg, Kris West, Carol Mullane, Connie Miller, and Gisela Richards for helping me above and beyond the call of duty. I have felt protected in the navigation of the administrative system by them in one way or the other. A special recognition to Emily, who has been available unconditionally when I needed her advice. I wish to thank the following agencies and institutions for their continuing financial support: CONACYT, UC MEXUS, UC Santa Cruz Regents, AOL Advertising, CITRIS. My graduate school experience would not have been possible without their generous support.

A big thank you to my former advisor friend and compadre Joaquin Salas and comadre Maria Elena Meza. Their support and experienced advice in the pursuit of the Ph.D. degree has been invaluable. I want to recognize my parents, David Barajas and Margarita Zamora, who have given me the power to decide since I was a child. Their advice, support, and trust have forged my character and have contributed tremendously to my professional development.

I am infinitely thankful to my daughter Dannah for lending her dad to the development of this Dissertation. Thank you Dannah for the countless nights that she saw me working, for being willing to be in daycare the longest time possible, and for being remarkably understanding of the stress of deadlines. In short, thank you for all the cost that represents having a dad in graduate school. The same appreciation I want to express to my two younger twin children: Minerva and Cesar. They all have been my support even at the deepest down of my journey.

Finally, I want to thank my wife and colleague Karla. I have enjoyed our Ph.D. journey together infinitely, as well as our parenthood journey. She has been incredibly supportive of me during deadlines and the up and downs of my Ph.D. I will always be immensely thankful for her support to move forward even from the deepest fall.

Part I

Background and Context

Chapter 1

Introduction

1.1 Motivation and Background

According to the Interactive Advertising Bureau (IAB) and PricewaterCoopers (PWC), internet display advertising revenues in the U.S. totaled US $6.5 billion during the first six months of 2014. This revenue amount represents 28% of the total online advertising ($23.1 billion) and constitutes an increase of 6% over the $6.1 billion reported over the same period of 2013. Due to the proliferation of the tracking of online user activity, performance-based Cost-Per-Action (CPA) campaigns accounted for 65% of the total campaigns. On the other hand, 34% of campaigns were run under the more traditional Cost-Per-Impression (CPM) business model[1]. In this context, determining the effectiveness of an online campaign in achieving increased user commercial actions is usually employed to give credit to CPA campaigns. This process is defined as *campaign attribution.*

Empirical evidence has demonstrated that improved attribution leads to more

[1] Source: IAB internet advertising revenue report. 2014 first six months' results. http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_HY_2014_PDF.pdf

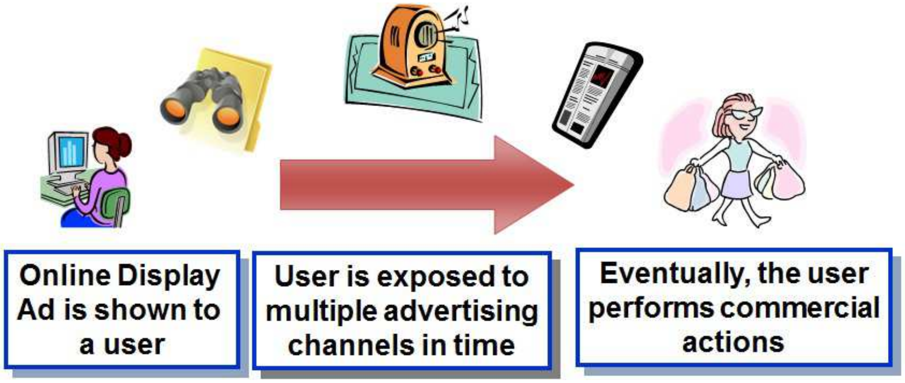

Online Display Ad is shown to a user
User is exposed to multiple advertising channels in time
Eventually, the user performs commercial actions

Figure 1.1: Online Display Advertising evaluation framework. How can the effectiveness of display advertising be measured when the user is exposed to various unobservable media channels?

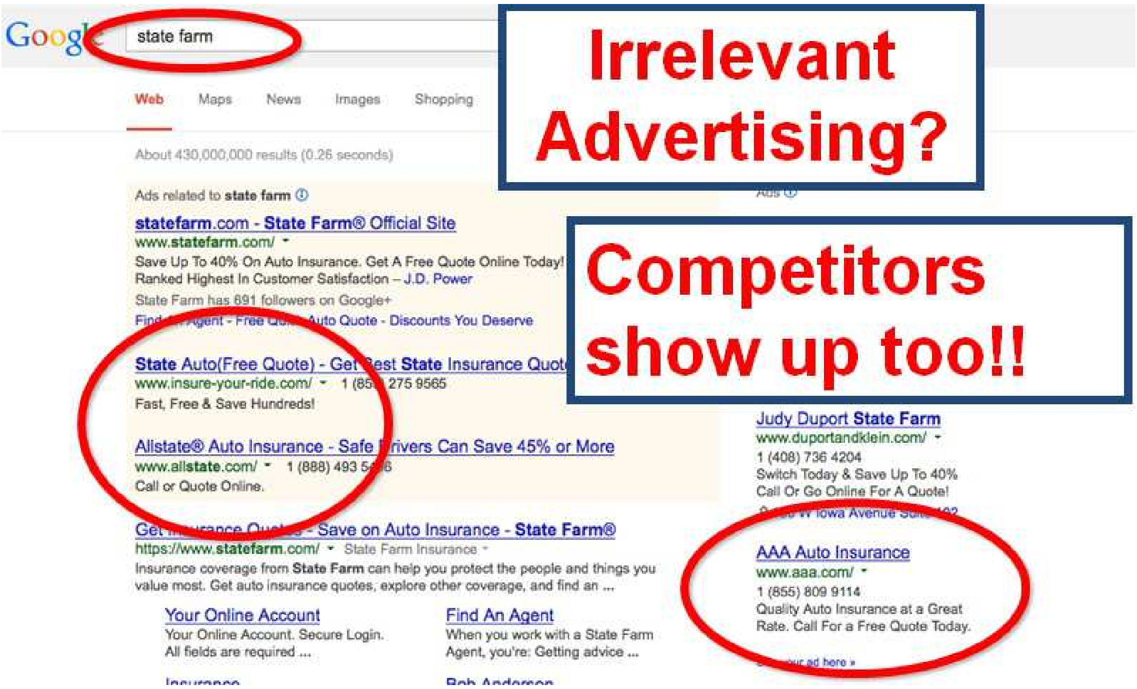
sales (or conversions) at a lower price for advertisers [80]. Also, a comprehensive survey of marketers and agencies, performed by Econsultancy and Google, reports that 83% of the respondents have engaged in attribution for less than two years by 2012 [32]. This gradual acceptance of attribution methods poses more challenges as firms and advertisers explore their benefits.

Recent studies show that display advertising often triggers online users to search for more information about commercial products [3]. Eventually, many of these users successfully convert at the advertiser's website. A key challenge is measuring the effectiveness of display advertising in such cases. Fig 1.1 shows what a user might see on visiting a site. Given the number of messages from different advertisers, how should an advertiser attribute credit for any conversion across multiple unobservable channels and media impressions?

To approximate the campaign effectiveness for thousands of campaigns, routinely run by large ad networks, treatment evaluation techniques for observational (post-campaign) studies have been developed and deployed. For instance, Google [22] and Yahoo! [84] have developed propensity scores based methods. Similarly, Turn Inc [78] implements a structural equation-based method by regression analysis of the user channel exposures and conversions. However, these approaches require user features, which are often missing and incomplete, and the availability of multiple user channel exposure information. Besides, these methods require a fixed time window, where ad networks track users and collect measures of interest. As a result, estimation is aggregated for that time window.

The use of randomized experiments, also known as A/B testing in industry or field experiments in economics, has demonstrated to be effective to evaluate the marginal effectiveness of marketing campaigns without over-estimation [55]. More importantly, randomized experiments guarantee the causal attribution of these campaigns when the experiment is designed correctly. The standard industry practice to design the randomized experiment assumes the ad creative is the *treatment* to evaluate. Thus, the users are randomly separated into two groups: study and control. Hence, when a targeting engine selects a visiting user for ad exposure, the campaign ad is displayed to users in the study group, or a placebo ad is displayed to users in the control group [87].

Recently, marketing campaigns are increasingly taking place on ad exchange platforms. These platforms facilitate marketplaces where advertising spaces on websites are bought and sold. Here, supply side platforms (SSP) provide visiting users to a publisher website, and demand-side platforms (DSP) bid on these opportunities to display an ad to these individual users in a real-time auction [79]. A survey of 49 media buyers indicates that 87.8% intended to purchase digital advertising via real-time bidding (RTB). Similarly, 48% of 101 publishers planned to switch to an ad exchange platform by 2011 [31]. Even



Irrelevant Advertising?

Competitors show up too!!

About 430,000,000 results (0.26 seconds)

Ads related to **state farm**

statefarm.com - State Farm® Official Site
www.**statefarm**.com/
Save Up To 40% On Auto Insurance. Get A Free Quote Online Today!
Ranked Highest In Customer Satisfaction – J.D. Power
State Farm has 691 followers on Google+
Find an Agent - Free Quote - Auto Quote - Discounts You Deserve

State Auto(Free Quote) - Get Best **State** Insurance Quote
www.insure-your-ride.com/ 1 (85 275 9565
Fast, Free & Save Hundreds!

Allstate® Auto Insurance - Safe Drivers Can Save 45% or More
www.all**state**.com/ 1 (888) 493 5 6
Call or Quote Online.

Get Insurance Quotes - Save on Auto Insurance - **State Farm**®
https://www.**statefarm**.com/ State Farm Insurance
Insurance coverage from **State Farm** can help you protect the people and things you value most. Get auto insurance quotes, explore other coverage, and find an ...

Your Online Account
Your Online Account. Secure Login. All fields are required ...

Find An Agent
When you work with a State Farm Agent, you're: Getting advice ...

Insurance

Rob Anderson

Ads

Judy Duport **State Farm**
www.duportandklein.com/
1 (408) 736 4204
Switch Today & Save Up To 40%
Call Or Go Online For A Quote!
W Iowa Avenue

AAA Auto Insurance
www.aaa.com/
1 (855) 809 9114
Quality Auto Insurance at a Great Rate. Call For a Free Quote Today.

your ad here »

Figure 1.2: Search advertising instance where competitors appear next to *state farm* brand keyword.

outside RTB exchanges, ad networks run internal auctions in regular basis [18].

In a competitive marketplace, the mere presence of the focal campaign has a potential effect on the user conversions (campaign presence effect) by preventing other ads from being displayed in this slot. As illustrated by Fig 1.2, competitors often target a similar user population, and consequently the ad slot is valuable. To display a placebo ad in the control group, the opportunity to advertise in a given ad slot must be *consumed*. However, this placebo ad display would not occur had the campaign not existed. Hence, the effect of the campaign presence in the marketplace represents a drawback of the standard industry practice of using placebo ads in the randomized design. In the absence of the focal campaign in the marketplace, other advertisers would have displayed their ads in the available ad slots, rather than a placebo ad. That is the relevant campaign counterfactual. Besides, the external validity of campaign effects estimated in an environment assumed to be free of competitors, to a marketplace with competitors, is likely to be inaccurate. Because media buying is performed endogenously in a competitive market, the user targeting becomes endogenous (or non-ignorable) and complicates the evaluation using placebo ads.

In the online advertising industry, user targeting is one of the most important decisions in running a campaign. A survey of 100 marketers, agencies, and media planners indicates that respondents perceive the user targeting and the campaign optimization capabilities as the main differentiators among ad networks [64]. This importance has motivated the development of a whole research area to manage the campaign ranging from user targeting to bidding policy development in ad exchanges [18, 43]. Also, online conversion attribution methods have been developed by industry. These methods include: Last-Touch Attribution (LTA) and Multi-Touch Attribution (MTA). LTA assigns the conversion credit to the last campaign exposure (touch point) to a user in the path to conversion. Similarly, MTA gives credit to the a set of touch points in this path [3].

The deployment of CPA campaigns, which generate revenue to ad networks based on these industry attribution practices, has produced increasingly sophisticated targeting engines. However, in these frameworks, credit must be given for an exposed and converter user even if this ad exposure does not influence the user. Therefore, user targeting development has focused primarily on optimizing online conversions by serving ads to the users who are most likely to convert based on machine learning techniques [1]. Often the evaluation of these algorithms is based on the prediction power of conversions (area-under-the-curve AUC or related methods [66]), which are likely to be not caused by the campaign [55]. These practices do not guarantee a causal impact optimization and incentivize the targeting of baseline users [14], those who convert regardless of the touch point (*always-buy* users). As a result, current user targeting practices provide limited value to advertisers when the objective is to increment conversion rates, as opposed to brand advertising via CPM campaigns.

Identifying user behavioral response features that correlate with causally generated conversions provides alternative signals to be optimized in the user targeting process. The user click on a given displayed ad represents one of the earliest behavioral features that advertisers consider as campaign success. However, the recent belief that user clicks are not informative of this success is increasingly gaining acceptance in the research community and industry. Dalessandro *et al.* (2012) concluded that user clicks do not correlate with user conversions and that user targeting based on clicks is statistically indistinguishable from random guessing [28]. These conclusions are drawn based on the power of user clicks to predict conversions in observational data. However, a large percentage of these conversions are likely to be unrelated to, and not caused by, the campaign, as it is standard in online advertising attribution analysis [55].

A more accurate approach is to measure the campaign effect on the conversion probability of the users who click on the ad (clickers) with a randomized experiment. Based on this estimation, we can determine the importance of the click in the user targeting optimization. However, to design such experiment one would need to randomize the users into control/study groups after finding the clickers. This randomized design is not feasible because, the campaign must display the ad to the users of the study and the control groups to observe the user selection introduced by the click event. Ideally, this user selection needs to be known before the campaign, or the placebo ad is displayed to randomize the clicking users properly.

1.2 Dissertation Contribution And Organization

We summarize the main contributions of this Dissertation to be the following:

1. An observational time series approach is proposed to estimate the campaign conversion attribution for thousands of campaigns based on aggregate conversion and campaign impression data.

2. We propose a new experimental design to estimate the effectiveness of online display advertising campaigns in marketplaces. We show that this design eliminates the risk of selection bias post-randomization for highly optimized performance based CPA campaigns.

3. We propose a causal inference method to estimate the campaign local average treatment effect (LATE) on the targeted users, and characterize this selection based on the probability of targeting influenceable users. We find evidence that performance-based CPA campaigns incentivize the targeting of users who buy regardless of the ad exposure when compared with impression based CPM campaigns.

4. Motivated by the CPA campaign targeting incentives, we propose to optimize causally generated conversions (campaign value) in the user targeting process. We propose an

offline evaluation methodology based on randomized experiment logged data and test different targeting policies.

5. We present a causal inference method to estimate the ad local treatment effect on the users who click on the ad. We find evidence that user clicks correlate with causally generated conversions, which suggests that the targeting optimization process should incorporate this behavioral signal.

This Dissertation is organized in five parts as follows:

Part I: Background and Context

We discuss Introduction and Related work in this part. Online display advertising general context and the challenges of campaign attribution are introduced. Also, causal inference concepts and relevant literature is discussed.

Part II: Observational Analysis: Campaign Evaluation at Scale

A method to estimate the value of display advertising campaigns post-campaign (observational) is proposed. This approach adjusts for some confounding factors using the predictions of a baseline Dynamic Linear time series Model. The analysis of thousands of campaigns semi-automatically is achieved by analyzing aggregate campaign impression and conversion time series only.

Part III: Campaign Attribution based on Randomized Experiments

We develop a dynamic analysis of the campaign conversion attribution based randomized experiments and aggregate conversion time series. We propose a novel randomized experimental design to model the appropriate campaign counterfactual, which includes the effect of the campaign presence in the marketplace. We contrast this design with current industry practices based on detailed analysis of the Targeted Display Advertising and Real-Time Bidding in ad networks and exchanges. We propose a causal inference approach in the Potential Outcomes causal framework, to analyze and compare the user targeting of two performance-based CPA and one display based CPM campaigns.

Part IV: From Prediction based to Causal based Targeting

We find evidence that standard user targeting objective of optimizing conversions incentivize the targeting of users who convert regardless of the ad exposure, and propose a user targeting offline evaluation based on a randomized experiment logged data. We present a method to optimize the targeting of causally generated conversions. Also, we present a causal inference method to estimate the value of the user clicks to incorporate them as part of the objective function of the targeting process.

Part V: Closing Remarks

We conclude by discussing the business implications of the methods and results we introduce here. Also, we discuss further research and field studies suggested by the current Dissertation contributions.

This Dissertation is partially based on the publications: [5, 6, 7, 8, 10, 13]. Additional publications developed as part of the current Dissertation include: [9, 11, 12, 53].

Chapter 2

Related Work

According to a literature survey performed by Ha (2008), online advertising as an advertising medium dates back to 1996, when Berthon *et al.* (1996) proposed a set of "success" metrics [47, 15]. This review shows that the majority of the literature before 2008 relied on lab experiments or the analysis of observational data with rudimentary data management. The author suggests the examination of the economic impact of online ads and the assessment of their effectiveness, among others, as part of the future research agenda.

The high dependency of the web on online advertising as a business model has surfaced the need for accurate assessment techniques of online campaigns. In this context, the advertising industry has developed methods for online conversion attribution. These methods include Last-Touch Attribution (LTA), Multi-Touch Attribution (MTA) and Algorithmic Attribution. LTA attributes all credit for a user conversion to the last ad viewed or clicked. Likewise, MTA heuristically split the conversion credit across the touch points in the path to conversion [3]. Algorithmic Attribution often refers to prediction models that measure the correlation of user exposures with sales. These attribution modeling tools are often well accepted by industry users as providers of data-driven insights [72]. Model driven MTA approaches have been proposed to model interacting channel effects and their attribution [78, 59]. However, these methods assign attribution credit to every user conversion-exposure co-occurrence while ignoring the counterfactual response without exposures. More importantly, they do not attempt to achieve causal attribution.

Online Display Advertising poses unique challenges to measuring the causal effect of online campaigns on the probability of user conversions. These challenges include small user conversion propensity (typically in the order of 1 in 10,000 or less [12]), small average causal effects and campaign lifts [58]. Moreover, a severe user selection bias is introduced by sophisticated campaign management that targets users and executes the bidding policy in ad exchanges [66, 24, 43, 18]. The analysis of observational studies (post-campaign) as an attribution tool has been proposed to approximate the causal effect of online ads and overcome the lack of a baseline in MTA. These approaches are categorized as: propensity-score-based approaches [22, 84], time-series-based inference [52, 8], and survival-based modeling [62]. Despite enabling the semi-automatic analysis of many campaigns, overall, observational studies are likely to over-estimate the campaign effectiveness [55]. In recognition of this, evaluators increasingly rely on randomized experiments to estimate campaign causal effects [27, 55, 87].

Previous work on online advertising evaluation based on randomized experiments shows the impact of ad impressions at a web portal on user search activity [55]. This design randomizes all user visits at ad serving time and does not consider the targeting engine.

The randomization of user visits limits the power of this randomized design to measure the effect on short-term observable signals, e.g. user search activity [55, 34]. The main drawback occurs when a given user is assigned to different treatment arms at various visits. For instance, the user conversion might take place at the advertiser website several hours after the multiple user visits to any of the publisher websites.

Current industry practice for campaign evaluation based on randomized experiments is to run a low-budget non-optimized CPM campaign and measure its effectiveness, which is assumed to hold for a larger budget and optimized CPA campaign [87, 27]. The experimental design of this practice randomizes users once and keeps them in the same group for the campaign duration. Also, this experimental framework runs a placebo campaign with a user targeting equivalent to that of the focal campaign. However, this design does not consider that the advertiser competes in a marketplace (ad exchange or ad network internal auction) to secure the publisher ad slot [24, 43]. This practice is standard within ad networks as well [18]. As a result, there is a selection effect due to the endogenous (non-ignorable) auction outcome and despite the exogenous (controllable) decision to target a given user. Therefore, the external validity of CPM campaign effects to CPA campaigns is prone to inaccuracies due to different user targeting incentives [14]. Also, the effect of the campaign presence in the marketplace, depicted by Fig 1.2, is not considered.

Test-control interference in marketplaces has been identified before when randomized users bid (demand) on scarce products (supply) [16]. The potential induced bias comes from spillover effects on the product demand and bidding between test-control user responses to the product supply. We note that in the marketplace of advertisers and publishers, where the experimental design randomizes online users, advertisers bid on publisher ad slots. This framework assumes enough product inventory (ad slots) to satisfy the demand (ads to be displayed). Thus, increasing ad slots demand will have a negligible effect on the user conversion probability.

In a recent economic literature review, Goldfarb (2014) surveys the online advertising literature based on the decreasing cost of user targeting [44]. Here, most of the literature on ad effectiveness based on field experiments evaluates focused targeting practices [54, 45, 57]. Similarly, the idea of a market target segment in campaign evaluation using randomized experiments has been addressed previously [34]. In practice, however, large ad networks target users based on complex user history and behavior with the objective of targeting converting users for CPA campaigns [1, 25]. Pandey *et al.* (2011) provide a comprehensive survey of "effective" targeting practices in terms of the standard objective of serving ads to the users who are more likely to convert [66]. In a theoretical analysis, Berman (2013) compares the user targeting of optimized CPA campaigns with the targeting of non-optimized CPM campaigns and suggests that CPA campaigns incentivize the targeting of *always-buy* users [14]. However, empirical evidence supporting this analysis has not been reported in the research or industrial literature. Overall, the comparison between CPA and CPM campaign performance based on their targeting has not been analyzed thoroughly.

The impact of online advertising on user clicks has been addressed mainly in the context of Search Advertising and with observational data [23, 60]. In online Display Advertising, Lewis and Reiley model the campaign effect on clicker conversions by analyzing observational data in a differences-in-differences approach, in spite of the availability of randomized data [57]. Dalessandro *et al.* (2012) perform an analysis to assess the effectiveness of optimizing user clicks in user targeting based on observational data. This analysis of observational data tends to report correlations between user clicks and a large amount of conversions not caused by the campaign [28]. Even when randomized experiments are performed, the user clicks are often discarded due to the lack of effective techniques to model them in the causal analysis [54].

In the causal inference literature, Potential Outcomes causal model analyzes the individual potential outcomes for each of the treatments [75]. For two treatment arms, this framework implies that half of the data is missing because we can never observe a unit response in both arms. In this causal model, Principal Stratification has been successfully employed to estimate the treatment effect when a selection bias is unavoidable in a randomized experiment [36]. The analysis of randomized experiments with noncompliance, where the randomly assigned individuals might opt out of the experiment due to treatment side effect [50, 51], is one of the most successful applications. Similarly, the analysis of right-censored data due to non-ignorable individual death [76, 37], and the education program assessment with truncated data due to student drop-out [88] are other problems addressed by Principal Stratification. A different approach when *intermediate* variables, such as user clicks, are observed in the "causal path" is to consider "causal mediation" or indirect effects [69, 68]. However, Rubin illustrates the risk of biased analysis when indirect effect modeling does not consider the potential outcomes adequately [74].

Other causal frameworks include the Structural Equation Model [68] and Econometric Causality [2, 48]. In this Dissertation, we approach the problem using Potential Outcomes to model post-treatment variables with the use of the experimental data. We note that the comparison of causal frameworks falls outside the scope of the current Dissertation focus, and we do not address it here thoroughly.

Part II

Observational Analysis: Campaign Evaluation at Scale

Chapter 3

Dynamic Conversion Attribution Based on Ad Impressions

3.1 Introduction and Problem Context

Recent research on campaign evaluation has focused primarily on two approaches: running randomized experiments (A/B testing), and bias correction based on user features for observational data. Based on A/B testing, a detailed method to compare the impact of user exposure to ad impressions is provided in [56]. Here, the authors verify the effect of ad exposure to users on their commercial actions. Lewis *et al.* (2011) also recommend running randomized experiments by addressing potential over-estimation issues due to user activity bias [55]. However, A/B testing presupposes that the experiment can be set up with the availability of tracking cookies to link ad impressions to the commercial actions. Besides, randomized experiments are often expensive to set up, and they require the display of placebo ads. These limitations prevent A/B testing from being deployed at scale for thousands of campaigns.

The authors of [22] propose to correct the bias in the user selection for ad exposure in observational studies. However, this approach is widely based on the user features that are often incomplete. Authors of [29] demonstrate the presence of long term and transient impact of marketing campaigns on sales. They suggest the concept of *persistence* or continuing memory as the actual impact. Since A/B testing and the correction in [22] do not consider the time lag between ad exposure and conversions, this continuing memory impact is not properly measured.

An essential requirement for the methods discussed is the use of reliable and stable tracking cookies. This requirement might not be highly relevant in scenarios where the ad network measures the signal of interest after a few minutes in the publisher website, as in the case of click through rates. However, user tracking is particularly important for commercial actions since the time between campaign exposure of a user and the final user conversion might be even days [61]. In practice, a large number of web users either reject tracking cookies outright or frequently delete such cookies. We have identified that approximately 17% of the *Advertising.com* users are not tracked based on cookies[1]. The best practice recommended is to discard those users [49]. Therefore, developing a time series approach, which incorporates persistence, and relates commercial actions to ad impressions shown to users with unreliable cookies, constitutes a significant contribution. Although there is extensive time series research in marketing [35, 67], we are not aware of any research to estimate dynamic campaign effects for online display advertising without user tracking.

[1] AOL Research and Development internal memo.

3.2 Chapter Contribution

We develop an approach to evaluate the impact of online marketing campaigns which differs from recent literature in three respects:

1. We focus on display advertising, where the actions are clearly commercial actions, rather than surrogates such as search terms or clicks.

2. We consider the context where user tracking information is not available.

3. The proposed method uses aggregate data and is simple to implement without expensive infrastructure.

We develop a Dynamic Linear Model (DLM) based method to model the impact of ad impressions on actions [8]. In the absence of user tracking information, we account for confounding effects by a base time series model. We rely on the prediction power of this base model to capture the effects of other factors on commercial actions. Unlike query terms or click through rates, user conversions are not immediately observed after user exposure, and these are performed in the advertiser website. Thus, estimating this relationship is challenging and provides the motivation for our work. Our contribution encompasses the following:

- A time series model to estimate the dynamic effect of ad impressions on commercial actions.

- A decay factor to model impression effects on actions which automatically provides different lags.

- The use of a base time series model of the actions to account for effects not attributed to campaigns.

- A fully automated approach to process outliers based on t-errors without setting a pre-determined threshold.

- A logarithmic model to account for the non-linear impact of ad impressions on daily actions.

3.3 Chapter Assumptions

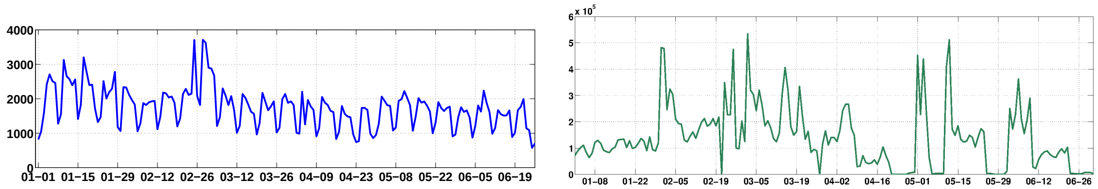
In this section, we state the main assumptions used in the current chapter. The results and conclusions of this chapter are valid to the extent that these assumptions are applicable.

1. A time series base model, which can be set by the analyzer, is assumed to account for the conversion time series evolution in the absence of advertising campaigns (counterfactual response).

2. The campaign aggregate dynamic effect on daily online conversions is assumed to be a transfer function response of the daily number of ad exposures (impressions).

3. An exponentially decaying campaign effect is assumed. The decay rate is fit automatically per product.

4. A dynamic evolving regression factor to account for the effect of the number of daily impressions is assumed.



(a)

(b)

Figure 3.1: (a) Commercial actions and (b) number of impressions through time. X-axis is the time in dates.

5. For multiple campaigns running at the same time, the campaign effects are modeled by additive (or superposition) DLMs.

6. A weekly seasonality effect, not attributable to online advertising campaigns, is assumed.

7. Model parameters are assumed to be random, with standard non-informative conjugate prior distributions given in Appendices 3.A and 3.B.

8. The prior state distributions of the DLMs developed in this Chapter are multivariate Normal distributions with zero mean and diagonal covariance matrix.

9. The above model components represent the structure we assumed in this Chapter. Fig 3.3 illustrates the dependencies between the random variables of this structure in a graphical model.

3.4 Methodology

This section is organized as follows: we define notation and the DLM. We illustrate the case of one campaign and discuss the insights of the model. Then, we generalize this model to multiple campaigns and show two possible base models to describe commercial actions when there are no active campaigns. We also present the log-transformation used in our approach and explain how we handle outliers based on t-errors. Finally, a Bayesian approach is detailed to fit the model using Gibbs sampling.

3.4.1 Notation and Definitions

Let T be the total number of days we observe commercial actions and impressions and N the number of advertising campaigns running during the observed period. The indices: t denotes discrete time in days, c indicates a particular campaign and k refers to the number of forecast look ahead steps. Let $X_t^{(c)}$ be the number of ad impressions at time t for campaign c, and Y_t be the number of actions at time t for a given product. $Y_{1:t}$ represents the vector of $[Y_1, \ldots, Y_t]$.

We define the evolution of a latent state $\theta_{\mathbf{t}}$ to be a stochastic process describing the underlying behavior of the series. Assuming Y_t is usually distributed conditional on the state $\theta_{\mathbf{t}}$, and $\theta_{\mathbf{t}}$ is Normally distributed conditional on the previous state $\theta_{\mathbf{t-1}}$, then we define:

$$\begin{aligned} Y_t &= F_t'\theta_t + \nu_t, \qquad &\nu_t \sim N(0, V_t), \\ \theta_t &= G_t\theta_{t-1} + w_t, &w_t \sim N(0, W_t), \end{aligned} \tag{3.1}$$

where G_t is the evolution, F_t is the observation matrix, ν_t is the observational noise with variance V_t, and w_t represents the state evolution with covariance matrix W_t. This is referred to as a Dynamic Linear Model (DLM) [85, 63]. Under this representation, we can update the distribution of $\theta_t|Y_{1:t}$ efficiently in closed form using Kalman Filtering equations [42, 71]. By choosing G_t and F_t, we can model different types of behavior of the time series.

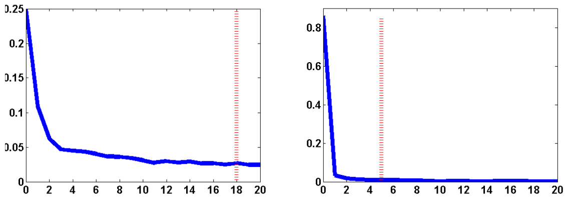


Figure 3.2: Time distribution of the last impression delivered to a user before performing a commercial action for two products. Dotted line represents 95% cumulative probability. x-axis is the number of days and y-axis the probability mass.

3.4.2 Modeling Campaigns

To model the effect of impressions on the number of commercial actions, we assume a persistence component of campaigns as suggested in previous work [29, 56]. This persistence modeling implies that impressions will affect the number of sales not only on the day they are displayed but also days after. To verify this effect in online commercial actions, we estimate the distribution of the number of days before the last impression that a user who performs a commercial action saw. We perform this analysis for two campaigns where user tracking is available. Fig 3.2 shows these distributions. If a user performs an action at day t, and he or she sees the last impression from a related campaign at day $t-k$, then k would be considered for this distribution. Note that for one product, just 0.25 of the probability mass is assigned to the same date of the action and up to 18 days for the 95% of cumulative probability.

We let ξ_t be the *effect* of ad impressions on the number of commercial actions at time t. We define the following:

$$\begin{aligned} Y_t = & \quad \xi_t + \nu_t, \\ \xi_t = & \quad \lambda \xi_{t-1} + \psi_t X_t + w_t^{(\xi)}, \\ \psi_t = & \quad \psi_{t-1} + w_t^{(\psi)}, \end{aligned} \tag{3.2}$$

where λ represents the constant rate of decay of this effect. ψ_t is the impact per impression which we allow to be dynamic over the duration of the campaign. We further constrain $\lambda \in [0, 0.88]$ to have a valid decay rate. This model can be expressed as a DLM as follows:

$$\begin{aligned} F' = [1,0], & \qquad \theta_t' = [\xi_t, \psi_t], \quad w_t' = [w_t^{(\xi)} + X_t w_t^{(\psi)}, w_t^{(\psi)}], \\ G_t = \begin{bmatrix} \lambda & X_t \\ 0 & 1 \end{bmatrix}, & \quad W_t = \begin{bmatrix} W_\xi + X_t^2 W_\psi & X_t W_\psi \\ X_t W_\psi & W_\psi \end{bmatrix}, V_t = V. \end{aligned} \tag{3.3}$$

Given the model parameters, $(\lambda, V, W_\xi, W_\psi)$ and a prior mean and variance for the latent state vector $\theta_1 \sim N(m_0, C_0)$, we estimate $(\theta_t | Y_{1:t}, X_{1:t})$ using Kalman Filtering equations.

The commercial actions evolution modeling of Eq 3.3 represents a transfer response function of the number of impressions X_t as defined by West and Harrison [85]. This model allows us to incorporate lagged effects of X_t into future values of Y_{t+r} at the state level. ξ_t represent these effects. On the other hand, standard dynamic regression typically integrates the aggregate effect at the observational level of the DLM [70, 71]. As a result, additional complexity in the evolution covariance matrix W_t is incorporated when compared to typical DLMs where this matrix is considered to be diagonal. Since the effect ξ_t is a function of the dynamic regression coefficient of impressions ψ_t and the previous effect ξ_{t-1}, a non-zero covariance term between ψ_t and ξ_t (equals to $X_t W_\psi$) is included in the state evolution. Similarly, the state variance of ξ_t depends on X_t and is equal to $W_\psi + X_t^2 W_\xi$.

We note that the model of Eq 3.2 does not include a bias or base model component. This limitation attributes *all* the actions to impressions which is not necessarily valid. We address this issue in section 3.4.3. By defining ξ_t, we automatically fit the number of previous days with impact on Y_t through λ unlike modeling a fixed lag in autoregressive models [29].

Typically, multiple campaigns run simultaneously for a given product. Here, we model each campaign independently and combine their effects linearly. Therefore, for N campaigns we have:

$$\begin{aligned} Y_t &= \sum_{c=1}^{N} \xi_t^{(c)} + \nu_t, \\ \xi_t^{(c)} &= \lambda^{(c)} \xi_{t-1}^{(c)} + \psi_t^{(c)} X_t^{(c)} + w_t^{(\xi,c)}, \\ \psi_t^{(c)} &= \psi_{t-1}^{(c)} + w_t^{(\psi,c)}. \end{aligned} \tag{3.4}$$

We incorporate an independent rate of decay, $\lambda^{(c)}$, and a dynamic impression to action conversion coefficient, $\psi_t^{(c)}$, for each campaign. In a similar manner to the analysis of one campaign, we re-write the model from Eq 3.4 as a DLM, based on the definition of Eq 3.1, as follows:

$$\begin{gathered} \theta_t'^{(c)} = [\xi_t^{(c)}, \psi_t^{(c)}], \qquad w_t'^{(c)} = [w_t^{(\xi,c)} + X_t^{(c)} w_t^{(\psi,c)}, w_t^{(\psi,c)}], \\ W_t^{(c)} = \begin{bmatrix} W_\xi^{(c)} + \left(X_t^{(c)}\right)^2 W_\psi^{(c)} & X_t^{(c)} W_\psi^{(c)} \\ X_t^{(c)} W_\psi^{(c)} & W_\psi^{(c)} \end{bmatrix}, \\ G_t^{(c)} = \begin{bmatrix} \lambda^{(c)} & X_t^{(c)} \\ 0 & 1 \end{bmatrix}, \qquad F'^{(1:N)} = [(1,0)^{(1)}, \cdots, (1,0)^{(N)}], \\ \theta_t'^{(1:N)} = [\theta_t'^{(1)}, \cdots, \theta_t'^{(N)}], \qquad G_t^{(1:N)} = \mathrm{diag}[G_t^{(1)}, \cdots, G_t^{(N)}], \\ w_t'^{(1:N)} = [w_t'^{(1)}, \cdots, w_t'^{(N)}], \qquad W_t^{(1:N)} = \mathrm{diag}[W_t^{(1)}, \cdots, W_t^{(N)}]. \end{gathered} \tag{3.5}$$

Finally, we define $M^{(1:N)}$ as the DLM model where we include all campaigns $1, \ldots, N$ in the analysis as follows:

$$M^{(1:N)} = \text{DLM}\left(F^{(1:N)}, G_t^{(1:N)}, V^{(1:N)}, W_t^{(1:N)}, \theta_t^{(1:N)}, w_t^{(1:N)}\right). \tag{3.6}$$

3.4.3 Base Model Definition

As discussed previously, model $M^{(1:N)}$ incorporates the effects of all N campaigns on the number of actions. However, there is no base model to describe the number of commercial transactions when there are no impressions. Also, the daily action time series provides prediction capability due to time dependencies of the observations. These aspects are highly relevant in separating the actions attributed to external factors from those we attribute to campaigns. Let $M^{(0)}$ be a base model with no campaign contributions. Thus, we define the full model $M^{(0:N)}$ as follows:

$$\begin{aligned} Y_t &= \widetilde{F}'\widetilde{\theta}_t + \widetilde{\nu}_t, & \widetilde{\nu}_t &\sim N(0, \widetilde{V}), \\ \widetilde{\theta}_t &= \widetilde{G}_t\widetilde{\theta}_{t-1} + \widetilde{w}_t, & \widetilde{w}_t &\sim N(0, \widetilde{W}_t), \end{aligned} \tag{3.7}$$

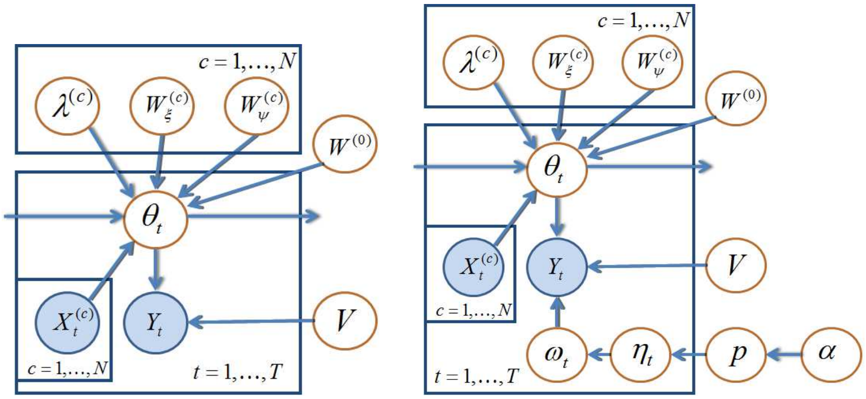
where:

$$\begin{aligned} \widetilde{F}' &= [F'^{(0)}, F'^{(1:N)}], & \widetilde{\theta}'_t &= [\theta_t'^{(0)}, \theta_t'^{(1:N)}], \\ \widetilde{w}'_t &= [w_t'^{(0)}, w_t'^{(1:N)}], & \widetilde{W}_t &= \text{diag}[W^{(0)}, W_t^{(1:N)}], \\ & & \widetilde{G}_t = \text{diag}[G^{(0)}, G_t^{(1:N)}], \end{aligned} \tag{3.8}$$

leading to:

$$M^{(0:N)} = \text{DLM}\left(\widetilde{F}, \widetilde{G}_t, \widetilde{V}, \widetilde{W}_t, \widetilde{\theta}_t, \widetilde{w}_t\right). \tag{3.9}$$

A key advantage of the model $M^{(0:M)}$ from Eq 3.9 is that it allows us to incorporate *any* DLM to attribute the time series of the actions. Thus, we can include any assumption about the dynamics of the actions and model the remaining variability of the data by the impressions. Fig 3.3(a) shows the graphical model for the model $M^{(0:N)}$. For this study,



(a) (b)

Figure 3.3: (a) Graphical model for multiple campaigns and a base model, $M^{(0:N)}$. (b) Model for multiple campaigns with outliers processing $M\omega^{(0:M)}$.

we test two base models for the time series: a random walk and a weekly seasonal model.

In traditional linear regression, a bias component is used to place the predictive variable around its expected value. Here, we use a dynamic bias, also known as the random walk, as the standard base model choice [70]. Thus, we define $M_b^{(0)}$:

$$\begin{aligned} Y_t &= \theta_{b,t}^{(0)} + \nu_{b,t}^{(0)}, & \nu_t &\sim N(0, V_b^{(0)}), \\ \theta_{b,t}^{(0)} &= \theta_{b,t-1}^{(0)} + w_{b,t}^{(0)}, & w_{b,t}^{(0)} &\sim N(0, W_b^{(0)}). \end{aligned} \tag{3.10}$$

In Fig 3.1, we observe a seasonal component of the number of actions that synchronizes with the day of the week [11]. Therefore, we define a base model to incorporate this component, based on the assumption that there are commercial actions produced simply because of the day of the week. To model this seasonality, we use the Fourier representation of DLMs [85, 71]. We set the seasonal period to 7, $\omega = 2\pi/7$, and define two harmonics to model this frequency as follows:

$$\begin{aligned}
F_s^{(0)} &= [1, (1,0)_\omega, (1,0)_{2\omega}], \quad \theta_{s,t}^{(0)} = [\theta_{b,t}^{(0)}, \theta_{\omega,t}, \theta_{\omega,t}^*, \theta_{2\omega,t}, \theta_{2\omega,t}^*], \\
G_\omega &= \begin{bmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{bmatrix}, \qquad G_s^{(0)} = \mathrm{diag}[1, G_\omega, G_{2\omega}], \\
w_{s,t}^{(0)} &= [w_{b,t}^{(0)}, w_{\omega,t}, w_{\omega,t}^*, w_{2\omega,t}, w_{2\omega,t}^*], \\
W_s^{(0)} &= \mathrm{diag}[W_b^{(0)}, W_\omega, W_\omega^*, W_{2\omega}, W_{2\omega}^*].
\end{aligned} \tag{3.11}$$

We refer to this model as $M_s^{(0)}$.

3.4.4 Log-Transformation

To relax the assumption of a linear impact of ad impressions, X_t, on actions, Y_t, we use the log transformation for both variables. Assuming $\xi_t = 0$ for illustrative purposes, we consider the following model for one campaign:

$$\begin{aligned}
Z_t &= \log(Y_t), & X_t^* &= \log(X_t), \\
Z_t &= \theta_t^{(0)} + \psi_t X_t^* + \widetilde{\nu}_t, & Y_t &= \exp\left\{\theta_t^{(0)} + \widetilde{\nu}_t\right\} X_t^{\psi_t}.
\end{aligned} \tag{3.12}$$

This model is multiplicative for Y_t. If $\psi_t < 1$, the effect of ad impressions on actions decreases as more of them are shown. As shown in Fig 3.6, the daily number of impressions is more dynamically changing than the number of commercial actions. By assuming this model, we smooth these changes of the time series of impressions. This smoothing supports the intuition of steady contributions through time.

Since the logarithm is a monotonic function $f(Y_t)$, the median of Y_t is the same as the inverse transform of the median of Z_t, $f^{-1}(Z_t)$. In general, the cumulative distribution for Y_t is the same as the cumulative distribution of the inverse transform of the Z_t, $cdf(Y_t) =$

Algorithm 1 Generative Model to Handling Outliers

```
Draw $p|\alpha \sim Dir(\alpha)$
for $t \leftarrow 1$ to $T$ do
    Draw $\eta_t|p \sim Mult(1, p)$
    Draw $\omega_t|\eta_t \sim \Gamma(\frac{\eta_t}{2}, \frac{\eta_t}{2})$
    Set $V_t^* = \omega_t^{-1} V$
end for
```

cdf$(f(Y_t))$. This is a crucial property for model fitting, described in section 3.4.6, as we simply transform $X_{1:T}, Y_{1:T}$ to the new space $X^*_{1:T}, Z_{1:T}$ for all campaigns and model $Z_{1:T}$ with $M^{(0:N)}$ as in section 3.4.3.

3.4.5 Handling Outliers in the Model

Given that advertisers collect the commercial action data, there is no control over this process by the Ad network [61]. As a consequence, very often we observe outliers or *drastic* changes in the daily number of actions. These sudden changes could be very problematic because they move our estimates and consequently increase their variance.

One approach to handling outliers is to give weights to the observations based on the variance modeled for each output Y_t [85]. For this analysis, we use a simplification of the model presented in [70]. We switch from using the Normal distribution for Y_t to using a t-distribution with a set of degrees of freedom to choose from. Those degrees of freedom range from highly long-tailed distributions (small number of degrees of freedom) to approximately Normal distributions (a large number of degrees of freedom). We use the Normal-Gamma mixture to represent this t-distribution [41].

Fig 3.3(b) shows the graphical model for all campaigns with outliers handling. In this model, all changes are introduced at the observation level with a hierarchical model to handle an individual number of degrees of freedom for each Y_t. Algorithm 1 shows the generative model for the observational variance when we process the outliers. As illustrated, we now have an independent variance for each observation. This variance is the product of a time-invariant V and ω_t that we draw from a Gamma distribution with η_t degrees of freedom. Conditional on $\omega_{1:T}$, we have a DLM model $M^{(0:M)}$ with $V_t^* = \omega_t^{-1} V$ for $t = 1, \ldots, T$. The possible values for η_t are predefined a priori. We fix these values to be the set $\{1, 2, \ldots, 10, 20, \ldots, 50\}$ whose cardinality equals the dimension of the multinomial and its Dirichlet prior.

3.4.6 Inferring the Model Parameters

In this section, we provide the details of the model fitting. We assume all parameters to be random variables and follow a Bayesian approach. We have a set of static variables (not indexed by time), and dynamic variables (stochastic processes). As observations, we have the daily number of actions for each product, and impressions for each campaign. Since we obtain the conditional posterior distribution of each random variable in closed form, we follow a Gibbs sampling approach. This method provides a set of samples used to estimate the statistics of interest. Algorithm 2 defines the static and dynamic variables in addition to the observations and illustrates the sampling procedure we develop. We refer to the latent variables, $\theta_{1:T}$, in the DLM as hidden states in this section.

Algorithm 2 Gibbs Sampling Algorithm

```
Define $D_{1:T} = \left\{Y_{1:T}, X_{1:T}^{(1:N)}\right\}$
Initial guess $\Phi^0 = \left\{\lambda^{(1:N)}, W_{\psi}^{(1:N)}, W_{\xi}^{(1:N)}, W^{(0)}, \widetilde{V}\right\}^0$
Initial guess $\Omega^0 = \{\omega_{1:T}, \eta_{1:T}, p\}^0$
for $s \leftarrow 1$ to $N_0 + N_s$ do
    Draw $\theta_{1:T}^s \sim p\left(\theta_{1:T}|\Phi^{s-1}, \Omega^{s-1}, D_{1:T}\right)$ using FFBS
    Draw $\Phi^s \sim p\left(\Phi|\theta_{1:T}^s, \Omega^{s-1}, D_{1:T}\right)$ using Eqs from Appendix 3.A.
    Draw $\Omega^s \sim p\left(\Omega|\theta_{1:T}^s, \Phi^s D_{1:T}\right)$ using Eqs from Appendix 3.B.
end for
```

Gibbs Sampling

The main focus to fit the model is to estimate the posterior distribution of the static variables and hidden states given the observed data. We define:

$$\begin{aligned} D_{1:T} &= \left\{Y_{1:T}, X_{1:T}^{(1:N)}\right\}, \quad \Omega = \{\omega_{1:T}\eta_{1:T}, p\}, \\ \Phi &= \left\{\lambda^{(1:N)}, W_{\psi}^{(1:N)}, W_{\xi}^{(1:N)}, W^{(0)}, \widetilde{V}\right\}. \end{aligned} \tag{3.13}$$

We sample $\theta_{1:T}|\Phi, \Omega, D_{1:T}$ based on Forward Filtering Backward Sampling (FFBS) method explained below. We provide the distributions to sample from $\Phi|\theta_{1:T}, \Omega, D_{1:T}$ and $\Omega|\theta_{1:T}, \Phi, D_{1:T}$, in Appendices 3.A and 3.B respectively.

To process outliers, we sample $\Omega|\theta_{1:T}, \Phi, D_{1:T}$ based on the generative model detailed in Algorithm 1. Here, conjugate priors are set for ω_t, p and a non-conjugate prior for η_t. Nonetheless, η_t is a discrete random variable and can take only a small number of values. This constraint on the number of possible values facilitates the estimation of the normalization constant for $\eta_t|\omega_t, D_{1:T}$ to estimate their posterior distribution. By fixing

$\omega_t = 1$ for $t = 1, \ldots, T$ in all Gibbs iterations, we obtain the model $M^{(0:N)}$ without handling outliers.

FFBS is a method to sample the hidden states θ_t conditional on the static variables or parameters in a DLM [70, 20]. This approach an alternative to sample from the joint random vector $\theta_{1:T}, \Phi | D_{1:T}$. In this framework, depicted in Fig 3.4, samples are generated backward after filtering conditional on the already generated states as follows:

1. Estimate $p(\theta_t | \Phi, D_{1:t}) = N(m_t, C_t)$ for $t = 1, \ldots, T$ using Kalman Filtering equations

2. Draw $\theta_T | D_{1:T} \sim N(m_T, C_T)$

3. For $t = T-1, \ldots, 1$ draw $\theta_t | \theta_{t+1}, D_{1:T} \sim N(h_t, H_t)$

$$\begin{aligned}
h_t &= m_t + C_t G'_{t+1} R^{-1}_{t+1} (\theta_{t+1} - a_{t+1}), \\
H_t &= C_t - C_t G'_{t+1} R^{-1}_{t+1} G_{t+1} C_t, \\
R_{t+1} &= G_{t+1} C_t G'_{t+1} + W_t, \\
a_{t+1} &= G_{t+1} m_t.
\end{aligned} \tag{3.14}$$

The initial state distribution is assumed to be approximately a non-informative prior distribution: $\theta_0 \sim N(\mathbf{0},\sigma_0^2\mathbf{I})$, where σ_0^2 =1e-12, $\mathbf{0}$ is a vector of zeros, and $\mathbf{I}$ represents the identity matrix.

In general, by using FFBS we estimate the distribution of the hidden states given the observations up to time t. Then, we sample the state variables at time T, and conditional on this value, we sample the state variables at time $T-1$ (backward). This procedure provides a randomly generated sample given the observations up to time T.

Often a question by advertisers is what percentage of actions can be *attributed* to campaigns. To address this question, we use an alternative model in which the campaign

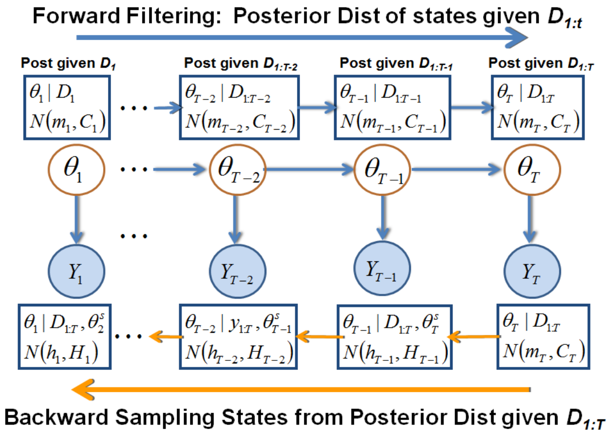

Forward Filtering: Posterior Dist of states given $D_{1:t}$
Post given D_1
Post given $D_{1:T-2}$
Post given $D_{1:T-1}$
Post given $D_{1:T}$
$\theta_1 \mid D_1$ $N(m_1, C_1)$
$\theta_{T-2} \mid D_{1:T-2}$ $N(m_{T-2}, C_{T-2})$
$\theta_{T-1} \mid D_{1:T-1}$ $N(m_{T-1}, C_{T-1})$
$\theta_T \mid D_{1:T}$ $N(m_T, C_T)$
θ_1
θ_{T-2}
θ_{T-1}
θ_T
Y_1
Y_{T-2}
Y_{T-1}
Y_T
$\theta_1 \mid D_{1:T}, \theta_2^s$ $N(h_1, H_1)$
$\theta_{T-2} \mid y_{1:T}, \theta_{T-1}^s$ $N(h_{T-2}, H_{T-2})$
$\theta_{T-1} \mid D_{1:T}, \theta_T^s$ $N(h_{T-1}, H_{T-1})$
$\theta_T \mid D_{1:T}$ $N(m_T, C_T)$
Backward Sampling States from Posterior Dist given $D_{1:T}$

Figure 3.4: FFBS representation. In forward filtering, $p(\theta_t|D_{1:t})$ for t=1, . . . , T are estimated. In backward sampling, $\theta^s_{1:T}|D_{1:T}$ are sampled recursively from $p(\theta_t|D_{1:T}, \theta_{t+1})$ for t=$T-1, \dots, 1$.

effects are constrained to be positive $\psi_t, \xi_t > 0$ in addition to the base model trend [11]. We constrain these components at the time we sample them in the Gibbs iterations. We use the approach from [73] to draw from a constrained multivariate Normal distribution.

Posterior Random Sample

Given a set of samples of the posterior distribution for the variables, $\theta_{1:T}, \Phi, \Omega$, we use the log-transformation and obtain their empirical distribution. As shown in section 3.4.4, we estimate $Z_{1:T}$ and $X^*_{1:T}$ from Eq 3.12 for all campaigns. Then, we draw a posterior distribution sample of $Z_{1:T}$ that we inversely transform to Y_t. Thus, no further changes are necessary to implement the model described.

To evaluate the model fitting, we use the median and 90% credible intervals. We also evaluate model prediction based on one-step ahead forecasting, $Y_t^{k=1}|D_{t-1}$, estimated using the samples we generate in FFBS at the filtering stage. Thus, we define:

$$\begin{aligned} &\hat{Y}_t|M^{(0:N)} \approx \text{Median}(F'\theta_t^s), \quad \hat{Y}_t|M_{\text{log}}^{(0:N)} \approx \text{Median}(\exp(F'\theta_t^s)), \\ &\hat{\omega}_t|_{\omega}M^{(0:N)} \approx \text{Median}(\omega_t^s), \quad \hat{Y}_t^{k=1}|M^{(0:N)} \approx \text{Median}(Y_t^{k=1,s}), \end{aligned} \tag{3.15}$$

for $s = 1, \ldots, N_s$ Gibbs samples. These estimates are used in section 3.5 and 3.6.1 to evaluate campaigns and model fitting.

3.4.7 Implementation Details

Numerical issues can cause a naive application of the Kalman filter to produce matrices $C_{1:T}$ that cannot be legitimate posterior covariance matrices, because they are not positive definite. This is also a problem in FFBS from Eq 3.14 as we could potentially get negative semi-definite matrices and singular matrices R_{t+1}^{-1} [70]. To overcome these issues, we use singular value decomposition (SVD) based approaches from [83] for filtering and [89] for backward sampling. During forward filtering, when a campaign becomes active we augment the state with the new campaign. Then, we set a large prior variance, and zero mean for this new campaign's state as if it were at the beginning of the time series. In the transition from active to inactive, we take the marginal distribution of the remaining active campaigns and base model by discarding the inactive campaign covariances.

3.5 Campaign Evaluation

In the absence of user tracking information, we account for confounding effects in campaign evaluation by a base time series model. This time series model accounts for commercial action attribution when campaigns are not active (control). In this framework, we present two campaign evaluation approaches. We interpret the model and evaluate the campaign performance dynamically. Also, we use the variability attributed to a campaign as a whole. This attributed variability is based on the dynamic model fitting with and without the campaign of interest.

To interpret the model for campaign evaluation, we estimate the proportion of actions attributed to a campaign as follows:

$$\begin{aligned} &Y_t^{(c)s}|M^{(0:N)} = F'\theta_t^s - F'^{\neg c}\theta_t^{\neg cs}, \quad \pi_t^{(c)s} = Y_t^{(c)s}/Y_t, \\ &Y_t^{(c)s}|M\log^{(0:N)} = \exp(F'\theta_t^s) - \exp(F'^{\neg c}\theta_t^{\neg cs}). \end{aligned} \tag{3.16}$$

Here, $\theta^{\neg cs}$ are the posterior state samples attributed to the base model and all campaigns except campaign c. For $M^{(0:N)}$ this is the same as $Y_t^{(c)s} = \xi_t^{(c)s}$. Nonetheless, for log-transform model $M\log^{(0:N)}$ this is not the case. This measure provides the expected difference attributed to campaign c, respect to other campaigns and confounding effects. Here, we have daily samples of the campaign effects $\pi_t^{(c)s}$.

We also estimate the variability attributed to a given campaign, $\mathrm{R}^2(c)$, by fitting the model without campaign c and calculating the difference with the full model. This approach measures the improvement in model fitting by campaign impressions. We estimate this variability attribution with respect to the data variance, the base model error variance, and error variance of the model with all campaigns except this campaign c as follows:

$$\begin{aligned} &\mathrm{R}^2(c|\mathrm{var}(Y_t)) = \frac{\mathrm{MSE}\big(M^{(0:N\neg c)}\big) - \mathrm{MSE}\big(M^{(0:N)}\big)}{\mathrm{var}(Y_t)}, \\ &\mathrm{R}^2(c|M^{(0)}) = \frac{\mathrm{MSE}\big(M^{(0:N\neg c)}\big) - \mathrm{MSE}\big(M^{(0:N)}\big)}{\mathrm{MSE}\big(M^{(0)}\big)}, \\ &\mathrm{R}^2(c|M^{(0:N\neg c)}) = 1 - \frac{\mathrm{MSE}\big(M^{(0:N)}\big)}{\mathrm{MSE}\big(M^{(0:N\neg c)}\big)}. \end{aligned} \tag{3.17}$$

Here, MSE stands for the mean squared error. These metrics estimate the variability attribution of campaign c given that all other campaigns are present with respect to: the data variance $var(Y)$, the base model $M^{(0)}$, and full model without campaign c, $M^{(0:N\neg c)}$. For these measures, we do *not* process outliers since this process *weights* the observations based on how they deviate from the others. Given that the model is run multiple times to estimate MSE for $M^{(0)}$ and $M^{(0:N\neg c)}$ these outliers are estimated differently.

3.6 Validation and Results

In this section, we describe the model fitting evaluation metric, data description and settings, and experimental results. We compare several versions of our model and discuss each of its components. Also, we compare our results with A/B testing. This comparison with results from randomized experiments is the standard approach used to evaluate the effects of marketing campaigns when controlled experiments can be run, and users can be tracked. Finally, we validate our model with a public synthetic dataset.

3.6.1 Model Fitting Evaluation

To measure model fitting, we use the mean squared error (MSE) per product time series. This approach provides a distribution of product MSEs. Then, we take the average MSE over products for model selection. However, since the product conversions are different in nature, ranging from email subscriptions to actual economic transactions, the time series of commercial actions have different central level and variability for various advertisers. Thus, we evaluate the model fitting base on the mean relative squared error

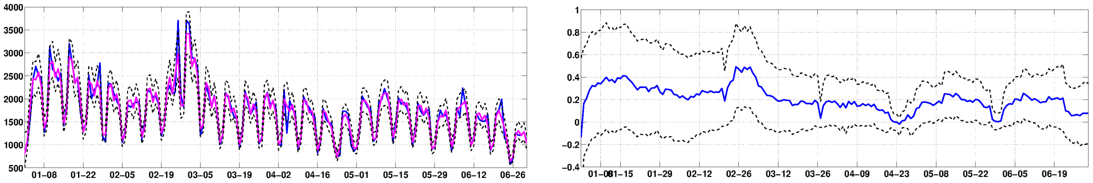
(MRSE). Given the model M with fitted $\hat{Y}_t$ and $\hat{\omega}_t$ we have for each product in the dataset:

$$\begin{aligned} &\text{MRSE}^f\left(M\right) = \tfrac{1}{T}\textstyle\sum_t \hat{\omega}_t|M\left(\frac{Y_t-\hat{Y}_t|M}{Y_t}\right)^2, \quad Y_t > 0, \\ &\text{MRSE}^f_{\omega=1}\left(M\right) = \tfrac{1}{T}\textstyle\sum_t \left(\frac{Y_t-\hat{Y}_t|M}{Y_t}\right)^2. \end{aligned} \tag{3.18}$$

This measure represents the mean squared error proportion relative to the observed Y_t. As a result, MRSE is normalized across different scales of the number of daily actions and eliminates any bias to high volume conversion series in the fitting evaluation. We provide these two evaluations given that our model processes outliers. If some outliers produce large errors for $\text{MRSE}_{\omega=1}$, then we might under-estimate the model fitting even though this is accounting for these sudden changes correctly. To evaluate model prediction, we use one-step-ahead forecast estimates $\hat{Y}_t^{k=1}$ to calculate MRSE, denoted as $\text{MRSE}^{k=1}$.

3.6.2 Data Description and Settings

For this study, we analyze all the transactions for $2,885$ marketing campaigns associated with $1,251$ products during six months, from January 1^{st} to June 31^{st}, 2011. We aggregate the daily actions and impressions by products and by campaigns. In general, a campaign is associated with a product when this is set up. We use these associations to relate actions with impressions. These relations suggest multiple campaigns targeting actions for the same product, and campaigns targeting actions for multiple products. In these experiments, we measure the performance of a campaign for each product independently. We define a 30-second threshold as the minimum time between two actions for the same user. This constraint prevents multiple clicks for a single action from being counted several times. We use $N_0 = 1000$ samples for burn-in and $N_s = 4000$ for the posterior distribution.



(a)

(b)

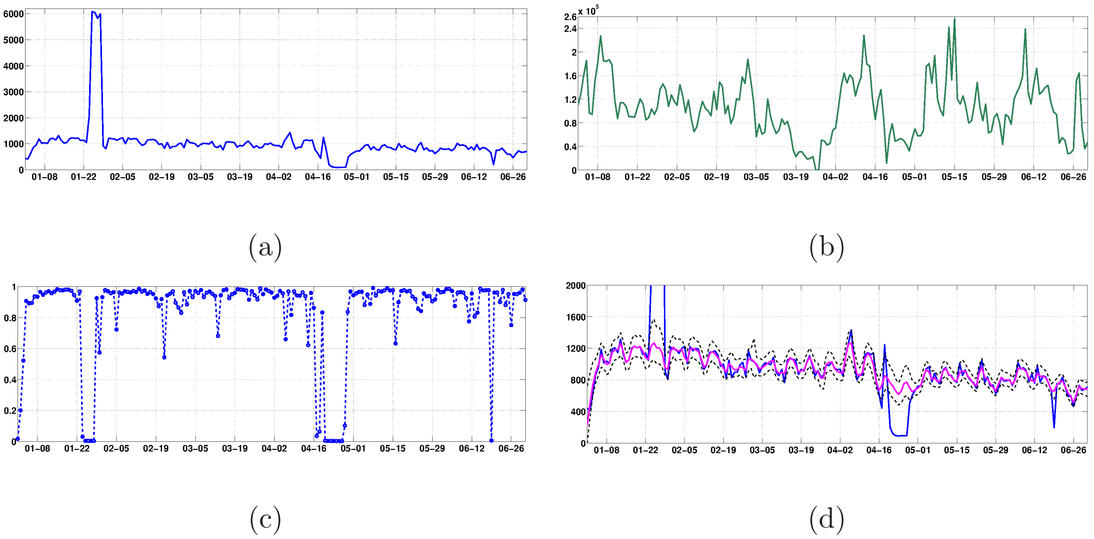
Figure 3.5: (a) Model fitting and (b) proportion of actions attributed to campaign in Fig 3.1. X-axis is time in dates.

3.6.3 Experimental Results

We are not aware of other approaches in the literature to estimate the campaign effects of online display advertising without user tracking. Thus, in this section we compare different variants of our model, based on model fitting and prediction. We evaluate the base model power by testing two models: random walk, and weekly seasonal model. The impact of the log transformation is evaluated based on fitting and prediction power. We also show qualitative results to analyze the model performance and the outlier handling. To illustrate the power of campaign impressions, we evaluate the fitting improvement provided by the incorporation of each campaign using R^2 as discussed above in section 3.5.

Qualitative Results

Fig 3.5 shows the model fitting and proportion of actions attributed to ad impressions shown in Fig 3.1. For this example, we use the model $M_s\omega\log$ (we process outliers, use the log transformation, and assume the weekly seasonal base model). As shown, the seasonal base model seems to be a good choice given the evident weekly periodicities in commercial actions. We observe that the method attributes the peak in daily actions during



(a) (b)

(c) (d)

Figure 3.6: (a) Commercial actions. (b) Ad impressions. (c) Median weights fitted for outliers. (d) Model fitting with 90% credible intervals. X-axis is the time in dates.

Table 3.1: Model variants used for experimentation.

Model	Process Outliers	Aggregate Camp	Log Transform	Positively Constrained
$M\omega$	X			
$M\omega$Agg	X	X		
$M\omega$+	X			X
$M\omega$+Agg	X	X		X
$M\omega\,$log	X		X	
$M\omega\,$logAgg	X	X	X	
$M\omega\,$log+	X		X	X
$M\omega\,$log+Agg	X	X	X	X

Table 3.2: Model evaluation results averaged over products. MRSE from Eq 3.18 is used for: fitted errors MRSE^f, one step forecast errors $\text{MRSE}^{k=1}$, non-weighted ($\omega_t = 1, t = 1 : T$) fitted and one step forecast errors $\text{MRSE}^f_{\omega=1}$, $\text{MRSE}^{k=1}_{\omega=1}$. 95% credible intervals are shown. Estimates scaled by 10^{-2}.

	Random Walk Base model $M_b^{(0)}$, MRSE			
Model	Fitted	Forecast	Fitted $\omega_t = 1$	Forecast $\omega_t = 1$
$M\omega$	7.91 ± 1.85	61.77 ± 7.13	14.87 ± 2.40	72.13 ± 7.58
$M\omega$Agg	9.79 ± 2.73	57.65 ± 8.75	16.63 ± 3.60	68.21 ± 9.31
$M\omega$+	15.78 ± 3.15	59.99 ± 8.06	21.79 ± 3.56	65.97 ± 7.89
$M\omega$+Agg	14.86 ± 2.74	53.41 ± 6.53	21.44 ± 3.15	62.03 ± 6.55
$M\omega$ log	1.33 ± 0.32	**13.25±2.21**	5.49 ± 1.02	20.00 ± 2.57
$M\omega$ logAgg	1.48 ± 0.32	**11.76±2.42**	5.75 ± 1.03	18.52 ± 2.76
$M\omega$ log+	1.87 ± 0.37	13.87 ± 2.77	5.70 ± 1.15	19.39 ± 3.08
$M\omega$ log+Agg	1.84 ± 0.42	12.70 ± 2.87	7.33 ± 1.52	21.10 ± 3.61
	Weekly Seasonal Base model $M_s^{(0)}$, MRSE			
$M\omega$	8.26 ± 2.13	61.13 ± 7.34	12.65 ± 2.11	70.75 ± 7.67
$M\omega$Agg	8.06 ± 2.00	58.56 ± 7.05	13.40 ± 2.47	71.47 ± 7.87
$M\omega$+	13.69 ± 3.23	62.11 ± 7.85	16.34 ± 3.10	68.55 ± 8.11
$M\omega$+Agg	10.32 ± 2.15	59.39 ± 7.06	14.18 ± 2.47	66.14 ± 7.33
$M\omega$ log	**0.72±0.14**	15.51 ± 2.81	**3.92±0.92**	21.20 ± 3.12
$M\omega$ logAgg	**0.64±0.13**	12.45 ± 2.20	**4.43±1.05**	19.14 ± 2.83
$M\omega$ log+	1.48 ± 0.27	15.07 ± 2.44	4.22 ± 0.95	19.39 ± 2.79
$M\omega$ log+Agg	1.20 ± 0.22	15.31 ± 2.43	**3.24±0.74**	19.16 ± 2.81

the first half of the time series to a gradual daily increase in the number of impressions.

Fig 3.6 shows the conversions and impressions series with the model fitting for other product. Similarly to Fig 3.5 and 3.1, there is a clear weekly seasonal component in the action series. In contrast to this previous instance, outliers are evident. The median of the posterior distribution for weights, $\hat{\omega}_t$, is also shown. The observation Y_t is likely to be an outlier when $\hat{\omega}_t$ is low. If $\hat{\omega}_t$ is zero, Y_t is treated as a missing value. Thus, if we observe sudden changes in the action series and these are not present in the impression series, the outlier handling *weights* them automatically.

Quantitative Results

To provide a review of the model, we test different variants and discuss the benefits of each component. Table 3.1 shows the various variants of the model we run. We incorporate the processing of outliers to interpret the elements of the model and estimate the actions attributed to impressions. Model MAgg stands for the model where we aggregate the impressions from all campaigns associated with a given product. Our goal is two-fold: 1)Modeling the impact of the set of campaigns as a whole on actions. 2)Hierarchically disaggregating campaigns to evaluate the effects of each one in detailed. In practice, these models are deployed in sequence.

Table 3.2 shows the results for the models based on fitting and prediction. Forecast errors for positively constrained models $M+$ should not be interpreted in the same manner as the other models because we estimate these errors in the filtering step of FFBS. In general, performance estimates based on MRSE instead of MRSE$_{\omega=1}$ are more robust because they diminish the effect of the outliers on the performance measure. When comparing the model performance for the aggregated campaigns versus modeling them separately, we observe comparable performance. We observe better performance when we include the log

Table 3.3: Averaged campaign evaluation results $\bar{\pi}^{(c)}$. Distribution of campaign effect with positive and negative 90% credible intervals (±), positive intervals (+), and negative intervals (-).

Model	% attributed	Campaign Effect Interval Sign		
		(+)	(-)	(±)
Random Walk Base model				
$M\omega$	14.07 ± 1.36	23.13	0.71	76.09
$M\omega$Agg	19.07 ± 2.63	40.96	2.01	57.03
$M\omega$ log	21.31 ± 1.63	18.65	0.58	**80.71**
$M\omega$ logAgg	24.75 ± 2.40	34.44	1.31	**64.25**
Weekly Seasonal Base model				
$M\omega$	10.39 ± 1.23	19.66	1.31	78.96
$M\omega$Agg	16.10 ± 1.91	34.33	1.56	64.11
$M\omega$ log	19.84 ± 1.64	14.83	0.60	**83.98**
$M\omega$ logAgg	21.09 ± 2.28	25.24	0.36	**73.18**

transformation consistently. This performance result clearly suggests that the relationship between actions and impressions is not linear. Comparing the two base models analyzed, we observe that weekly seasonal base model results show better fitted MRSE than the random walk. However, the predictive power of both models is equivalent. Based on fitted MRSE, the performance of the constrained models is lower than without this constraint consistently. We expect this performance since the imposed constraint might not be optimal for the MRSE measure.

Table 3.4: Attributed variability results.

	RandWalk $M_b^{(0)}$		WeekSea $M_s^{(0)}$	
Measure	Mean	Std Dev	Mean	Std Dev
$\mathrm{R}^2(c\vert\mathrm{var}(Y_t))$	**0.1241**	0.2704	**0.0667**	0.1750
$\mathrm{R}^2(c\vert M^{(0)})$	0.2804	0.3827	0.3002	0.3701
$\mathrm{R}^2(c\vert M^{(0:N\neg c)})$	0.4967	0.4114	0.4703	0.3729

Table 3.3 shows the campaign performance based on the models fitted. We observe that the models with the best fitting show the largest percentage of campaigns with positive and negative values in the 90% credible interval of the impressions effect on actions. These results illustrate the challenge of discarding the zero effect of advertising on online conversions, when compared to the effect on other signals as online search activity [55]. Table 3.4 depicts the mean and variance over all the campaigns, for the variability attribution based on R^2 (Eq 3.17). This attribution metric represents the fitting power improvement provided by the incorporation of campaign c. We observe that the variability attribution with respect to the base model residuals $\mathrm{R}^2(c|M^{(0)})$, and the other campaigns residuals, $\mathrm{R}^2(c|M^{(0:N\neg c)})$, are of similar values. However, the variability attribution with respect to the variability in the data is lower for the weekly seasonal base model. This variability lower attribution is because, under the seasonal base model, the campaign effect varies based on the day of the week.

3.6.4 Comparison with A/B Testing

One of the motivations for a dynamic model is to reduce the impact of time as a confounder on campaign evaluation. In this context, running a randomized experiment to test the effects of a campaign (A/B testing) has been suggested recently in the context of online advertising [55]. In this section, we compare our results, assuming no user tracking, with A/B testing for two campaigns in the UK during 27 days. These are independent campaigns run for independent products. For campaign 1, we use daily action data from 5 previous months, in addition to the duration of the campaign, in the estimation. For campaign 2, we only use the 27 days of action and impression data. Unlike the experiments presented in [55], we choose the users based on a targeting algorithm (considered to be a black box). We then randomly decide whether or not to expose the user to the ad impression to measure the effect on the targeted population.

We count the users who are first exposed to the campaign and later perform the action of interest (exposed and actor, N_{Exp}^{Act}) or not (exposed and non-actor, $N_{Exp}^{\neg Act}$). For the control group, we select the users who are targeted first but the ad is delivered to them randomly. We find if the user performs an action (actor and non-exposed, $N_{\neg Exp}^{Act}$) or not (non-actor and non-exposed, $N_{\neg Exp}^{\neg Act}$). We use standard Beta conjugate prior distribution to estimate the distribution of the probability of action given the data. Thus, the posterior distribution becomes:

$$P(\eta_1|\cdot) \propto P(\eta_1)\eta_1^{N_{Exp}^{Act}}(1-\eta_1)^{N_{Exp}^{\neg Act}} = Beta(1+N_{Exp}^{Act}, 1+N_{Exp}^{\neg Act}). \tag{3.19}$$

Here, η_1 is the likelihood of a user acting given exposure and $Beta$ is the Beta distribution. Similarly, we estimate the posterior distribution for the probability of action given non-

Table 3.5: A/B testing comparison with the attribution given by Mωlog and Mωlog+ for the RandWalk model.

Method	Campaign 1			Campaign 2		
	Low	Med	High	Low	Med	High
A/B	0.009	0.199	0.458	-0.034	0.115	0.312
Mωlog	0.076	0.289	0.421	-0.049	0.347	0.809
Mωlog+	0.133	0.272	0.623	0.094	0.180	0.519

exposure η_0. We estimate the change in probability of action respect to the control group $\gamma_{A/B}$=$(\eta_1 - \eta_0)/\eta_0$. This estimation is achieved by sampling from the distribution in Eq 3.19 for η_0 and η_1, and calculating the distribution for the statistic $\gamma_{A/B}$. For our method, we estimate the mean increase of actions attributed to the campaign respect to those attributed to other campaigns and the base model.

Table 3.5 shows the results for A/B testing and our method. We note that credible intervals in A/B testing are not tight as the probability of action is small, and the actions are sparse. Even when we run a randomized experiment, the sparsity of user conversions is an issue when we measure the campaign effects with a proper targeting algorithm as in a real scenario. We have one positive effect campaign in the 90% credible interval and one leaning towards positive effect in this interval. We compare these results with those obtained by our method incorporating the log transformation method and handling outliers. We obtain similar credible interval to that of A/B testing for Campaign 1. For Campaign 2 the zero effect is included in the 90% credible intervals but the effect is also leaning towards positive values. For positively constrained contributions, we have similar estimates for Campaigns 1 and 2, but with a larger credible interval. We do not compare with the weekly seasonal base model as few daily data points (4 weeks of data) are used. Overall, the median estimates for the two variants tested fall in the credible interval of A/B testing, except Mωlog for campaign 2. Similarly, the median estimates for A/B testing fall in the credible intervals of the proposed method. Larger intervals are the major difference between methods.

3.6.5 Synthetic Data Evaluation

To test our method with a ground truth, we use a public synthetic dataset, PROMO [86]. This dataset assumes a multiplicative model and different kinds of seasonality in sales not generated by campaigns. As only active/inactive indicators are available for each day, we incorporate inactive campaign days with zero impressions and active days with a fixed number of impressions. We use products with less than 6 relevant campaigns with the first 365 days of data. For this study, we consider 39 campaigns and 14 products. Employing our method, we recover the days the campaigns are active based on ability to discard the zero daily effect of the 90% credible interval, $\xi_t^{(c)}$. We estimate the following:

effDays Effective campaign days detected correctly

NeffDays Non-effective campaign days detected correctly

effDays-Prod Effective campaign days detected correctly per product

NeffDays-Prod Non-effective campaign days detected correctly per product

effCamp Effective campaigns detected correctly

Table 3.6 provides the results for the PROMO dataset. We detect around **85%** of effective campaigns. As a baseline, the winning team of the competition, which generates this

Table 3.6: Results of Mωlog and Mωlog+ for the RandWalk base model in the PROMO dataset.

Method	effDays	NeffDays	effDays Prod	NeffDays Prod	effCamp
$M\omega$log	0.633	0.833	**0.732**	0.862	**0.846**
$M\omega$log+	0.636	0.787	**0.735**	0.833	0.821

dataset, reported **78%** of recall [86]. In this context, recall is equivalent to **effCamp**. We also detect **73%** of the days a campaign is active per product. Note that in this scenario, there is no user tracking. Therefore, the use of A/B testing, or any population-based method, is *not* possible.

We have presented a time series based approach to measure the effects of online display marketing campaigns when user tracking is not available. We have modeled the impact of ad impressions on commercial actions through a DLM and provided daily effect estimates. We have incorporated persistence of campaign effects, through a decay factor, and accounted for outliers automatically without any threshold. We have presented several different campaign evaluation measures: 1)R^2, the standard measure in marketing. 2)The linear model typically assumed in regression analysis. 3)The dynamic bias base model, the standard choice in regression. 4)Positively constrained campaign impact. Although some measures perform better than others under certain circumstances, they are intended to provide a spectrum of choices for the practitioner at the time of evaluating a campaign. Nonetheless, we have found that a model in the log-scale is more effective in representing the relationship between ad impressions and commercial actions. Results indicate that a seasonal base model will give less attribution to campaigns.

Despite being able to estimate the campaign attribution at scale for thousands of campaigns, multiple factors might potentially confound the campaign attribution results. Thus, combining randomized experiments and the dynamic analysis of campaign effects proposed in this chapter enables us to estimate the campaign causal attribution at different points in time. We provide this methodology and analysis in Chapter 4.

Appendix

3.A Sampling distributions for Φ

We assume approximate non-informative priors for Φ. $IG(\alpha, \beta)$ represents Inverse Gamma distribution with shape α and rate β. $TN(m, C, a, b)$ refers to the Normal distribution truncated at $[a, b]$. T_0^c and T_f^c represent the start and end times of campaign c. $T^c{=}T_f^c - T_0^c + 1$. For priors, $[\alpha_{v0}, \alpha_{w0}^{(0)}, (\alpha_{\xi0}, \alpha_{\psi0})^{(1:N)}]{=}0.5$; $[\beta_{v0}, \beta_{w0}^{(0)}, (\beta_{\xi0}, \beta_{\psi0})^{(1:N)}]{=}10^{-6}$. Thus, the sampling distributions become:

$$\begin{aligned} &\widetilde{V} \sim IG(\alpha_v, \beta_v), && \alpha_v = \alpha_{v0} + \tfrac{T-1}{2}, \\ &SS_y = \textstyle\sum_{t=1}^{T} \omega_t (Y_t - \widetilde{F}'\theta_t)^2, && \beta_v = \beta_{v0} + \tfrac{1}{2} SS_y, \end{aligned} \tag{3.20}$$

$$\begin{aligned} W^{(0)} \sim IG(\alpha_w^0, \beta_w^0), \quad &\alpha_w^0 = \alpha_{w0}^0 + \tfrac{T-2}{2}, \\ &\beta_w^0 = \beta_{v0}^0 + \tfrac{1}{2} \textstyle\sum_{t=1}^{T-1} (\theta_{t+1}^{(0)} - G^{(0)} \theta_t^{(0)})^2, \end{aligned} \tag{3.21}$$

$$\begin{aligned} &W_\xi^{(c)} \sim IG(\alpha_\xi^{(c)}, \beta_\xi^{(c)}), && \alpha_\xi^{(c)} = \alpha_{\xi0}^{(c)} + \tfrac{T^c-2}{2}, \\ &\hat{\xi}_{t+1}^{(c)}{=}\lambda^{(c)} \xi_t^{(c)} {+} \psi_{t+1}^{(c)} X_{t+1}^{(c)}, && \beta_\xi^{(c)} = \beta_{\xi0}^{(c)} + \tfrac{1}{2} \textstyle\sum_{t=T_0^c}^{T_f^c-1} (\xi_{t+1}^{(c)} - \hat{\xi}_{t+1}^{(c)})^2, \end{aligned} \tag{3.22}$$

$$\begin{aligned} W_\psi^{(c)} \sim IG(\alpha_\psi^{(c)}, \beta_\psi^{(c)}), \quad &\alpha_\psi^{(c)} = \alpha_{\psi0}^{(c)} + \tfrac{T^c-2}{2}, \\ &\beta_\psi^{(c)} = \beta_{\psi0}^{(c)} + \tfrac{1}{2} \textstyle\sum_{t=T_0^c}^{T_f^c-1} (\psi_{t+1}^{(c)} - \psi_t^{(c)})^2, \end{aligned} \tag{3.23}$$

$$\begin{aligned} \lambda^{(c)} &\sim TN\left(m^{(c)}, C^{(c)}, 0, 0.88\right), \\ m^{(c)} &= \frac{\sum_{t=T_0^c}^{T_f^c-1} (\xi_{t+1}^{(c)} - \psi_{t+1}^{(c)} X_{t+1}^{(c)}) \xi_t^{(c)}}{\sum_{t=T_0^c}^{T_f^c-1} (\xi_t^{(c)})^2 + 1}, \quad C^{(c)} = \frac{W_\xi^{(c)}}{\sum_{t=T_0^c}^{T_f^c-1} (\xi_t^{(c)})^2 + 1}. \end{aligned} \tag{3.24}$$

We truncate the rate of decay λ to $[0, 0.88]$ which is equivalent to $[0, 5.44]$ days for the effect to decay by 50%. Note from Eq 3.20 that ω_t *weights* the squared difference in the observation variance.

The Inverse Gamma prior distribution of the variance parameters that we assume considers a prior sample size of 1 ($\alpha = 0.5$), and an approximately negligible prior sum-of-squared errors ($\beta = 10^{-6}$). This prior distribution choice is a proper conjugate prior and guarantees that the posterior distribution is proper. The effect of the prior distribution of the observational variance on the posterior fitting of the attribution results is negligible because of the series length (six of months of daily observations, $T = 182$). The effect of the prior choice on the fitting of the campaign-specific variance parameters depends on the duration of the campaign. We find that the impact of this prior distribution is not significant as long as the model is *observable*. For the current model, a campaign should be active at least for four days with non-zero impressions given that other campaigns and the base models are observable.

3.B Sampling distributions for Ω

We sample $\Omega | \theta_{1:T}, \Phi, D_{1:T}$ based on the generative model in Algorithm 1. $\Gamma(\alpha, \beta)$ is the Gamma distribution with shape α and rate β. We use a Dirichlet prior, α=1, for p.

This procedure leads to the following sampling distributions:

$$\omega_t \sim \Gamma(\alpha_\omega, \beta_\omega), \quad \alpha_\omega = \tfrac{\eta_t+1}{2}, \quad \beta_\omega = \tfrac{1}{2}\widetilde{V}^{-1}(Y_t - \widetilde{F}'\theta_t)^2, \tag{3.25}$$

$$\eta_t \sim p(\eta_t = i), \quad p(\eta_t = i) \propto \Gamma(\omega_t|\tfrac{i}{2}, \tfrac{i}{2})p_i, \tag{3.26}$$

$$p \sim Dir(\alpha + N_y), \quad N_y = [N_{y1}, \ldots, N_{yL}]', \quad N_{yi} = \textstyle\sum_{t=1}^{T}(\eta_t = i). \tag{3.27}$$

We set η_t to be $\{1, 2, \ldots, 10, 20, \ldots, 50\}$ whose cardinality, L, equals the dimension of the multinomial and its Dirichlet prior. We note that the prior distribution of η_t (the number of degrees of freedom of the t-distributed weighted errors) is not a conjugate prior. However, given that the number of possible values for η_t is countable and small, we sample η_t from Eq 3.26 using inverse transform sampling [30].

Part III

Campaign Causal Attribution based on Randomized Experiments

Chapter 4

Dynamic Causal Attribution: An Aggregated Approach

4.1 Introduction and Problem Context

The allocation of a given budget to online display advertising as a marketing channel has motivated the development of statistical methods to measure its effectiveness accurately. The use of randomized experiments, also known as A/B testing in the industry, has demonstrated to be effective to evaluate marketing campaigns without over-estimating their effects [55, 12]. These methods require a time window where users are tracked, and the metrics of interest are collected. As a result, the estimation is aggregated for that time window. This aggregation is a limitation as often sales are affected by seasonal effects. For instance, detecting which days of the week a given campaign is more effective provides insights to understand and improve the campaign.

4.2 Chapter Contribution

We propose a time series approach to estimate the effects of marketing campaigns on the daily number of sales or conversions. We consider the randomized design proposed by Barajas *et al.* (2012) in targeted display advertising [12], which is detailed in Chapter 5 [7]. In this framework, users are randomized into control and study groups before any decision has been made in the targeting process, as in the case of the standard Intention-to-Treat effect estimation [50].

We aggregate the daily number of conversions over all the users and consider these sales time series for the control and the study groups. Fig 4.3 shows the observed sales series for both treatment groups of two campaigns. We decompose these series jointly into weekly and trend components using Dynamic Linear Models (DLM) [70]. Based on this framework, we infer the daily mean causal effect as the sales trend differences between both series. We model both series jointly and estimate the average causal effect directly. We smooth effects during the sales evolution diminishing the sparsity issues of online sales.

4.3 Chapter Assumptions

In this section, we state the main assumptions used in the current chapter. The results and conclusions of this chapter are valid to the extent that these assumptions are applicable.

1. We refer to tracking cookies as users in the experimental design and estimation. We consider stable user cookies born before the campaign starts and that are active in the entire ad network.

2. The treatment assignment is assumed to be independent of the treatment effect, i.e. random assignment.

3. We do not consider interference or spillover effects between control and study groups. User interference might occur, but the impact of this interference on the user conversion probability is assumed to be negligible.

4. We assume a DLM to model the evolution of daily number of conversions for both treatment groups. The conversion time series and the latent evolving state are assumed to be Normally distributed random variables.

5. We assume that both control and study groups share a common background conversion series evolution. Thus, we consider the causal effect of the series trend component only.

6. The prior state distribution of the DLM developed in this chapter is a multivariate Normal distribution that is fitted automatically by maximum likelihood estimation.

7. The above model components represent the structure we assumed in this Chapter. Eq 4.1 illustrates the model equations of this structure.

4.4 Methodology

We define y_t^{ct} and y_t^{st} as the total number of online conversions observed for users in the control and study groups respectively. We model both series jointly and assume a weekly seasonal component to be the same for both groups. Thus, we analyze the campaign effects on the sales trend only. We assume a latent space model, using a DLM, where we model a

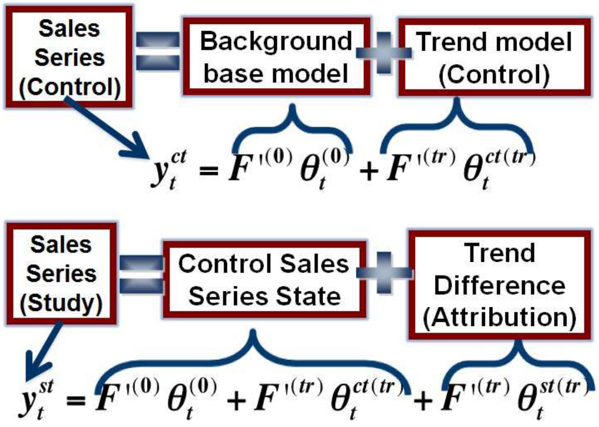

Sales Series (Control)
Background base model
Trend model (Control)
$y_t^{ct} = F'^{(0)}\theta_t^{(0)} + F'^{(tr)}\theta_t^{ct(tr)}$
Sales Series (Study)
Control Sales Series State
Trend Difference (Attribution)
$y_t^{st} = F'^{(0)}\theta_t^{(0)} + F'^{(tr)}\theta_t^{ct(tr)} + F'^{(tr)}\theta_t^{st(tr)}$

Figure 4.1: Dynamic linear model assumed for campaign attribution. This model assumes a background base model shared by both control and study treatment groups. We consider the dynamic campaign attribution at the sales trend level.

seasonal and trend sales components for both treatment groups at the state evolution. We define:

$$\begin{aligned} y_t^{ct} &= F'^{(0)}\theta_t^{(0)} + F'^{(tr)}\theta_t^{ct(tr)}, \\ y_t^{st} &= F'^{(0)}\theta_t^{(0)} + F'^{(tr)}\theta_t^{ct(tr)} + F'^{(tr)}\theta_t^{st(tr)}. \end{aligned} \tag{4.1}$$

$\theta_t^{(0)}$ represents the state of a shared (background) base model, which we assume to be a weekly seasonal model. $\theta_t^{ct(tr)}$ is the trend model for the control group, and $\theta_t^{st(tr)}$ is the difference in sales trends attributed to the campaign. $F^{(0)}$ and $F^{(tr)}$ represent observational matrices to model the trend and the base components respectively. This model is depicted by Fig 4.1.

We consider the case of unbalanced probabilities of user assignment to the study group $z = 1$, and to the control group $z = 0$, which the randomized design fixes. We write this model as a 2-D DLM as follows:

$$\begin{aligned} Y_t &= F'\theta_t + \nu_t, & \nu_t &\sim N(0, V), \\ \theta_t &= G\theta_{t-1} + w_t, & w_t &\sim N(0, W), \end{aligned} \tag{4.2}$$

where:

$$
\begin{aligned}
Y_t &= [y_t^{ct}, y_t^{st}]', \qquad\qquad \theta_t = [\theta_t^{ct(tr)}, \theta_t^{st(tr)}, \theta_t^{(0)}]', \\
F' &= \begin{bmatrix} p(z=0) & 0 \\ 0 & p(z=1) \end{bmatrix} \times \begin{bmatrix} F'^{(tr)} & 0 & F'^{(0)} \\ F'^{(tr)} & F'^{(tr)} & F'^{(0)} \end{bmatrix}.
\end{aligned}
\tag{4.3}
$$

We set $F^{(tr)}$ and $F^{(0)}$ to model a random walk trend and a weekly seasonal component. We use the Fourier representation for two harmonics defined defined by Eq 3.11 of Chapter 3. Similarly, G is constructed as the superposition of the basic components assumed. Thus, these matrices are fixed based on these simpler models[1]. This representation allows us to model the expected trend difference between the treatment groups in the evolution. Also, we enforce the seasonal base model, in this case, a weekly seasonal component.

The parameters of the model of Eqs 4.2 and 4.3 that need to be fitted are the observational and evolution covariance matrices $\{V,W\}$, and the initial state prior distribution, $\{m_0, C_0\}$ where $\theta_0 \sim N(m_0, C_0)$. Thus, we define the model parameter set to be $\Phi = \{V, W, m_0, C_0\}$. To guarantee a unique optimal fitting of the parameters, we consider the matrices W and V to be diagonal. We calculate the Maximum Likelihood (ML) estimate of these parameters using an Expectation-Maximization (EM) approach [42]. Given the parameters $\Phi = \{V, W, m_0, C_0\}$, we estimate the distribution of the latent states $P\left(\theta_t|Y_{1:T}\right)$ for $t = 1, \ldots T$ using the Kalman filtering and backward smoothing equations (E-step). We then optimize the augmented likelihood after replacing the expected values for each state (M-step). For details of the optimization see [42]. These steps are performed iteratively until convergence. Figure 4.2 illustrates this EM-based fitting process.

Given the ML estimates Φ^*, we smooth the time series to calculate the expected

[1] See [70] pages 89-95 for the random walk trend, and 102-106 for the Fourier seasonal models to set these components.

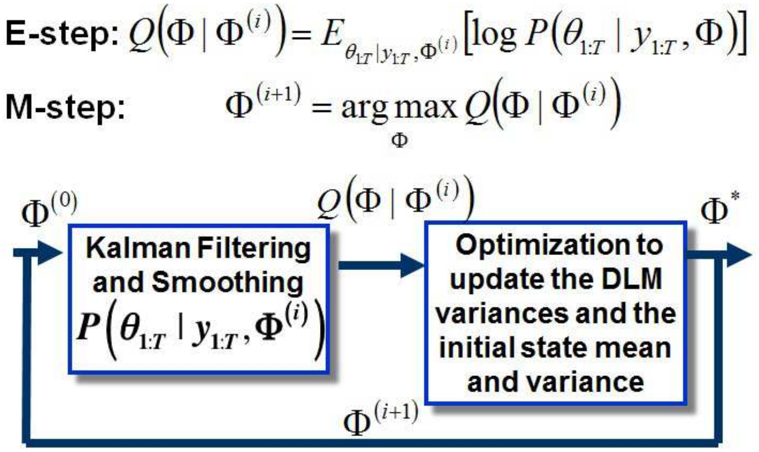
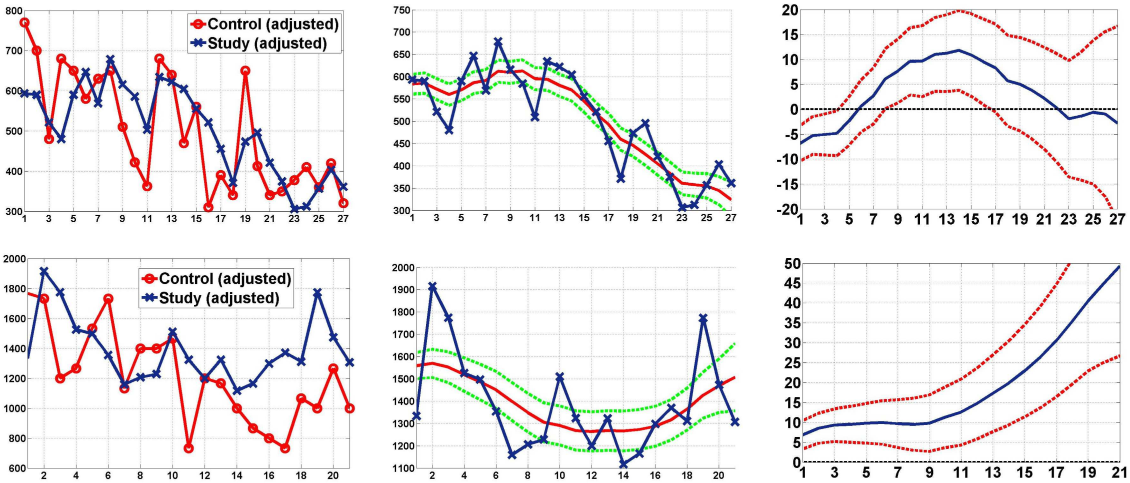

E-step: $Q(\Phi \mid \Phi^{(i)}) = E_{\theta_{1:T} \mid y_{1:T}, \Phi^{(i)}}[\log P(\theta_{1:T} \mid y_{1:T}, \Phi)]$
M-step: $\Phi^{(i+1)} = \arg\max_{\Phi} Q(\Phi \mid \Phi^{(i)})$
$\Phi^{(0)}$
$Q(\Phi \mid \Phi^{(i)})$
Φ^{*}
Kalman Filtering and Smoothing $P(\theta_{1:T} \mid y_{1:T}, \Phi^{(i)})$
Optimization to update the DLM variances and the initial state mean and variance
$\Phi^{(i+1)}$

Figure 4.2: EM iterations used to calculate the Maximum Likelihood estimator of the DLM model parameters where $\Phi = \{V, W, m_0, C_0\}$.

causal trend difference attributed to the campaign. Fig 4.3 shows the trend component fitted for the study group for two campaigns. We estimate the causal lift (CL_t) as the percentage change in sales trends, due to the campaign, with respect to the control trend: $\mathrm{CL}_t = 100 \times F'^{(tr)}\theta_t^{st(tr)}/F'^{(tr)}\theta_t^{ct(tr)}$. We use the Delta method to approximate the distribution of the ratio of two Normal random variables [21]. We set the initial parameter values of the EM fitting process randomly. We run the model multiple times with different random initialization values without significant changes in the average attribution results.

4.5 Results

Fig 4.3 shows the results for two real campaigns. As illustrated, the attribution is not evident from the observed data. This lack of clarity of the campaign attribution is a consequence of the seasonal component that affects both series, and typical noisy conversion data.



(a) (b) (c)

Figure 4.3: Dynamic Causal Attribution for Campaign 1 (top) and Campaign 2 (bottom). (a) Observed conversions adjusted based on $p(z)$ (y-axis represents the number of conversions). (b) Series trend fitted for the study group (y-axis represents the number of conversions). (c) Dynamic causal attributed lift CL_t in percentage (%). x-axis represents days.

We observe from the causal lift evolution that there are positive and negative effects of Campaign 1 at different times. Even when the observed data suggests this positive impact, comparing point by point is highly problematic, and it does not provide any statistical support. This behavior shows a campaign with immediate effects where at the beginning of the campaign users wait to buy, probably to survey the competition. Then, the campaign effects peak to fade gradually to the prior campaign sales level. Thus, no significant brand advertising effect is observed in the short term.

For Campaign 2, positive attribution results are evident from the observed data towards the end of the series. As a consequence, the causal campaign lift shows an increasing tendency. Note that even when several points of the study series are greater than those in the

Table 4.1: Mean attribution lift (%) estimated from the trend differences and the raw data.

Method	Campaign 1			Campaign 2		
	Low	Med	High	Low	Med	High
MCL - Trend	1.31	3.11	4.91	17.03	19.47	21.90
MCL - Raw	-5.03	1.31	7.65	8.29	14.50	20.71

control series, the actual increase is hard to obtain by inspection. This analysis illustrates that Campaign 2 provides delayed effects after the campaign is finished, as opposed to Campaign 1.

Table 4.1 shows the average campaign effects estimated from the series trends, and from the raw data. We obtain the mean causal lift (MCL - Trend) as the average CL_t for the campaign duration for both treatment groups. We compare this measure with the raw estimation (MCL - Raw), obtained from the sample mean of the observed data points without using the time sequence. As depicted, this raw measure is noisier and does not provide any insight into the time when the campaign is more effective.

4.6 Impact and Limitations

We have presented a time series based approach to attribute trend differences to marketing campaigns. We attribute these differences using causal estimates based on a randomized experiment. We constrain the evolution to be smooth to avoid sudden changes in the attribution. This method provides disaggregated estimates that allow us to obtain marketing insights about the time that the campaign is effective.

The analysis of two campaigns shows different campaign attribution levels at various points in time while the campaigns are running. Overall, we observe a small performance at the beginning of those campaign followed by a performance boost. These results show a typical cold-start campaign phase that eventually improves as more data is available from user campaign exposures.

The current Chapter approach complements the observational analysis of Chapter 3 to estimate the campaign causal effect evolution when randomized data is available. We focus the current aggregate analysis on calculating the causal attribution of the online display advertising channel for budget allocation[2]. However, a user-level evaluation analysis is required to guide and improve the campaign user targeting. We address the campaign evaluation at the user level in Chapters 5 and 6.

[2] See [52] for a budget allocation application of campaign attribution.

Chapter 5

User-Level Causal Evaluation: Defining the Campaign Counterfactual

5.1 Introduction and Problem Context

User-level campaign evaluation is often performed with the objective of improving future ad exposures. In this context, running randomized experiments (or field experiments) is becoming the standard approach to measuring the marginal effectiveness of online campaigns, and guaranteeing a causal attribution [27, 55, 87]. In this practice, the ad creative is assumed to be the *treatment* to evaluate, and users are randomly separated into two groups, study and control. Hence, when a targeting engine selects a visiting user for ad exposure, the campaign ad is displayed to users in the study group, or a placebo ad is displayed to users in the control group [87].

The full deployment of this framework is limited by the cost of showing placebo ads, which typically consist of charity ads, and the potential revenue loss resulting from yielding the opportunity to advertise to control users. As a consequence, a low-budget randomized experiment is often performed, followed by a larger-budget investment that purports to generalize (external validity) the measured effect of this campaign without further testing [27].

Recently, marketing campaigns are increasingly taking place on ad exchange platforms. These platforms facilitate marketplaces where advertising spaces on websites are bought and sold. A survey of 49 media buyers indicates that 87.8% intended to purchase digital advertising via real-time bidding (RTB) by 2011 [31]. Similarly, outside RTB exchanges, ad networks run internal auctions in regular basis [18]. Consequently, the external validity of campaign effects estimated in an environment assumed to be free of competitors, to a marketplace with competitors, is likely to be inaccurate. Because media buying is performed endogenously in a competitive market, the user targeting complicates the evaluation using placebo ads. Moreover, to display a placebo ad, the opportunity to advertise must be *consumed*, and the campaign must exist in the marketplace (campaign presence effect). Otherwise, its absence introduces competitor (synergizer) effects.

5.2 Chapter Contribution

We focus on the marginal causal attribution of single-product online conversions to online display campaigns (single channel) run on hundreds of publisher websites, given

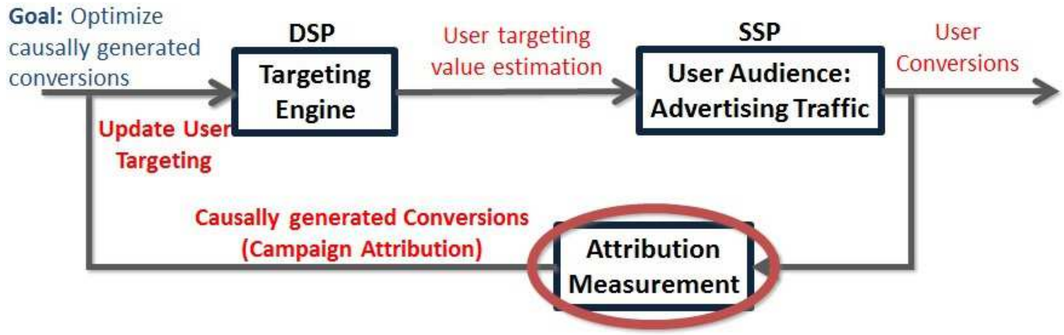

Goal: Optimize causally generated conversions
DSP
Targeting Engine
User targeting value estimation
SSP
User Audience: Advertising Traffic
User Conversions
Update User Targeting
Causally generated Conversions (Campaign Attribution)
Attribution Measurement

Figure 5.1: Online advertising optimization loop. The focus of this chapter is the measurement (attribution) box.

all other advertising channel exposures or prior branding effect. We report that the current industry practice confounds three campaign effects. These effects are the ad effect on exposed users, the strategic impact of the campaign presence in a competitive market, and the targeting (selection) effect of the media buyer. We summarize the elements of our contribution below in this context.

Expand the scope of attribution in marketplaces to the overall campaign We propose to perform continuing evaluation and estimate the campaign attribution for the current running conditions instead of isolating the ad creative effect. In this new perspective, the entire campaign, which includes: 1)The campaign presence in the marketplace and 2)The ad creative, is now the treatment to evaluate. Consequently, we propose a new randomized design that considers all the visiting users where the control treatment arm is not exposed to placebo ads. We argue that this is the right campaign counterfactual in a marketplace. This design cost, which is minimal in terms of revenue loss, enables us to perform continuing evaluation and attempts to close the feedback loop for campaign causal optimization displayed by Fig 5.1. The proposed design is simple to implement and estimate and does not suffer from endogenous targeting.

Capture the effect of the campaign presence in the marketplace We propose a second randomized design that separates the ad effect from the impact of the campaign presence in the marketplace. We show the risks of a selection effect for the standard campaign evaluation practice of using placebo ads in a marketplace, which is a consequence of endogenous user targeting. Contrary to the case of paid search effectiveness analyzed by Blake *et al.* (2014), we find evidence of a campaign presence effect [17]. This effect, which is an ineluctable consequence of running the campaign, is ignored in the standard practice, and significantly change the campaign attribution.

5.3 Chapter Assumptions

In this section, we state the main assumptions used in the current chapter. The results and conclusions of this chapter are valid to the extent that these assumptions are applicable.

1. We refer to tracking cookies as users in the experimental design and estimation. We consider stable user cookies born before the campaign starts and that are active in the entire ad network.

2. The treatment assignment is assumed to be independent of the treatment effect, i.e. random assignment.

3. No interference or spillover effects are considered between control and study groups.

User interference might occur, but the impact of this interference on the user conversion probability is assumed to be negligible.

4. Stable Unit Treatment Value Assumption (SUTVA). We assume that the treatment status of any user does not affect the potential outcomes of the other users, i.e. no interference between users is assumed.

5. We model the user conversions as a random events, with a predetermined probability for each treatment arm. Thus, the converting users are conditionally independent of each other given this probability.

6. The user ad exposure is assumed to be binary (targeted or non-targeted) without considering the number ad exposures. Similarly, the visiting user indicator and the converting user indicator do not consider the multiple instances of these events for a given user (see table 5.4).

7. Model parameters are assumed to be random, with standard Jeffrey's conjugate prior distribution. Indicator random variables are assumed to be Bernoulli distributed with prior distribution: Beta(0.5,0.5).

5.4 Experimental Design for Attribution in Marketplaces

We discuss the framework of online Targeted Display Advertising in a marketplace to design the randomized experiment with the right counterfactual. Then, we analyze the randomized design using placebo ads as in the current industry practice. We show the risk of selection bias of this design in Targeted Display Advertising. We present our proposed

Table 5.1: Description of the variables used in chapter 5.

Term	**Description**
Z∈{C,P,S}	Random treatment assignment
Y∈{0,1}	Converting user indicator
B∈{No,Yes}	Decision to bid indicator
A∈{Lose,Win}	Auction output indicator
D∈{0,1}	Targeted user indicator
W∈{0,1}	Ad exposure indicator
$i \in \mathbb{N}$	Variable index for the i-*th* user
$\mathrm{N}_{dz}^{y} \in \mathbb{N}$	User count given $Y = y, D = d, Z = z$
$\Delta^{select} \in [-1, 1]$	Statistic to test for equivalent user selection for placebo ads
$\Delta^{convert} \in [-1, 1]$	Statistic to test for equivalent user populations for placebo ads

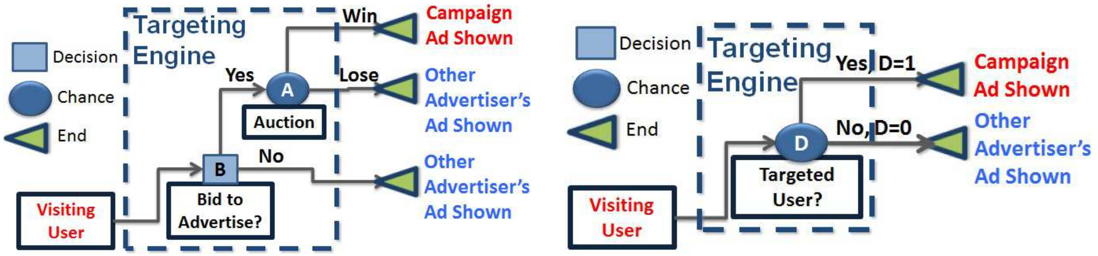
model and indicate the campaign attribution estimated by this design in terms of the ad creative effect and the impact of the campaign presence in the marketplace. Finally, we show that the proposed method is cost-effective for continuous campaign evaluation.

5.4.1 Targeted Display Advertising in Marketplaces: Overview

In Targeted Display Advertising, marketing campaigns are often run by advertisers working closely with a given ad network. The mechanism for displaying an ad is depicted by the decision tree of Fig 5.2(a). This mechanism is based on conducting an auction for every visiting user who is provided by a supply-side platform (SSP) or publisher websites.

Table 5.2: Performance metrics analyzed chapter 5. Lifts are in the range: $\in [-1, \infty)$. Average Treatment Effects (ATE) are in the range: $\in [-1, 1]$.

Metric	**Lift**	**Description**
ATE_{Camp}	lift_{Camp}	Overall campaign average effect on all visiting users
ATE_{Ad}	ACL_{Ad}	Average effect of the ad creative on targeted users
ATE_{Market}	ACL_{Market}	Average effect of the campaign presence in the marketplace on targeted users
LATE_{Ad}	lift_{ad}	Local average treatment effect of the campaign on targeted users (developed in Chapter 6)



(a) (b)

Figure 5.2: Online targeted display advertising flow for a given user visit.

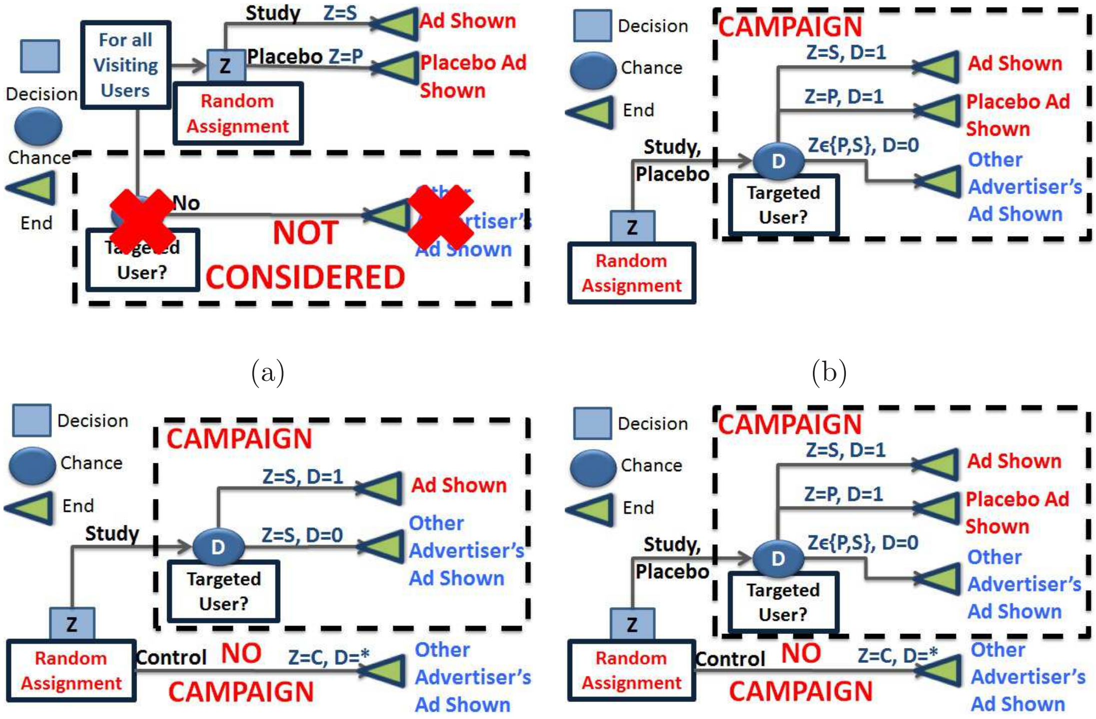
To target users, advertisers develop user profiles of the target market segment based on demographics and other features. In practice, however, the ad network employs a highly sophisticated algorithm, illustrated by the decision node B of Fig 5.2(a), to determine if a user should be targeted. In performance-based Cost-per-Action (CPA) campaigns, this decision is based on user behavior and history, and how likely the user is to convert, among other features [66, 1]. If the campaign decides to bid through a demand-side platform (DSP) in the ad exchange (B=Yes), it submits the bid through Real Time Bidding (RTB) [79]. This action illustrated by the chance (endogenous) node A of Fig 5.2(a). If the campaign wins the advertising slot (A=Win), the campaign ad is displayed to the user. Otherwise, other advertiser shows an ad. For cost-per-display (CPM) campaigns, the decision to bid is set to B=Yes, and the bidding strategy is determined by guaranteed delivery contracts or the spot market [43] [1]. Outside of ad exchanges, these targeting and auction processes are routinely run by large ad networks [18]. For the effects of the current analysis, we consider the aggregate targeting engine output (chance node D of Fig 5.2(b)) to refer to targeted users. Here, D=1 if the user is targeted, *i.e.* if B=Yes and A=Win, and D=0 otherwise.

5.4.2 Campaign Evaluation using Placebos: The Standard Practice

The standard approach to evaluating an online marketing campaign is to use randomized experiments assuming the creative ad design, including the advertising message, is the *treatment* to evaluate. Lewis *et al.* (2011) propose randomly assigning the visiting users at serving time. Their goal is to see the focal ad (study), or the placebo ad in the form of a charity ad assumed to be an unrelated ad (control) [55]. Fig 5.3(a) illustrates this process. This framework is too limited to be applied to the standard Targeted Display Advertising in a marketplace. In this method, none of the components discussed earlier in this Section are considered. Also, the randomizing user visits limit the power of this design because a given user might be assigned to both treatment arms during different visits.

The current industry practice is to randomize the visiting users once and keep

[1] Recently a minimum performance level constraint has been imposed on CPM campaigns. However, the optimization of the user targeting is minimal here when compared with CPA campaigns.



(a)

(b)

(c)

(d)

Figure 5.3: (a) User randomization framework proposed by authors of [55] without user targeting selection. (b) Standard industry randomization practice with placebo ads. (c) Proposed randomization design for campaign attribution. (d) Randomization framework with disaggregated campaign effects.

them in the same arm throughout the whole experiment, as depicted by Fig 5.3(b) [87]. Because the media buying is performed endogenously in a competitive market, the user targeting indicator D becomes a post-treatment variable. Conditioning the analysis on its realization might introduce a post-treatment bias[2]. Moreover, the targeting engine routinely incorporates user activity feedback, such as user clicks and visits, to improve the targeting [1], which would not be the case for the placebo ad.

More generally, these practices focus on the ad evaluation, without considering

[2] A post-treatment variable is a random variable whose realization is available after the randomization assignment has been performed. As a result, the treatment can potentially affect this realization [36]

the effect of the campaign presence in the marketplace. Also, the ad is often evaluated with a low-budget CPM campaign, and the effects are assumed to hold for larger-budget CPA campaigns (Chittilappilly in [27] describes a general industry practice). However, the external validity of CPM campaign effects to CPA campaigns is prone to inaccuracies due to the CPA targeting incentives [14], and the market interactions [64].

5.4.3 Proposed Randomized Design

We propose evaluating the overall campaign, including the ad and the campaign presence in the marketplace. This new perspective implies that the campaign is now the *treatment* to evaluate. We randomize the visiting users before making any decision in the decision tree of Fig 5.3(c), and keep them in the same group for the campaign duration. As a result, users in the control group are not exposed to placebo ads. This design aggregates the ad and campaign presence in the marketplace effects analyzed in detailed below. Our goal for this randomized design is *not* to predict or generalize the campaign performance for future long-term exposures that are the objective of randomized experiments. Our goal is to evaluate the campaign performance under the current conditions to attribute credit to its overall performance, which is the key attribution problem of interest to online advertisers. In the context of the campaign loop of Fig 5.1, our focus is short-term (mid-flight) ad prediction where both effects are stable.

To disaggregate the proposed design of Fig 5.3(c), we consider the design of Fig 5.3(d), where $Z \in \{\text{Control}, \text{Placebo}, \text{Study}\} = \{C, P, S\}$. To avoid a selection effect, two assumptions of the observed targeting in the study and placebo arms need to be tested:

Assumption 1. *Statistically equivalent user selection; the marginal probability of user targeting is the same for both treatment arms*

Assumption 2. *Statistically equivalent targeted populations; the marginal conversion probability of the non-targeted users is the same for both treatment arms*

Testing Assumption 1 indicates whether the marginal targeting policy (aggregated over all user segments) is the same for both placebo and study arms in the average. Testing Assumption 2 indicates whether the user marginal targeting process (aggregated over all user segments) provides statistically equivalent populations. If the non-targeted populations are equivalent, in terms of conversion probability, then the complementary populations are statistically equivalent as a consequence of user randomization. Although rejecting Assumption 1 suggests non-equivalent user targeting, testing Assumption 2 is what determines the presence of a selection effect (bias) in the observed data.

Let $Y_i(Z_i)$ be the i^{th} user conversion indicator under the treatment Z_i, and assume Assumption 2 holds. Similarly, assume $P(Y_i(C)|D_i = 1, Z_i = C)$ is known for the control group, in which the targeted user indicator D_i is not observed; we address this estimation in Section 6.4 of Chapter 6. Thus, the ad average treatment effect $\text{ATE}_{Ad,i}$, and the average treatment effect of the campaign presence in the marketplace $\text{ATE}_{Market,i}$ are defined as follows:

$$\begin{aligned}\text{ATE}_{Ad,i} &= \text{E}[Y_i(S)|D_i = 1, Z_i = S] - \text{E}[Y_i(P)|D_i = 1, Z_i = P],\\ \text{ATE}_{Market,i} &= \text{E}[Y_i(P)|D_i = 1, Z_i = P] - \text{E}[Y_i(C)|D_i = 1, Z_i = C].\end{aligned} \tag{5.1}$$

The proposed randomized design of Fig 5.3(c) takes the entire campaign as *treatment*, and estimates the campaign average treatment effect ($\mathrm{ATE}_{Camp,i}$) as follows:

$$\begin{aligned}\mathrm{ATE}_{Camp,i} &= \mathrm{E}[Y_i(S)|Z_i = S] - \mathrm{E}[Y_i(C)|Z_i = C] \\ &= \sum_{\forall d} P(D_i = d) \times \mathrm{E}[Y_i(S)|D_i = d, Z_i = S] \\ &- \sum_{\forall d} P(D_i = d) \times \mathrm{E}[Y_i(C)|D_i = d, Z_i = C].\end{aligned} \tag{5.2}$$

Given that Y_i is affected only for the users whom the ad is displayed to, *i.e.* $\{\forall i : D_i{=}1\}$, all other terms of Eq 5.2 cancel out except for this sub-population. Thus, by substituting for $\mathrm{ATE}_{Ad,i}$ and $\mathrm{ATE}_{Market,i}$ from Eq 5.1 we have:

$$\begin{aligned}\mathrm{ATE}_{Camp,i} &= P(D_i = 1) \times \{\mathrm{E}[Y_i(S)|D_i = 1, Z_i = S] - \mathrm{E}[Y_i(C)|D_i = 1, Z_i = C]\} \\ &= P(D_i = 1) \times \{\mathrm{ATE}_{Ad,i} + \mathrm{ATE}_{Market,i}\}.\end{aligned} \tag{5.3}$$

Therefore, the campaign effect of the proposed design, depicted by Fig 5.3(c), provides the aggregated ad and campaign presence effects. These are weighted by the probability of displaying the campaign/placebo ad. This weighting term is a consequence of a larger user population considered by the campaign (of all visiting users), rather than the sub-population of exposed users required to compute $\mathrm{ATE}_{Ad,i}$.

The standard evaluation using placebo ads identifies $\mathrm{ATE}_{Ad,i}$ as the “campaign” effect. However, the estimation of the campaign economic value (campaign attribution) based on $\mathrm{ATE}_{Ad,i}$ alone does not incorporate $\mathrm{ATE}_{Market,i}$, which is a consequence of displaying the ad. Therefore, the summation of these *two effects* must be considered. We analyze the value of $\mathrm{ATE}_{Market,i}$, for different scenarios in Appendix 5.A. Similarly, we discuss the cost of the proposed design in terms of potential revenue loss and targeted advertising in Appendix 5.B. We show that this randomized design is the one with lowest potential revenue loss when compared with the standard practice, and the most suitable for continuing evaluation.

Remark 1. *The design of Fig 5.3(c) identifies the right counterfactual to estimate* ATE_{Camp} *(Eq 5.2) when the objective is to calculate the campaign attribution. Fig 5.3(d) design disaggregates* ATE_{Camp} *into* ATE_{Ad} *and* ATE_{Market} *(Eq 5.3), which are both campaign effects.*

Remark 2. *To estimate* ATE_{Market} *(Eq 5.1), the expected conversion probability of users who would be targeted if in the study group but are actually in the control group,* $E[Y_i(C)|D_i = 1, Z_i = C]$*, needs to be inferred. The user's targeting indicator,* D_i*, is not observed for the control group. Section 6.4 addresses this estimation problem.*

Remark 3. *One might be inclined to believe that the three-arm design described by Fig 5.3(d) can be easily analyzed, as an extension of the standard randomized experiment of Fig 5.3(b), which includes a placebo arm. We reiterate that the error in that logic, and the reason for a different counterfactual and estimation method, is that the publisher slot must be captured and assigned to the campaign or placebo ad.*

5.5 Results: Campaign Evaluation using Placebo Ads

To illustrate the effect of the campaign presence in the marketplace, and the risk of conditioning the ad effect on post-treatment (endogenous) variables, we ran a large-scale experiment. Here we consider three treatment groups, $Z_i \in \{$Control,Placebo,Study$\}$=$\{C, P, S\}$ (Fig 5.3(d) design), collaboratively with an advertiser in the financial information services sector. We implemented the standard practice to evaluate online campaigns and ran a low-budget CPM campaign, where user conversions are economically equivalent. The advertiser

Table 5.3: Campaign data based on the experimental design of Fig 5.3(d) to disaggregate the campaign effects. Campaign active duration: 16 days. $\{0, 1\}$ represents unobserved user selection for ad exposure indicator.

$\mathbf{W_i}$	$\mathbf{D_i}$	$\mathbf{Y_i}$	$\mathbf{Z_i}$	**Count**	**CPM Campaign**			
0	{0,1}	0	C	$N^0_{\{0,1\}C}$	57,492,247			
0	{0,1}	1	C	$N^1_{\{0,1\}C}$	8,131	$\mathbf{Z_i}$	**Count**	**Placebo**
0	0	0	S	N^0_{0S}	9,938,896	P	N^0_{0P}	9,817,552
0	0	1	S	N^1_{0S}	1,246	P	N^1_{0P}	1,182
1	1	0	S	N^0_{1S}	3,618,467	P	N^0_{1P}	3,713,430
1	1	1	S	N^1_{1S}	607	P	N^1_{1P}	583

sells multiple products, and other campaigns were run simultaneously to market those products. Table 5.3 shows the aggregated data (Campaign CPM) based on the notation of Table 5.3, and Table 5.4 shows user activity statistics. Given that the users were randomized once before the user targeting were performed, there was a selection effect induced by conditioning the analysis on the observed targeting indicator. Here, the auction process prevented this indicator from being controllable (endogenous media buying). To verify that there was no selection effect, we now test the Assumptions 1 and 2 of Section 5.4.3.

Define the targeting indicator D_i under the treatments, Z_i=$\{P, S\}$, to be $\{D_i^P, D_i^S\}$. To estimate the ad effect conditional on the observed D_i^z, we define Δ_i^{select} and $\Delta_i^{convert}$ as:

$$\begin{aligned}\Delta_i^{select} &= P(D_i^S = 1|Z_i = S) - P(D_i^P = 1|Z_i = P),\\ \Delta_i^{convert} &= P(Y_i(S) = 1|D_i^S = 0, Z_i = S) - P(Y_i(P) = 1|D_i^P = 0, Z_i = P).\end{aligned} \tag{5.4}$$

Then, we define the hypotheses: $H_0^{select} : \Delta_i^{select} = 0$, $H_0^{convert} : \Delta_i^{convert} = 0$. We test

Table 5.4: User activity statistics for CPM Campaign of Table 5.3. Mean and standard deviation (Std) are displayed. **Visits/user** is the number of visits per user. **Convs**$|Y_i = 1$ is the number of conversions per converting user. **Imps/user** is the number of ad exposures per targeted user ($D_i = 1$)

	$Z_i = C$		$Z_i = P$		$Z_i = S$	
Variable	Mean	Std	Mean	Std	Mean	Std
Visits/user	18.18	93.67	18.22	93.40	18.22	94.04
Convs$\|Y_i = 1$	1.03	0.36	1.03	0.18	1.04	0.33
	$Z_i = P, D_i = 1$		$Z_i = S, D_i = 1$			
	Mean	Std	Mean	Std		
Visits/user	54.42	132.91	54.40	133.12		
Convs$\|Y_i = 1$	1.02	0.16	1.05	0.46		
Imps/user	1.68	1.35	1.70	1.39		

these hypotheses, and estimate their lifts (Δ_i^{select} Lift, $\Delta_i^{convert}$ Lift), by sampling the Beta distribution as in the case of the lift$_{Camp}$ estimation of Section 6.4.2 [3]. The testing results of Table 5.5 suggest rejecting H_0^{select} (Δ_i^{select} Lift= $[-2.84\%, -2.75\%, -2.65\%]$), and not rejecting $H_0^{convert}$ ($\Delta_i^{convert}$ Lift= $[-2.80\%, 4.12\%, 11.41\%]$). As a result, the change of the user targeting probability was not enough to reject the assumption that the sampled placebo and campaign populations are equivalent in conversion rates[4].

We estimate the lift effect of the ad ACL$_{Ad}$, based on ATE$_{Ad}$ of Eq 5.1 in Section 5.4.3, which is the standard "campaign" attributed effect. We report a positively leaning effect (ACL$_{Ad}$ = $[-2.78\%, 6.74\%, 17.97\%]$). We perform the analysis of Section 6.4.1 to

[3] We calculate the t-statistic for these conversion probability differences and the results are equivalent. However, the estimation of the lifts requires other approximations.

[4] We expect larger effects for CPA campaigns where the targeting of placebo ads must be equally optimized.

Table 5.5: Campaign disaggregated results, and validation of the placebo campaign based on 90% credible intervals. {*Low, Med, High*} are the {0.05, 0.5, 0.95} quantiles

Δ^{select}(1e-3)			$\Delta^{convert}$(1e-6)					
Low	Med	High	Low	Med	High			
-7.81	-7.53	-7.26	-3.50	4.95	13.30			
Δ^{select} Lift(%)			$\Delta^{convert}$ Lift(%)					
Low	Med	High	Low	Med	High			
-2.84	**-2.75**	-2.65	-2.80	4.12	11.41			
ATE$_{Ad}$(1e-5)			ATE$_{Market}$(1e-5)			LATE$_{Ad}$(1e-5)		
Low	Med	High	Low	Med	High	Low	Med	High
-0.46	1.06	2.70	-4.80	-2.79	-0.62	-3.83	-1.69	0.44
ACL$_{Ad}$(%)			ACL$_{Market}$(%)			lift$_{ad}$(%)		
Low	Med	High	Low	Med	High	Low	Med	High
-2.78	**6.74**	17.97	-24.02	**-15.06**	-3.70	-18.88	**-9.15**	2.62

calculate E$[Y_i(C)|D_i = 1, Z_i = C]$, and estimate ATE$_{Market}$ lift, ACL$_{Market}$, based on Eq 5.1. We estimate a negative effect of the campaign presence in the marketplace and discard the zero effect of the 90% credible interval (ACL$_{Market} = [-24.02\%, -15.06\%, -3.70\%]$). We know that the focal campaign competed in the marketplace against interacting campaigns run to advertise other products of the same brand. As a result, the presence of the current campaign alone prevented the other ads of the same advertiser from being displayed. Similar spillovers across product campaigns have been detected before by Sahni *et al.* (2014) in the context of email coupon promotions [77]. Note that this negative effect moves the campaign effect significantly based on the local campaign effect of Eq 6.4: LATE$_{ad}$, lift$_{ad}$

($\text{lift}_{ad} = [-18.88\%, -9.15\%, 2.62\%]$). Therefore, assuming that ATE_{Market} is a confounding effect and should be eliminated misses an important component of the campaign attribution. In this context, the campaign must exist in the marketplace to obtain the benefits of the ad, which makes $\text{ATE}_{Market} \neq 0$.

5.6 Impact and Limitations

We have shown that evaluating an online advertising campaign involves more than evaluating just the ad. As we have discussed in Section 5.4.3, the marketplace interactions imply that the final decision to display the campaign/placebo ad is not entirely controllable (endogenous) in the randomized experiment. We demonstrate this phenomenon with the evaluation of a campaign using placebo ads in Section 5.5. We can not expect that an ad tested in a controlled environment, as assumed by the exploratory evaluation of CPM campaigns, will have the same performance in a real marketplace. Similarly, the effects of being in the marketplace are ineluctable if the ad is to be displayed. Consequently, the right placebo is the complete absence of the campaign, and the randomized experiment becomes the measuring tool that runs with minimal cost.

The proposed experimental design assumes the overall campaign to be the treatment to evaluate. Based on this design, all visiting users are exposed to the campaign, and the campaign effect on the users exposed to ad becomes a local treatment effect (LATE). This framework is similar to the Intention-to-Treat standard analysis where the randomization of individuals occurs before the delivery of the treatment [50]. For the case of targeted display advertising, analyzing the LATE of the campaign on the exposed users enables us to characterize the campaign user targeting. We address this estimation in Chapter 6 for targeting-optimized CPA campaigns and non-targeted optimized CPM campaigns.

Appendix

5.A Effect of the Campaign Presence in the Marketplace Analysis

Based on the three-arm design of Fig 5.3(d), $Z_i \in \{\text{Control}, \text{Placebo}, \text{Study}\} = \{C, P, S\}$, we define $\pi_i(Z_i)$ to be the competitors' targeting policy. Let $\pi_{0,i}$ denote the competitors policy if the focal campaign does not exist ($\pi_i(C) = \pi_{0,i}$). Let $\pi_{1,i}$ be the alternative policy competitors execute with probability α as a consequence of the campaign presence in the marketplace. If competitors are not interested in user i with probability $1 - \alpha$, they will not compete to target this user and $\pi_i(Z_i) = \pi_{0,i} : \forall Z_i \in \{P, S\}$. Let β represent the probability that competitors would win the opportunity to advertise in the control group, but lose against the focal or placebo campaigns, and their ads have an effect on Y_i. These definitions lead to the distributions:

$$\begin{gathered} P(\pi_i(Z_i) = \pi_{0,i}|Z_i) = \begin{cases} 1 & \text{if } Z_i = C \\ 1-\alpha & \text{if } Z_i \in \{P, S\} \end{cases}, \\ P(\pi_i(Z_i) = \pi_{1,i}|Z_i) = 1 - P(\pi_i(Z_i) = \pi_{0,i}|Z_i), \\ P\left\{\mathrm{E}\left[Y_i(C)|\pi_i(C) = \pi_{0,i}\right] - \mathrm{E}\left[Y_i(P)|\pi_i(P) = \pi_{1,i}\right] \neq 0\right\} = \beta. \end{gathered} \tag{5.5}$$

The parameter $\beta \in [0, 1]$ is related to $\alpha \in [0, 1]$ through a competitors policy change function, $\beta = f_\pi(\alpha) \in [0, 1]$. Similarly, the effect $\text{ATE}_{Market,i}$ is related to β based on a competitors effect function, $\text{ATE}_{Market,i} = f_{ATE}(\beta) \in [-1, 1]$. Some special cases include (proof of these cases is trivial based on Eq 5.5):

- If $\alpha = 0 \Rightarrow \beta = f_\pi(0) = 0 \Rightarrow \text{ATE}_{Market,i} = 0$: average competitors policy is not affected by the campaign.
- If $\alpha > 0 \wedge \beta = f_\pi(\alpha) = 0 \Rightarrow \text{ATE}_{Market,i} = 0$: competitors advertising will not have any effect on Y_i.
- If $\beta = f_\pi(\alpha) > 0 \Rightarrow \alpha > 0$: a competitors effect greater than zero is likely only if the focal campaign is likely to affect their average ad delivery policy
- $\beta = f_\pi(\alpha) > 0 \iff \text{ATE}_{Market,i} \neq 0$: an average campaign presence effect implies a non-zero probability of competitors effect on Y_i and vice versa.

5.B The Cost of the Randomized Design

We analyze the cost of the proposed design of Fig 5.3(c) where no placebo ad is displayed, and $Z_i \in \{\text{Control}, \text{Study}\} = \{C, S\}$. Let N_{Exp} be the number of users for whom the opportunity to advertise is won. For the control group, there is a potential revenue loss, proportional to the campaign effect value ($Val(\text{ATE}_{Camp,i})$), if these users were exposed to the campaign. Because no ad impression is displayed to these users, a campaign budget surplus remains from not displaying these ads ($Cost(\text{AdDisplay})$). Thus, for all visiting users N_T, the design cost ($Cost(\text{Design})$) becomes:

$$Cost(\text{Design}) = P(Z_i = C) \times [N_T \times Val(\text{ATE}_{Camp,i}) - N_{Exp} \times Cost(\text{AdDisplay})] . \quad (5.6)$$

If this budget surplus is used to display campaign ads to a larger population in the study group, we have $\text{ATE}^{\Delta}_{Ad,i}$ and $\text{ATE}^{\Delta}_{Market,i}$ to be the average campaign effects on these additional exposed users. As a result, substituting Eq 5.3 in Eq 5.6 and given $N_{Exp} = P(D_i = 1) \times N_T$, the design cost ($Cost(\text{Design}^{\Delta})$) results into:

$$\begin{aligned} Cost(\text{Design}^{\Delta}) = \quad & P(Z_i = C) \times N_{Exp} \\ & \times Val([\text{ATE}_{Ad,i} + \text{ATE}_{Market,i}] - [\text{ATE}^{\Delta}_{Ad,i} + \text{ATE}^{\Delta}_{Market,i}]). \end{aligned} \tag{5.7}$$

Let $[\text{ATE}_{Ad,i} + \text{ATE}_{Market,i}] - [\text{ATE}^{\Delta}_{Ad,i} + \text{ATE}^{\Delta}_{Market,i}] = \epsilon$. Given an optimal user targeting policy, where the users with highest potential causal impact are most likely to be targeted, then $\epsilon > 0$ and $\epsilon << \text{ATE}_{Ad,i} + \text{ATE}_{Market,i}$. Therefore, the cost of experimentation is reduced to a *function of a small number*.

Chapter 6

Campaign Local Effect on the Targeted Users: Evaluating User Targeting Business Models

6.1 Introduction and Problem Context

In online advertising, user targeting is one of the most important decisions in running a campaign. A survey of 100 marketers, agencies, and media planners indicates that survey respondents perceive the user targeting and the campaign optimization capabilities as the main differentiators among ad networks [64]. This importance has motivated the development of online conversion attribution methods by the industry. These methods include: Last-Touch Attribution (LTA) and Multi-Touch Attribution (MTA). LTA assigns the conversion credit to the last campaign exposure (touch point) to a user in the path to conversion. Similarly, MTA gives credit to the a set of touch points in this path [3]. As a result, the deployment of cost-per-action (CPA) campaigns, which generate revenue to ad networks based on these attribution practices, has produced increasingly sophisticated targeting engines. These targeting engines mostly aim to display ads to converting users [66, 1]. However, these practices do not guarantee a causal impact optimization and incentivize the targeting of baseline users [14], those who convert regardless of the touch point (*always-buy* users).

Ad exposures are often considered to be a consequence of user activity [22, 55], or even a potential "coincidence" [87] in the ad effectiveness literature. However, in reality these decisions are frequently optimized by the targeting engine. Although the recent literature has addressed the evaluation of focused targeting practices using field experiments [54, 45], the attention to assessing the user selection effect of standard targeting engines for CPA campaigns has been minimal.

6.2 Chapter Contribution

Given the randomized design of Fig 5.3(c) of Chapter 5, the estimation of the campaign attribution is straightforward (Eq 6.3). However, by Remark 2 the conversion probability of the users of the control group who would be targeted needs to be inferred.

We develop a methodology to estimate the user conversion probability of the users in the control group who are statistically equivalent to those targeted in the study group using Potential Outcomes causal model and Principal Stratification [75, 36] [1]. The proposed

[1] Although other causal frameworks have been developed, mainly the Structural Equation Model [68] and Econometric Causality [2, 48], we approach the problem using Potential Outcomes. This framework allows us to model post-treatment variables with the use of the experimental data formally. There has been a long

approach allows us to estimate the effect of the campaign presence in the marketplace analyzed in Chapter 5. Also, we estimate the local treatment effect of the campaign on the targeted users and characterize this campaign targeting in terms of influenceable user classes [26].

Compared to prior literature that evaluates focused targeting practices [54, 45], we analyze the aggregated targeting performance of CPM and CPA campaigns. By comparing the probability of targeting *always-buy* users, we find evidence supporting the hypothesis that CPA campaigns incentivize the targeting of these users [14]. This evidence raises questions concerning the external validity of ad effects estimated in a standard evaluation CPM experiment to the CPA campaign deployment scenario.

6.3 Chapter Assumptions

In this section, we state the main assumptions used in the current chapter. The results and conclusions of this chapter are valid to the extent that these assumptions are applicable.

1. We refer to tracking cookies as users in the experimental design and estimation. We consider stable user cookies born before the campaign starts and that are active in the entire ad network.

2. The treatment assignment is assumed to be independent of the treatment effect, i.e. random assignment. This independence condition is a requirement for the methodology developed in this Chapter.

debate in the comparison between causal models, which falls outside the scope of this analysis and is not discussed here.

3. Stable Unit Treatment Value Assumption (SUTVA). We assume that the treatment status of any user does not affect the potential outcomes of the other users, i.e. no interference between users is assumed.

4. We model the user conversions and ad exposures as a random events, with predetermined probabilities for each treatment arm. Thus, the converting users are conditionally independent of each other given a predetermined probability, and the ad-exposed users are conditionally independent of each other given a predetermined probabilities.

5. The user ad exposure is assumed to be binary (targeted or non-targeted) without considering the number ad exposures. Similarly, the visiting user indicator and the converting user indicator do not consider the multiple instances of these events for a given user (see table 6.5).

6. Only the targeted users, those who are exposed to the ad, are subject to the campaign effect as illustrated by Fig 6.1. Section 6.5.2 validates this assumption.

7. The probabilities of positively affected users and negatively affected users are the same (non-zero value) for the non-targeted population.

8. Model parameters are assumed to be random, with standard Jeffrey's conjugate prior distribution. Indicator random variables are assumed to be Bernoulli distributed with prior distribution: Beta(0.5,0.5).

9. The above model components represent the structure we assumed in this Chapter. Eq 6.2 illustrates the joint distribution between the random variables of this structure.

Table 6.1: Description of the variables used in chapter 6.

Term	**Description**
Z∈{C,P,S}	Random treatment assignment
Y∈{0,1}	Converting user indicator
D∈{0,1}	Targeted user indicator
W∈{0,1}	Ad exposure indicator
$\theta_{dz} \in [0, 1]$	Probability of $Y = 1$ given $D = d, Z = z$
$p_{sel} \in [0, 1]$	Probability of $D = 1$
$\theta^{(s)}_{0z}, p^{(s)}_{sel,z} \in [0, 1]$	Parameters obtained by repeated randomization for validation
$\Delta^{(s)}_{psel}, \Delta^{(0)}_{\theta 0} \quad \in [-1, 1]$	Difference statistics between repeated randomized groups
$i \in \mathbb{N}$	Variable index for the i-*th* user
$\mathrm{N}^{y}_{dz} \in \mathbb{N}$	User count given $Y = y, D = d, Z = z$
N_{obs},N_{samp}	Observed/sampled count sets
N_{burnin},$\mathrm{N}_{s} \in \mathbb{N}$	Burn-in/Gibbs number of samples
$a_0, b_0 \in (0, \infty]$	Beta prior parameters
U∈{Per$^+$,Per$^-$, AB,NB}	Influenceable user category indicator defined by Eq 6.8
Θ	Parameter Set of Eq 6.2: $\{\theta_0, \theta_{1C}, \theta_{1S}, p_{sel}\}$

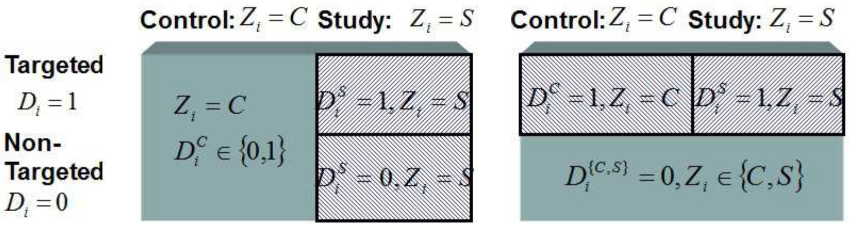
6.4 Estimation Methodology

6.4.1 Campaign Causal Effect on the Users Exposed to the Ad

The Potential Outcomes Causal Model analyzes the potential individual outcomes for each of the treatments [75]. For two treatment arms, this framework implies that half

Table 6.2: Performance metrics analyzed in chapter 6. Lifts are in the range: $\in [-1, \infty)$. Average Treatment Effects (ATE) are in the range: $\in [-1, 1]$.

Metric	**Lift**	**Description**
ATE_{Camp}	lift_{Camp}	Overall campaign average effect on all visiting users
LATE_{Ad}	lift_{ad}	Local average treatment effect of the campaign on targeted users
SelEff	lift_{sel}	User selection effect introduced by the targeting engine
$P(D = 1 \vert U)$	Probability of targeting user influenceable category U	
ATRB_{Camp}	Campaign attributed converting users, with respect to $\text{N}^1_{0S}+\text{N}^1_{1S}$, estimated based on ATE_{Camp} (left of Eq 6.6)	
ATRB_{Ad}	Campaign attributed converting users, with respect to $\text{N}^1_{0S}+\text{N}^1_{1S}$, estimated based on LATE_{Ad} (right of Eq 6.6)	



(a) (b)

Figure 6.1: User segments based on control/study (Z_i) and non-selected/selected (D_i) groups. (a) Observed segments. (b) Idealized segments to estimate the campaign effects on the targeted users.

of the data is missing because we can never observe a unit response in both arms. If the treatment assignment is independent of the treatment effect (i.e. random assignment), then the causal estimates are unbiased. A necessary assumption of this causal model is the Stable

Table 6.3: Observed user counts based on the user potential outcomes. N^y_{dz}, where $D_i = d$, $Z_i = z$, $Y_i = y$, are user counts for the given values of Y, Z, D. Missing values are presented as *.

User	**Potential Outcomes**				**Treatment**	**Principal**	
Counts	Control		Study		**Assignment**	**Stratum**	
N^y_{dz}	$W_i(C)$	$Y_i(C)$	$W_i(S)$	$Y_i(S)$	Z_i	$(W_i(C), W_i(S))$	D_i
$N^0_{\{0,1\}C}$	0	0	*	*	C	(0,*)	*
$N^1_{\{0,1\}C}$	0	1	*	*	C	(0,*)	*
N^0_{0S}	0	*	0	0	S	(0,0)	0
N^1_{0S}	0	*	0	1	S	(0,0)	0
N^0_{1S}	0	*	1	0	S	(0,1)	1
N^1_{1S}	0	*	1	1	S	(0,1)	1

Unit Treatment Value Assumption (SUTVA), which implies that the treatment status of any unit does not affect the potential outcomes of the other units. Thus, the user responses are exchangeable given the induced treatment effect.

The Principal Stratification modeling provides a framework to estimate treatment effects conditional on post-treatment (non-ignorable) variables, which might be affected by the treatment [36]. The key element in this context is the identification of user classes, or strata, with equal treatment effects and probability of treatment assignment. Given the proposed randomized design of Fig 5.3(c), where $Z_i \in \{\text{Control}, \text{Study}\} = \{C, S\}$, the user exposure to the ad is a post-treatment variable. Here the targeting process is performed for the study group and is not for the control group [2]. Let W_i indicates if a user is exposed

[2] Note that the current analysis holds for the randomized design of Fig 5.3(d) as well if only the control

to the ad ($W_i = 1$) or not ($W_i = 0$). To define the principal strata, we model the potential outcomes for $W_i(C)$, $W_i(S)$. Since the ad is never shown to the users of the control group ($W_i(C) = 0$), the user principal strata, W_i^P, are defined as follows:

$$W_i^P = \left\{ \begin{pmatrix} W_i(C) \\ W_i(S) \end{pmatrix} \right\} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, \quad D_i = \begin{cases} 0 & \text{if } W_i^P = (0,0)' \\ 1 & \text{if } W_i^P = (0,1)' \end{cases} \quad . \tag{6.1}$$

Table 6.3 depicts the observed and missed data in the potential outcomes notation. This definition guarantees that the selection effect in the control group is the same as that of the study group (ignorable), $D_i(C) = D_i(S) = D_i$, based on the potential conversion probability (Assumption 1 of Chapter 5). D_i indicates whether the user is targeted had he/she been assigned to the study group (*targeted-if-assigned*, $D_i = 1$), or not (*never-targeted*, $D_i = 0$). Consequently, we do not observe D_i in the control group, which is illustrated by Fig 6.1(a).

We define the probability of D_i to be Bernoulli distributed with parameter p_{sel}, and the probability of user conversion Y_i to be Bernoulli distributed with parameters θ_{dz} for the four combinations $D_i = d$, $Z_i = z$, and $Y = \{Y_i : \forall i\}$, $Z = \{Z_i : \forall i\}$, $D = \{D_i : \forall i\}$. Let the targeted user indicator for those assigned to the control arm be D_i^C, and for those assigned to the study arm be D_i^S. Therefore, assuming $\Theta = \{\theta_{dz}, p_{sel} : \forall d \in \{0,1\}, \forall z \in \{C,S\}\}$ are random variables, we have:

$$P(Y, Z, D, \Theta) = P(\Theta) \prod_{\forall i} P(D_i = d | p_{sel}) P(Y_i(Z_i) | D_i = d, Z_i = z, \theta_{dz}) P(Z_i = z) \ . \tag{6.2}$$

One concern of the model of Eq 6.2 is that distribution parameters $\{\theta_{0C}, \theta_{1C}\}$ are not identifiable, and a constraint based on $\{\theta_{0S}, \theta_{1S}\}$ needs to be imposed. That is, for given values of θ_{0C} and θ_{1C} the same likelihood value is produced if we switch these parameter

and study arms are analyzed.

values. Given that the randomized assignment is independent of the potential outcomes of *never-selected* users ($Y_i \perp Z_i | D_i = 0$), we do not consider any campaign effect on this sub-population as depicted by Fig 6.1(b) (Assumption 2)[3]. This constraint ensures that the model parameters are identifiable and leads to: $\theta_{0S} = \theta_{0C} = \theta_0, \Theta = \{\theta_0, \theta_{1C}, \theta_{1S}, p_{sel}\}$.

We note that in a sequential setting, the targeting engine employs the user conversion data to estimate the user targeting likelihood of the next visiting user: $P(D_{+i}|Y, D)$. However, based on SUTVA and the predefined probability of targeting assumed, the user targeting indicator is exchangeable, $P(D_i|p_{sel}) : \forall i$.

The inference objective of the joint distribution of Eq 6.2 is to estimate the posterior distribution of the parameters Θ given the observed data from Table 6.3. Calculating this posterior distribution in closed form is intractable because D^C must be observed. Thus, we implement a Markov Chain Monte Carlo (MCMC) based approach using Gibbs sampling depicted by Algorithm 3. We denote the set of observed counts as N_{obs} (step 1). Given an initial guess for Θ^0 (step 4), we sample D^C and estimate the counts N^y_{dC}: $\forall d \in \{0, 1\}, \forall y \in \{0, 1\}$ based on the probability of D_i^{Cy} (steps 6-8). We denote these sampled counts as $N_{samp} = \left\{N^y_{dC} : \forall d \in \{0, 1\}, \forall y \in \{0, 1\}\right\}$ (step 2). Given the augmented user counts, $\{N_{obs}, N_{samp}\}$, we sample each parameter of Θ conditional on $\Theta_{-\theta}$, which is the set Θ without θ (steps 9-11). The sampling distributions of the parameters $\{\theta_0, \theta_{1C}, \theta_{1S}, p_{sel}\}$, are Beta$(a_0, b_0)$ distributions with a Jeffreys conjugate prior parameters, $\{a_0 = 0.5, b_0 = 0.5\}$ (step 3). We test other prior parameters in Appendix 6.A. This sampling process is repeated for $N_{burnin} + N_s$ times (steps 5-12). After discarding a set of burn-in samples, N_{burnin}, a set

[3] This constraint is also known as *exclusion restriction* in the context of the causal inference based on randomized experiments with non-compliance [50].

Algorithm 3 Gibbs Sampling Algorithm based on the joint distribution of Eq. 6.2

```
Input: $N_{obs} = \Big\{N^y_{dS}, N^y_{\{0,1\}C} : \forall d \in \{0,1\}, \forall y \in \{0,1\}\Big\}$ from Table 6.3
Define $N_{samp} = \big\{N^y_{dC} : \forall d \in \{0,1\}, \forall y \in \{0,1\}\big\}$
Set $a_0 = 0.5, b_0 = 0.5$
Initial guess $\Theta^0 = \{\theta_{1z}, \theta_0, p_{sel}\}^0, \quad \forall z \in \{C, S\}$
for $i \leftarrow 1$ to $N_{burnin} + N_s$ do
    Set $P(D_i^{Cy} = 1|\Theta, N_{obs}) = \frac{p_{sel}(\theta_{1C})^y(1-\theta_{1C})^{(1-y)}}{p_{sel}(\theta_{1C})^y(1-\theta_{1C})^{(1-y)} + (1-p_{sel})(\theta_0)^y(1-\theta_0)^{(1-y)}}$, $\forall y \in \{0,1\}$
    Draw $N^y_{1C}|\Theta, N_{obs} \sim \text{Binomial}\Big(N^y_{\{0,1\}C}, P(D_i^{Cy} = 1|\Theta, N_{obs})\Big), \quad \forall y \in \{0,1\}$
    Set $N^y_{0C} = N^y_{\{0,1\}C} - N^y_{1C}, \quad \forall y \in \{0,1\}$
    Draw $\theta^{(i)}_{1z}|\Theta_{-\theta_{1z}}, N_{samp}, N_{obs} \sim \text{Beta}\big(a_0 + N^1_{1z}, b_0 + N^0_{1z}\big), \quad \forall z \in \{C, S\}$
    Draw $\theta^{(i)}_0|\Theta_{-\theta_0}, N_{samp}, N_{obs} \sim \text{Beta}\big(a_0 + N^1_{0C} + N^1_{0S}, b_0 + N^0_{0C} + N^0_{0S}\big)$
    Set $a_{psel} = a_0 + \sum_{\forall z \in \{C,S\}, \forall y \in \{0,1\}} N^y_{1z}, \quad b_{psel} = b_0 + \sum_{\forall z \in \{C,S\}, \forall y \in \{0,1\}} N^y_{0z}$
    Draw $p^{(i)}_{sel}|\Theta_{-p_{sel}}, N_{samp}, N_{obs} \sim \text{Beta}(a_{psel}, b_{psel})$
end for
return $\Theta^{N_{burnin}+1:N_{burnin}+N_s}$
```

of samples of the posterior distribution is obtained, $\Theta^{1:Nsamples}$. These samples $\Theta^{1:Nsamples}$ are used to estimate the variability (or heterogeneity) of the local campaign effect and lift of Eq 6.4 below, and the targeting analysis of Section 6.4.3.

Remark 4. *We use the power of user randomization to estimate the conversion probability of the statistically equivalent users in the control group, to those targeted in the study group, without relying on user features. We take advantage of the fact that there is no campaign* *effect on the non-targeted users. Given randomized user treatment assignments, the proportion of users statistically equivalent to those targeted in the study group must be the same in both treatment groups for large populations in average. Therefore, the proposed model guarantees that the targeting and the conversion probabilities are balanced for both control and study groups.*

6.4.2 Campaign Effect Estimation

We estimate the average campaign treatment effect (ATE_{Camp}) on the overall visiting users and the lift (lift_{Camp}) as follows:

$$\text{ATE}_{Camp} = \text{E}(Y_i(S)|Z_i = S) - \text{E}(Y_i(C)|Z_i = C), \quad \text{lift}_{Camp} = \frac{\text{ATE}_{Camp}}{\text{E}(Y_i(C)|Z_i = C)}. \tag{6.3}$$

Assuming a Jeffreys conjugate prior distribution, $\{a_0 = 0.5, b_0 = 0.5\}$, the posterior distribution becomes $\text{Beta}(a_0 + N_z^1, b_0 + N_z^0)$ where N_z^1, N_z^0 are the number of converting and non-converting users of the z group. We sample from these posterior distributions to provide credible intervals for both ATE_{Camp} and lift_{Camp}.

The local average treatment effect by the campaign on the targeted users (LATE_{Ad}, and the lift (lift_{ad}) are estimated from the posterior distribution of Θ as follows:

$$\begin{aligned} &\text{LATE}_{Ad} = \text{E}(Y_i(S)|D_i = 1, Z_i = S) - \text{E}(Y_i(C)|D_i = 1, Z_i = C), \\ &\text{LATE}_{Ad} = \theta_{1S} - \theta_{1C}, \qquad\qquad \text{lift}_{ad} = \frac{\theta_{1S} - \theta_{1C}}{\theta_{1C}}. \end{aligned} \tag{6.4}$$

Based on the samples $\Theta^{(1:N_s)}$ obtained by the Gibbs sampling procedure of Section 6.4.1, credible intervals are estimated from the set $\{\text{LATE}_{Ad}, \text{lift}_{ad}\}^{(1:N_s)}$. The analysis of Section 5.4.3 and Eq 5.3 leads to:

$$\text{ATE}_{Camp} = P(D_i = 1) \times \{\text{ATE}_{Ad} + \text{ATE}_{Market}\} = P(D_i = 1) \times \text{LATE}_{Ad}. \tag{6.5}$$

Here, $\text{ATE}_{Ad,i}$ is the ad effect that would be estimated using ad placebos, and $\text{ATE}_{Market,i}$ is the effect of the campaign presence in the marketplace. Therefore, LATE_{Ad} is the campaign effect on the targeted users of the study group.

We estimate the proportion of converting users attributed to the campaign with respect to those in the study group, based on ATE_{Camp} and LATE_{Ad} (ATRB_{Camp}, ATRB_{Ad}):

$$\text{ATRB}_{Camp} = \text{ATE}_{Camp} \times \frac{\sum_{\forall y \in \{0,1\}, \forall d \in \{0,1\}} N_{dS}^{y}}{N_{0S}^{1} + N_{1S}^{1}}, \quad \text{ATRB}_{Ad} = LATE_{Ad} \times \frac{N_{1S}^{0} + N_{1S}^{1}}{N_{0S}^{1} + N_{1S}^{1}}. \tag{6.6}$$

Given that the campaign impacts only the targeted users, these metrics have to match. These attribution metrics provide the campaign value in terms of causally generated user conversions and represent the output of the measurement block of Fig 5.1.

6.4.3 User Targeting Characterization

To characterize the user targeting of converting users performed by the targeting engine, we estimate the user selection effect (SelEff) and the lift (lift_{sel}) as follows:

$$\begin{aligned} &\text{SelEff} = \text{E}(Y_i(C)|D_i = 1, Z_i = C) - \text{E}(Y_i(C)|D_i = 0, Z_i = C), \\ &\text{SelEff} = \theta_{1C} - \theta_0, \qquad \text{lift}_{sel} = \frac{\theta_{1C} - \theta_0}{\theta_0}. \end{aligned} \tag{6.7}$$

Note that targeting converting users, whose performance is measured by SelEff, is a common objective of the targeting engine [66]. The business model in the industry for CPA campaigns, namely last-touch and multi-touch attribution [3], incentivizes this targeting practice. Thus, being part of a converting user path is enough to attribute credit to the campaign.

To characterize the causal user targeting process, we partition the users into four influenceable categories [26], U_i as follows: Per$^+$, positively influenced user, *persuadable*; Per$^-$, negatively influenced user, *anti-persuadable*; AB, converting user with no effect, *always-buy*; NB, non-converting user with no effect, *never-buy*. Given the targeting indicator D_i, the probability of a user category U_i is defined as:

$$\begin{aligned}
&P(U_i = \text{Per}^+|D_i, \Theta) \propto \; P(Y_i(S) = 1|D_i, Z_i = S, \Theta)P(Y_i(C) = 0|D_i, Z_i = C, \Theta),\\
&P(U_i = \text{Per}^-|D_i, \Theta) \propto P(Y_i(S) = 0|D_i, Z_i = S, \Theta)P(Y_i(C) = 1|D_i, Z_i = C, \Theta),\\
&P(U_i = \text{AB}|D_i, \Theta) \propto P(Y_i(S) = 1|D_i, Z_i = S, \Theta)P(Y_i(C) = 1|D_i, Z_i = C, \Theta),\\
&P(U_i = \text{NB}|D_i, \Theta) \propto P(Y_i(S) = 0|D_i, Z_i = S, \Theta)P(Y_i(C) = 0|D_i, Z_i = C, \Theta).
\end{aligned} \tag{6.8}$$

Since Y_i and D_i are Bernoulli distributed, we have:

$$P(Y_i(Z_i) = y|Z_i, D_i, \Theta) = \theta_{dz}^{y}(1 - \theta_{dz})^{1-y}, \quad P(D_i = d|, \Theta) = p_{sel}^{d}(1 - p_{sel})^{1-d}. \tag{6.9}$$

We estimate the probability of targeting a user given U_i by Bayes theorem as follows:

$$P(D_i = 1|U_i, \Theta) = \frac{P(D_i = 1|\Theta)P(U_i|D_i = 1, \Theta)}{\sum_{\forall d \in \{0,1\}} P(D_i = d|\Theta)P(U_i|D_i = d, \Theta)}. \tag{6.10}$$

This estimation provides the basis to characterize the targeting engine[4].

Remark 5. *We estimate the probabilities of* persuadable*,* anti-persuadable*,* always-buy*, and* never-buy *user categories, despite not using user features, because we observe the counterfactual user response in both control and study treatment groups. We incorporate user features to characterize these categories in Chapter 7.*

[4] Since the campaign effect is not considered for the non-targeted users, $P(U_i$ $=$Per$_i^+|D_i$ $=$ $0)$=$P(U_i$ =Per$_i^-|D_i = 0)$.

6.5 Results

In this Section, we discuss the data collection and processing and validate the model assumptions based on user randomization. We then present the analysis of two CPA campaigns (Fig 5.3(c) design[5]). Finally, the user targeting is analyzed for these two CPA campaigns, and the CPM campaign analyzed in Chapter 5 (Table 5.3), based on the targeting policy they executed.

6.5.1 Data Collection and Description

We ran two large-scale randomized (or field) experiments (Fig 5.3(c) design) collaboratively with two European advertisers in the mobile communications and the public transportation service sectors. The user targeting was optimized in real time by a sophisticated targeting engine that valued the user and managed the bidding process for both CPA campaigns. User conversions were economically equivalent for both campaigns. We are not at liberty to disclose the ad content or the identity of the advertiser.

We randomly assigned the visiting users using the last two digits of the timestamp their cookies were born. This rule separated the users and kept them in their assigned group while the campaign was active and is validated in Section 6.5.2. We only consider those users whose cookies were born before the campaign started and remained active in the ad network[6]. We perform this selection to avoid user contamination and guarantee that we do not miss user tracking due to cookie deletion

[5] We present a power analysis of the campaign effect estimation in Appendix 6.B. This study illustrates the difficulty in measuring this effect in Targeted Advertising even when even when tens of millions of users are part of the experiment.

[6] We assume that the cookie deletion event is independent of the campaign effect (ignorable or exogenous). Thus, no bias is introduced by focusing on users with stable cookies.

Table 6.4: Campaign data based on notation of Table 6.3. Duration for CPA Campaign 1: 30 days, CPA Campaign 2: 28 days.

Count	$N^0_{\{0,1\}C}$	$N^1_{\{0,1\}C}$	N^0_{0S}	N^1_{0S}	N^0_{1S}	N^1_{1S}
Campaign 1	1,560,146	400	12,010,058	2,387	5,708,558	2,599
Campaign 2	2,803,640	734	18,681,097	3,170	2,584,728	2,685

Table 6.5: User activity statistics for the campaigns of Table 6.4. Mean and standard deviation (Std) are displayed. **Visits/user** is the number of visits per user. **Convs**$|Y_i = 1$ is the number of conversions per converting user. **Imps/user** is the number of ad exposures per targeted user ($D_i = 1$).

	Campaign 1					
	$Z_i = C$		$Z_i = S$		$Z_i = S, D_i = 1$	
Variable	Mean	Std	Mean	Std	Mean	Std
Visits/user	36.25	162.21	36.49	175.98	83.50	332.74
Convs$\|Y_i = 1$	1.14	0.40	1.19	0.59	1.19	0.59
Imps/user	-	-	-	-	3.47	8.41
	Campaign 2					
	$Z_i = C$		$Z_i = S$		$Z_i = S, D_i = 1$	
	Mean	Std	Mean	Std	Mean	Std
Visits/user	37.32	218.13	37.16	223.23	160.93	637.71
Convs$\|Y_i = 1$	1.32	0.73	1.33	0.84	1.35	0.87
Imps/user	-	-	-	-	2.63	5.71

Table 6.6: Validation of model conditions expressed by Fig 6.1(b). Testing procedure is detailed by Eq 6.11. Results reported based on 90% credible intervals. {*Low, Med, High*} are the {0.05, 0.5, 0.95} quantiles

	Campaign 1			**Campaign 2**		
	Low	Med	High	Low	Med	High
Δ_{psel}(1e-3)	-2.65	0.02	2.71	-0.89	0.01	0.91
	Campaign 1			**Campaign 2**		
	Low	Med	High	Low	Med	High
$\Delta_{\theta 0}$(1e-5)	-1.71	0.03	1.70	-1.22	0.02	1.23

Given a user timeline of events, we focus on those events recorded after the first visit to any publisher website where the ad was potentially displayed. We mark the user as targeted in the study group ($Z_i = S, W_i = 1, D_i = 1$) if at least one ad exposure was recorded (otherwise $W_i = 0, D_i = 0$). If one conversion has been registered after one ad exposure, and before the campaign ended, the user is considered to be targeted and converter ($W_i = 1, D_i = 1, Y_i = 1$). No ad exposure was performed for the users in the control group ($Z_i = C, W_i = 0$), and consequently the user targeting indicator is missing ($D_i = *$). The user counts based on the notation of Table 6.3 are depicted by Table 6.4, and Table 6.5 shows user activity statistics.

6.5.2 Model Validation: Randomization

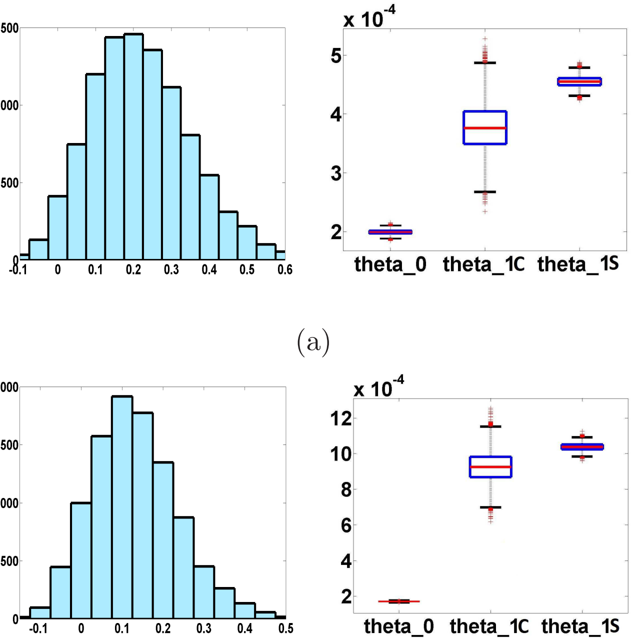
The estimation methodology of Section 6.4.1 relies on user randomization, and the condition of no campaign effects on the non-targeted users illustrated by Fig 6.1(b). To test these conditions, we analyze the CPA campaigns of Table 6.4 where the design of Fig 5.3(c) is implemented. We randomly partition the users of the study group into two sub-groups, where the targeting indicator D_i^S is observed. This process generates both simulated control ($Z_i = C$) and study ($Z_i = S$) groups, where D_i^C and D_i^S are observed. We define $p_{sel,z}$ to be the targeting probability, p_{sel}, for the z random group. We perform this partition $3,000$ times, obtain the method-of-moments estimate for $\{p_{sel,z}^{(s)}, \theta_{0z}^{(s)}\}$, and calculate:

$$\Delta_{psel}^{(s)} = p_{sel,S}^{(s)} - p_{sel,C}^{(s)}, \quad \Delta_{\theta 0}^{(s)} = \theta_{0S}^{(s)} - \theta_{0C}^{(s)}. \tag{6.11}$$

Zero values for Δ_{psel} and $\Delta_{\theta 0}$ verify the conditions of the model (user selection Assumptions 1 and 2 of Chapter 5), and the randomization procedure. Table 6.6 reports the credible intervals for these statistics, and shows that they are centered at 0 for both campaigns. Therefore, we conclude that p_{sel} is equal for both treatment arms, and that no campaign effect is present in the non-targeted users. We also conclude that the user randomized assignment is independent of the treatment effects and provides the basis for the effect estimations to be causal.

6.5.3 Campaign Effect Results

Fig 6.2 depicts the estimation results for the CPA campaigns of Table 6.4. Here, we use $N_{burnin} = 2,000$ burn-in iterations and $N_s = 10,000$ samples for the Gibbs sampling framework of Algorithm 3. As illustrated, the posterior distribution for lift_{ad} is skewed because lift_{ad} is a ratio of random variables. The posterior distributions for $\{\theta_0, \theta_{1C}, \theta_{1S}\}$ are illustrated by the boxplots of Fig 6.2(a) and (b). A significant difference is evident between the conversion rates for the targeted (θ_{1C}, θ_{1S}) and the non-targeted (θ_0) groups,



(a)

(b)

Figure 6.2: Model fitting results for: (a) Campaign 1, (b) Campaign 2. From left to right, posterior distribution for lift_{ad}, and the box plot for θ_0, θ_{1C}, θ_{1S} where y-axis is the conversion probability.

which is measured by SelEff and lift_{sel} of Eq 6.7. As indicated by Table 6.7, we obtain a median $\text{lift}_{sel} = \{89\%, 444\%\}$ for Campaign 1 and 2 respectively.

For comparison purposes, we estimate the campaign effect on the targeted users by assuming that we do not observe the control group response, LATE_{Ad}^{I2C}. This naive effect estimation is used by last-touch or multi-touch attribution when only the focal campaign is run (single channel). Similarly, we estimate the campaign effect without correcting for post-treatment bias (or its endogeneity), LATE_{Ad}^{post}. These effects are defined as follows:

$$\begin{aligned}\text{LATE}_{Ad}^{I2C} &= \text{E}[Y_i|W_i(S) = 1, Z_i = S] - \text{E}[Y_i|W_i(S) = 0, Z_i = S],\\ \text{LATE}_{Ad}^{post} &= \text{E}[Y_i|W_i(S) = 1, Z_i = S] - \text{E}[Y_i|W_i(C) = 0, Z_i = C].\end{aligned} \tag{6.12}$$

Table 6.7: Attribution results based on 90% credible intervals. {*Low, Med, High*} are the {0.05, 0.5, 0.95} quantiles.

	Campaign 1			**Campaign 2**		
	Low	Med	High	Low	Med	High
LATE_{Ad}^{I2C}(1E-4)	2.40	2.56	2.73	8.34	8.68	9.01
lift_{ad}^{I2C}(%)	-	129	-	-	**511**	-
LATE_{Ad}^{post}(1E-4)	1.73	1.99	2.24	7.39	7.76	8.13
lift_{ad}^{post}(%)	-	77.54	-	-	**296**	-
ATE_{Camp}(1E-5)	0.23	2.49	4.64	-0.37	1.35	3.01
lift_{Camp}(%)	0.84	9.71	19.61	-1.35	5.15	12.19
LATE_{Ad}(1E-5)	0.85	7.90	14.48	-3.21	11.42	25.49
lift_{ad}(%)	1.89	**21.04**	46.33	-3.00	12.36	32.43
SelEff(1E-5)	1.12	1.77	2.48	6.16	7.55	8.97
lift_{sel}(%)	55	**89**	126	359	**444**	534
ATRB_{Camp}(%)	0.85	**8.90**	16.51	-1.36	**4.91**	10.96
ATRB_{Ad}(%)	0.96	**9.05**	16.59	-1.42	**5.05**	11.26

Table 6.7 shows the campaign effects on the overall user population, ATE_{Camp}, and on the targeted population, LATE_{Ad}. Here, the zero effect is not included in the 90% credible intervals for Campaign 1. Campaign 2 is leaning towards positive values but with a small negative range in the credible interval. In addition, we observe variations of less than 0.2% between median ATRB_{Ad} and ATRB_{Camp}: $\{9.05\%, 8.90\%\}$ for Campaign 1; $\{5.05\%, 4.91\%\}$ for Campaign 2. This result shows consistency between LATE_{Ad} and ATE_{Camp}, and confirms the campaign effect analysis of Section 5.4.3. We note the severe

Table 6.8: User selection median probabilities based on Eqs 6.8–6.10. Campaign 1 and Campaign 2 are CPA optimized campaigns (Table 6.4). Campaign 3 is CPM non-optimized campaign (Table 5.3).

	Campaign 1	**Campaign 2**	**Campaign 3**
$P(D_i = 1 \mid U_i = \text{Per}^+)$	**0.5211**	0.4583	**0.3276**
$P(D_i = 1 \mid U_i = \text{Per}^-)$	0.4732	0.4296	**0.3497**
$P(D_i = 1 \mid U_i = \text{AB})$	**0.6728**	**0.8217**	**0.4180**
$P(D_i = 1 \mid U_i = \text{NB})$	**0.3221**	0.1215	0.2669
$P(U_i = \text{Per}^+ \mid D_i = 1)$	**4.55E-4**	**1.04E-3**	1.68E-4
$P(U_i = \text{Per}^- \mid D_i = 1)$	**3.76E-4**	**9.24E-4**	1.85E-4
$P(U_i = \text{AB} \mid D_i = 1)$	1.71E-7	9.59E-7	3.11E-8
$P(U_i = \text{NB} \mid D_i = 1)$	0.9992	0.9980	0.9996

over-estimation by the last-touch attribution effect (and by the effect without correcting for post-treatment bias) when compared with the causal lift (lift_{Ad}^{I2C} and lift_{Ad}^{post} vs lift_{Ad}); that is, for Campaign 1: 129% and 77.54% vs 21.04%; for Campaign 2: 511% and 296% vs 12.36%.

6.5.4 User Targeting Characterization Results

We analyze the user targeting process of the CPA campaigns of Table 6.4 (Campaign 1 and Campaign 2) based on the analysis of Section 6.4.3, and compare them with the targeting of the CPM campaign of Table 5.3 (Campaign 3). Table 6.8 shows the user targeting characterization results. The probability of *never-buy* users, is large in the targeted population ($P(U_i = \text{NB}|D_i = 1) > 0.99$ for all campaigns), which is a consequence of low conversion rates. Using Bayes theorem as in Eq 6.10, we observe that the probability of targeting a *never-buy* user is the lowest as there is no incentive to target this user category ($P(D_i = 1|U_i = \text{NB}) = \{0.32, 0.12, 0.27\}$ for Campaign {1,2,3} respectively). Similarly, the probability of targeting a *persuadable* user is significantly lower for the CPM Campaign 3 than for the CPA Campaigns 1 and 2 by as much as 37% ($0.52-0.33 = 0.19$ with respect to 0.52, where $P(D_i = 1|U_i = \text{Per}^+) = \{0.52, 0.46, 0.33\}$ for campaigns $\{1, 2, 3\}$ respectively), showing the positive effect of the targeting optimization.

As discussed in Section 6.4.3, lift_{sel} provides the conversion probability change in the targeted population (selection effect). CPA last-touch business model suggests that increasing this difference is beneficial for the overall campaign effect. We estimate that Campaign 2 (lift_{sel} = 444%) has a superior performance to Campaign 1 (lift_{sel} = 89%) under the CPA policy of targeting converting users. However, we estimate a significantly larger probability of targeting an *always-buy* user for Campaign 2 than for Campaign 1 ($P(D_i = 1|U_i = \text{AB}) = \{0.82, 0.67\}$ for campaigns {1,2} respectively). Campaign 2 is more effective in optimizing user conversions than Campaign 1 by a factor of five (444% vs. 89%). However, Campaign 2 is 22% ($0.82 - 0.67 = 0.15$ with respect to 0.67) more likely to target *always-buy* users. This analysis shows that the well-accepted policy of targeting users with the highest conversion probability does not necessarily improve the campaign value to the advertiser. Moreover, we estimate that this probability of targeting *always-buy* users is as much as 96% larger for CPA Campaign 2 when compared to the CPM Campaign 3 ($0.82 - 0.418 = 0.402$ with respect to 0.418). Therefore, the external validity (or extrapolation) of the ad effect estimated for a CPM campaign to a CPA campaign, assumed under the standard evaluation practice, is highly prone to inaccuracies. Because we observe that the targeted populations are fundamentally different.

6.6 Impact and Limitations

By characterizing the campaign targeted population of CPM and CPA campaigns in Section 6.5.4, we have demonstrated that the external validity of ad effects tested under CPM based targeting to CPA based targeting is inaccurate. This inaccuracy is because the targeted populations between these campaign business models are fundamentally different.

In CPA campaigns, the decision to target users is often driven by the user propensity to convert. As a result, we have found evidence showing that CPA campaigns incentivize the targeting of users who are going to buy in any case, which does not add value to the advertiser. On the other hand, purely non-optimized CPM campaigns are less effective than CPA campaigns to target users with positive effect.

The current analysis and results provide a potential opportunity to advertisers to act upon and improve the user targeting policy to optimize causal estimates. We develop a user targeting method to maximize the causal effect of the campaign in Chapter 7 diminishing the targeting of users who buy regardless of the ad exposure.

Appendix

6.A The Prior Probability and a Method of Moments: Robustness Checks

Given the Bayesian method of Section 6.4, we analyze the effect of different Beta prior parameters, and compare them with a method of moments which is derived now. Since D_i is observed for the study group, the estimation of p_{sel} and θ_{1S} in the study group is straightforward based on the method of moments. Similarly, θ_0 is approximated based on the observed conversions of the non-targeted users in the study group. As the observed conversion probability of the control group is a mixture of θ_0 and θ_{1C} weighted by $1 - p_{sel}$ and p_{sel} respectively, and $\{\theta_0, p_{sel}\}$ are shared by both arms (approximation), the estimation of θ_{1C} becomes:

$$\begin{gathered}\hat{p}_{sel} = \frac{N^1_{1S} + N^0_{1S}}{N^1_{1S} + N^0_{1S} + N^1_{0S} + N^0_{0S}}, \quad \hat{\theta}_{1S} = \frac{N^1_{1S}}{N^1_{1S} + N^0_{1S}}, \\ \hat{\theta}_0 = \frac{N^1_{0S}}{N^1_{0S} + N^0_{0S}}, \quad \hat{\theta}_{1C} = \frac{1}{\hat{p}_{sel}} \left[\frac{N^1_{\{0,1\}C}}{N^1_{\{0,1\}C} + N^0_{\{0,1\}C}} - \hat{\theta}_0 (1 - \hat{p}_{sel}) \right].\end{gathered} \tag{6.13}$$

This approach does not account for the data sample size and requires several approximations. Despite these limitations, we provide a robustness check based on this estimator. Table 6.9 compares this point estimator with the Bayesian method of Section 6.4 for different prior rates: $a_0/(a_0 + b_0)$; assuming a prior sample size: $a_0 + b_0 = 1$. Results show that more intuitive prior rate choices for low conversion rates {0.01,0.001}, do not affect results more than 0.9% in median lift$_{Ad}$ and its credible interval. We use the Jeffreys prior, $\{a_0 = 0.5, b_0 = 0.5\}$, because increasingly skewed prior distribution are more likely to be numerically unstable in the Gibbs sampling. The method of moments of Eq 6.13 shows

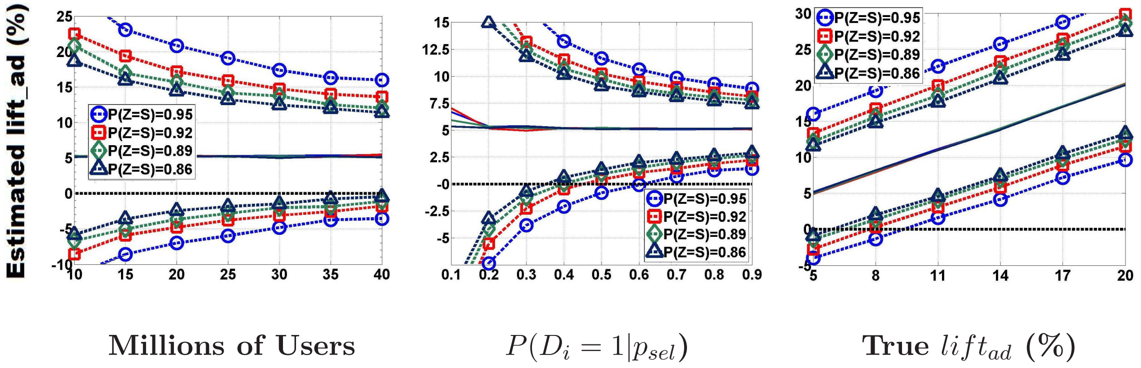
Table 6.9: Prior rate effect on the lift_{Ad} (%) estimation given a prior sample size: $a_0+b_0 = 1$, based on Algorithm 3, compared with the method of moments of Eq 6.13 (Moments). $N_{burnin} = 2,000$. $N_s = 10,000$. {*Low, Med, High*} are the {0.05, 0.5, 0.95} quantiles.

Prior Rate	**Campaign 1**			**Campaign 2**			**Campaign 3**		
$\frac{a_0}{a_0+b_0}$	Low	Med	High	Low	Med	High	Low	Med	High
0.5	2.15	21.15	46.51	-2.64	12.10	31.10	-19.09	-9.32	2.79
0.01	2.63	21.89	47.20	-2.26	12.99	31.86	-18.98	-9.01	3.10
0.001	2.51	21.53	47.55	-2.34	12.57	31.37	-19.58	-9.48	2.51
Moments	-	20.55	-	-	11.99	-	-	-9.59	-

discrepancies of less than 1% lift_{Ad} when compared with this prior choice.

6.B Estimation Power Analysis

In major firms, we have observed that the proportion of users used as the control group is typically determined based on the belief that large user populations are easily available by intuition. However, we show that poorly designed experiments lead to wide credible intervals containing the zero effect in the targeting advertising framework studied by the current chapter. Given the parameter values of Fig 6.3, we estimate lift_{ad} as a function of: the total user population, the user targeting probability p_{sel}, and a set of true lift_{ad} values. We generate the counts of Table 6.3 assuming the point estimate from Eq 6.13 is perfect. Given these count sets, we fit the model using the Bayesian approach of Section 6.4. Fig 6.3(a) shows that even when the user population is 40 million, the credible interval includes zero for all the randomized designs analyzed, $P(Z_i = S) = \{0.95, 0.92, 0.89, 0.86\}$. If we naively set 5% of the users as control group ($P(Z_i = S) = 0.95$), a typical industrial



(a) (b) (c)

Figure 6.3: Estimation power as a function of: (a) total user population in millions, (b) user targeting probability, (c) campaign lift on the targeted users (%). 90% credible intervals are displayed. y-axis represents estimated lift$_{ad}$. Parameters: Population: 37,158,296, θ_{1C}=1.48E-3, θ_{1S}=1.56E-3, lift$_{ad}$=5.40%, $P(Z_i = S)$=0.95, p_{sel}=0.3, θ_0=1E-3. Parameter source: [55].

practice, the experiment will be useless. When the user targeting probability is $p_{sel} = 0.4$, we observe that the zero effect is discarded of the 90% credible interval when 11% ($P(Z_i = S) = 0.89$) or higher user population is used as control group, which is depicted by Fig 6.3(b). Fig 6.3(c) displays that true lift$_{ad}$ values as low as 6% are detected when 14% ($P(Z_i = S) = 0.86$) of users are assigned to the control group. This analysis indicates the need for performing a similar analysis at the time of designing the experiment based on parameter predictions.

Part IV

Changing the User Targeting Paradigm: From Prediction based to Causal based Targeting

Chapter 7

Campaign Mid-flight Causal Optimization

7.1 Introduction and Problem Context

User targeting development has focused largely on optimizing user conversions by serving ads to the users who are most likely to convert [66]. Often the evaluation of these algorithms is based on the prediction power of conversions, which are liable to be not caused by the campaign [55]. This standard framework has lead to a limited targeting effect of performance-based campaigns, cost-per-action (CPA) campaigns, on causally generated (or incremental) conversions [7]. Moreover, as illustrated in Chapter 6 and by Berman (2013) in [14], the optimization of user conversions increases the probability of targeting *always-buy* users who do not contribute the campaign causal effect. Besides, this practice often leads to large discrepancies when one tests these algorithms with a randomized experiment.

The use of randomized experiments is becoming the standard practice to measure accurately the casual ad effect on user conversions [55]. Given that randomized experiments are expensive, the generated data should be leveraged as much as possible. However, the use of this dataset has been limited to the ad effectiveness estimation only without leveraging its power on user targeting.

7.2 Chapter Contribution

We propose a user simulator that leverages the data of randomized experiments by considering all the visiting users to the publisher websites [12, 7]. We analyze the effect of the user targeting policy measured by 1) the campaign/placebo exposure difference, and 2) the overall campaign effect. We take advantage of the user influenceable categories of Chapter 6 in Section 6.4.3 to optimize the overall campaign effect.

To evaluate the impact of the user targeting policy, we predict the user conversion response of the campaign and placebo ad exposures (targeted users) and the response of those who are not targeted. Based on the data of a randomized experiment for 37 million users, 8 million targeted users, and demographic user features, we simulate the standard conversion optimization policy and three targeting algorithms. We simulate the user targeting for both the focal and the placebo campaigns, and estimate the ad average causal effect. We estimate that the conversion optimization provides similar effectiveness to a uniform targeting and significantly inferior the causally optimized targeting.

We show the value of continuing evaluation by optimizing the user targeting assuming short-term ex-ante external validity of the overall campaign effects. Also, we take advantage of the user influencable categories of Chapter 6 in Section 6.4.3 in this optimization. Thus, we expand the effect estimation model of Chapter 6 in Section 6.4.1 to incorporate user features in the effect estimation. Based on this estimation, we test different user targeting policies for mid-flight overall campaign optimization, in the context of the control loop of Fig 5.1. Our results suggest that optimizing user targeting significantly impacts campaign effects.

7.3 Chapter Assumptions

In this section, we state the main assumptions used in the current chapter. The results and conclusions of this chapter are valid to the extent that these assumptions are applicable.

1. We refer to tracking cookies as users in the experimental design and estimation. We consider stable user cookies born before the campaign starts and that are active in the entire ad network. We assume the demographic user features to be finite and countable (e.g. *Male, 35-44 years old, 50K-75K income*). Data providers infer these features based on user online activities and associate them to tracking cookies.

2. The treatment assignment is assumed to be independent of the treatment effect, i.e. random assignment. This independence condition is a requirement for the methodology developed in this Chapter.

3. Stable Unit Treatment Value Assumption (SUTVA). We assume that the treatment status of any user does not affect the potential outcomes of the other users, i.e. no

interference between users is assumed.

4. We model the user conversions and ad exposures as a random events, with predetermined probabilities for each treatment arm. Thus, the converting users are conditionally independent of each other given a predetermined probability, and the ad-exposed users are conditionally independent of each other given a predetermined probabilities.

5. The user demographic missing features are independent of the campaign causal effect and balanced between randomized treatment arms.

6. The user demographic feature stratification is assumed to capture all the important heterogeneity between individuals. Thus, the individual causal effect based on this features is not biased [38, 39].

7. The user ad exposure is assumed to be binary (targeted or non-targeted) without considering the number ad exposures. Similarly, the visiting user indicator and the converting user indicator do not consider the multiple instances of these events for a given user.

8. Model parameters are assumed to be random variables. The user propensity to convert or to be targeted as a function of user demographics is assumed to be a probit function fitted by probit regression. The posterior distribution of the regression parameters is assumed to be Multivariate Normal distributed by Laplace approximation.

9. The above model components represent the structure we assumed in this Chapter. Eq 7.2 illustrates the joint distribution between the random variables of this structure.

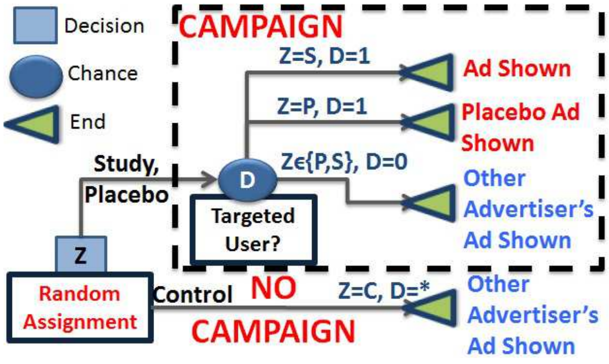


Figure 7.1: Randomized design used for user targeting simulations.

7.4 Methodology

7.4.1 User Targeting Simulation based on Ad Effectiveness Optimization

The standard practice to estimate the ad causal effect is to run a randomized experiment where the online visiting users are randomly assigned to the placebo or the study treatment arms. For those assigned to the study group, the campaign ad is displayed, while a placebo ad is displayed to the users of the placebo group. In practice, a placebo campaign, which replicates the targeting performed by the focal campaign, is run to show the placebo ads as discussed by Chapter 5 in Section 5.4. Fig. 7.1 depicts this process that considers the Study/Placebo treatment arms. To analyze the effect of the user targeting on the Ad effectiveness optimization, we simulate the user targeting for the focal and the placebo campaigns.

We define the following indicator variables for each user i: Z_i for {Control, Placebo, Study} assignments $\{C, P, S\}$; D_i for non-targeted/ targeted users $\{0, 1\}$; Y_i for non-converting/converting users $\{0, 1\}$; and X_i for feature segments defined to be finite and countable. For the current analysis, we consider the Placebo/Study treatment arms ($Z_i \in \{P, S\}$). We calculate the user counts by segments, $N^y_{dz}|X_i$ given $D_i = d$, $Z_i = z$, $Y_i = y$, X_i, leading to the set:

$$N^{obs} = \left\{ N^y_{dz}|X_i : \forall d \in \{0,1\}, \forall z \in \{P,S\}, \forall y \in \{0,1\}, \forall X_i \right\}. \tag{7.1}$$

Thus, the cardinality of N^{obs} becomes: $\#\{N^{obs}\} = 8 \times \#\{X_i \forall i\}$.

We use the data, N^{obs}, to simulate a given targeting function, $F_{targ}(X_i)$, based on Algorithm 4. We model the user response to the campaign and the placebo ads, $P(Y_i = 1|D_i = 1, Z_i = z, X_i) = \theta_{1z}|X_i : \forall z \in \{P, S\}$, as well as the response of the non-targeted population, $P(Y_i = 1|D_i = 0, Z_i = z, X_i) = \theta_{0z}|X_i : \forall z \in \{P, S\}$, through a probit transformation as illustrated by steps 3–4 of Algorithm 4. Here, glmfit($[N^1|X, N^0|X]$) represents standard probit regression fitting given the vectors of successes and failures $N^1|X_i, N^0|X_i$, and feature vector X_i.

We consider the audience-by-segment constraint $N^{Visit}_z|X_i$, and the observed targeted users as a fixed campaign budget N^{budget}_{1z} (steps 5–6). We define a budget multiplier λ to guarantee that all this budget is consumed by the user targeting, which includes the probability of user segments $P(X_i)$ (steps: 10–11). The *min* function enforces the visiting population segment constraints ($N^{Visit}_{remain}|X_i$). The while loop of steps 9–15 re-distributes the remaining budget in case $N^{Visit}_{remain}|X_i$ is exhausted for any segment. We aggregate the user counts over X_i to generate the four counts given $Z_i = z$: $N^{new}_{z,agg} = \left\{ N^{y,new}_{dz} : \forall d \in \{0,1\}, \forall y \in \{0,1\} \right\}$.

This simulation is run for both treatment arms $z \in \{P, S\}$ independently, and the ad effect is measured based on a t-test of ATE=E$(Y_i|D_i = 1, Z_i = S)$−E$(Y_i|D_i = 1, Z_i =$

Algorithm 4 User Targeting Campaign Simulation

```
Input: Targeting function $F_{targ}(X_i)$, User Counts
    $N_z^{obs} = \{N_{dz}^{y}|X_i : \forall d \in \{0,1\}, \forall y \in \{0,1\}, \forall X_i\}$.
Output: Aggregated User Counts After Targeting $N_{z,agg}^{new} = \{N_{dz}^{y,new} : \forall d \in \{0,1\}, \forall y \in \{0,1\}\}$
Set $[\hat{\gamma}_{0z}, \hat{\gamma}_{1z}]$=[glmfit$\left(\left[N_{0z}^{1}|X_i, N_{0z}^{0}|X_i\right], \forall X_i\right)$, glmfit$\left(\left[N_{1z}^{1}|X_i, N_{1z}^{0}|X_i\right], \forall X_i\right)$]
    // Probit Approximation
Set $[\hat{\theta}_{0z}, \hat{\theta}_{1z}]|X_i = [\Phi(X_i'\hat{\gamma}_{0z}), \Phi(X_i'\hat{\gamma}_{1z})], \quad \forall X_i$ // Observed Conversion Propensity
Set $N_z^{Visit}|X_i = N_{1z}^{1} + N_{1z}^{0} + N_{0z}^{1} + N_{0z}^{0}|X_i, \quad \forall X_i$ // Audience per Segment $X_i$
Set $N_{1z}^{budget} = \sum_{\forall X_i}(N_{1z}^{1} + N_{1z}^{0})|X_i$ // Observed Budget
Set $N_{1z}^{1,new}|X_i = N_{1z}^{0,new}|X_i = 0, \quad \forall X_i$ // Set Counts
Set $N_{remain}^{budget} = N_{1z}^{budget}$ // Initialize Remaining Budget
while $N_{remain}^{budget} > 0$ do
    Set $P(X_i) = N_{remain}^{Visit}|X_i / \sum_{\forall X_i} N_{remain}^{Visit}|X_i, \quad \forall X_i$
    Set $\lambda = N_{remain}^{budget} / \left(\sum_{\forall X_i} N_{remain}^{budget} \times F_{targ}(X_i) \times P(X_i)|X_i\right)$ // Budget Multiplier
    Set $\left[N_{1z}^{1,new}, N_{1z}^{0,new}\right]|X_i = \left[N_{1z}^{1,new}, N_{1z}^{0,new}\right]|X_i$
        $+\min\left(\lambda \times F_{targ}(X_i) \times N_{remain}^{budget} \times P(X_i), N_{remain}^{Visit}|X_i\right) \times \left[\hat{\theta}_{1z}, 1-\hat{\theta}_{1z}\right]|X_i, \quad \forall X_i$
        // User Targeting
    Set $N_{remain}^{Visit}|X_i = N_z^{Visit} - (N_{1z}^{1,new} + N_{1z}^{0,new})|X_i, \quad \forall X_i$ // Remaining Audience
    Set $N_{remain}^{budget} = N_{1z}^{budget} - \left(\sum_{\forall X_i}[N_{1z}^{1,new} + N_{1z}^{0,new}|X_i]\right)$ // Remaining Budget
end while
Set $\left[N_{0z}^{1,new}, N_{0z}^{0,new}\right]|X_i = N_{remain}^{Visit} \times \left[\hat{\theta}_{0z}, 1-\hat{\theta}_{0z}\right]|X_i, \quad \forall X_i$ // Non-Targeted User Counts
Set $N_{z,agg}^{new} = \left\{\sum_{\forall X_i} N_{dz}^{y,new}|X_i : \forall d \in \{0,1\}, \forall y \in \{0,1\}\right\}$ // Aggregate User Counts
```

$P) = \theta_{1P} - \theta_{1S}$, using $\{N_{1z}^{y,new} : \forall z \in \{P,S\}, \forall y \in \{0,1\}\}$. We use the average treatment effect for each user feature segment, ATE(X_i) as the optimal ad effectiveness targeting rules.

7.4.2 User Targeting based on Campaign Effectiveness Optimization

To optimize the overall campaign mid-flight, we estimate the campaign effect given the user feature segments (X_i). This estimation provides the user targeting rules to construct the user targeting function $F_{targ}(X_i)$. We address this development by replacing the Bernoulli distributions of the estimation model of Chapter 6 in Section 6.4.1 (Eq 6.2) with probit regressions conditional on X_i. In this manner, the campaign effects conditional on $X_i : \forall X_i$ are estimated and used to guide the target engine. For the current Section analysis, we consider the Control/Study treatment arms of the experimental design of Fig 7.1 ($Z_i \in \{C, S\}$). Let $\Phi(x)$ be the standard Normal cumulative density function, and $\Theta_X = \{\gamma_0, \gamma_{1C}, \gamma_{1S}, \beta_{sel}\}$, then:

$$P(Y, Z, D, \Theta_X|X) = P(\Theta_X) \prod_{\forall i} P(D_i = d|\beta_{sel}, X_i) P(Y_i(Z_i)|D_i = d, Z_i = z, \gamma_{dz}, X_i) P(Z_i),$$
$$P(D_i|\beta_{sel}, X_i) = \Phi(\eta_i^\beta), \quad \eta_i^\beta = X_i' \beta_{sel}, \quad P(Y_i|D_i, Z_i, \gamma_{dz}, X_i) = \Phi(\eta_i^{dz}), \qquad \eta_i^{dz} = X_i' \gamma_{dz}. \tag{7.2}$$

This model exploits the power of user randomization, and balances the treatment groups in the inference of $D_i^C|X_i$ based on the propensity of being targeted.

To estimate the model of Eq 7.2, we provide a Gibbs-sampling based approach depicted by Algorithm 5. We calculate the user counts of Control and Study treatment arms (Table 6.3) for all user feature combination segments, assumed to be finite and countable:

$$N_{Camp}^{obs} = \left\{ N_{dS}^y|X_i, N_{\{0,1\}S}^y|X_i : \forall d \in \{0,1\}, \forall y \in \{0,1\}, \forall X_i \right\}, \tag{7.3}$$

whose cardinality becomes $\#\{N_{Camp}^{obs}\} = 6 \times \#\{\forall X_i\}$.

We sample the missing targeting indicator, $D_i^C|X_i : \forall X_i$, following a similar logic to that of Algorithm 3 (steps: 5-9). We fit binomial probit regression functions

Algorithm 5 Gibbs Sampling Algorithm based on the joint distribution of Eq. 7.2

```
1: Input: $N^{obs}_{Camp} = \Big\{N^y_{dS}|X_i, N^y_{\{0,1\}S}|X_i : \forall d \in \{0,1\}, \forall y \in \{0,1\}, \forall X_i\Big\}$
2: Define $N_{samp}|X_i = \{N^y_{dC}|X_i : \forall d \in \{0,1\}, \forall y \in \{0,1\}\}, \forall X_i$
3: Initial guess $\Theta^0_X = \{\gamma_0, \gamma_{1z}, \beta_{sel}\}^0, \quad \forall z \in \{C, S\}$
4: for $i \leftarrow 1$ to $N_{burnin} + N_s$ do
5:     Set $P(D_i = d|\beta_{sel}, X_i) = (\Phi(\eta^\beta_i))^d(1 - \Phi(\eta^\beta_i))^{1-d}, \quad \eta^\beta_i = X_i'\beta_{sel}, \quad \forall X_i$
6:     Set $P(Y_i(Z_i) = y|D^z_i = d, Z_i = z, \gamma_{dz}, X_i) = (\Phi(\eta^{\gamma_{dz}}_i))^y(1 - \Phi(\eta^{\gamma_{dz}}_i))^{1-y}$,
       $\eta^{dz}_i = X_i'\gamma_{dz}, \quad \forall X_i$
7:     Set $P(D^{Cy}_i = 1|\Theta_X, D^s, Y, Z, X_i)$
       $= \dfrac{P(D_i = 1|\beta_{sel}, X_i)P(Y_i(C) = y|D_i = 1, Z_i = C, \gamma_{dz}, X_i)}{\sum_{\forall d \in \{0,1\}} P(D_i = d|\beta_{sel}, X_i)P(Y_i(C) = y|D_i = d, Z_i = C, \gamma_{dz}, X_i)}, \quad \forall X_i$
8:     Draw $N^y_{1C}|\Theta_X, N_{obs}, X_i \sim \text{Binomial}\Big(N^y_{\{0,1\}C}|X_i, P(D^{Cy}_i = 1|\Theta, N_{obs}, X_i)\Big)$,
       $\forall y \in \{0,1\}, \forall X_i$
9:     Set $N^y_{0C}|X_i = N^y_{\{0,1\}C}|X_i - N^y_{1C}|X_i, \quad \forall y \in \{0,1\}, \forall X_i$
10:    Set $\{\hat\gamma_{1z}, \hat\Sigma_{1z}\}$=glmfit$\big(\big[N^1_{1z}|X_i, N^0_{1z}|X_i\big], \forall X_i\big), \quad \forall z \in \{C, S\}$
11:    Set $\{\hat\gamma_0, \hat\Sigma_0\}$=glmfit$\big(\big[N^1_{0C} + N^1_{0S}|X_i, N^0_{1C} + N^0_{0S}|X_i\big], \forall X_i\big)$
12:    Set $\{\hat\beta_{sel}, \hat\Sigma_{sel}\}$=glmfit$\Big(\Big[\sum_{\forall z\in\{C,S\},\forall y\in\{0,1\}} N^y_{1z}|X_i, \sum_{\forall z\in\{C,S\},\forall y\in\{0,1\}} N^y_{0z}|Xi\Big], \forall X_i\Big)$
13:    Draw $\gamma^{(i)}_{1z}|\Theta_{X,-\gamma_{1z}}, N_{samp}, N_{obs}, X \sim \text{MVN}\Big(\hat\gamma_{1z}, \hat\Sigma_{1z}\Big), \quad \forall z \in \{C, S\}$
14:    Draw $\gamma^{(i)}_0|\Theta_{X,-\gamma_0}, N_{samp}, N_{obs}, X \sim \text{MVN}\Big(\hat\gamma_0, \hat\Sigma_0\Big)$
15:    Draw $\beta^{(i)}_{sel}|\Theta_{X,-\beta_{sel}}, N_{samp}, N_{obs}, X \sim \text{MVN}\Big(\hat\beta_{sel}, \hat\Sigma_{sel}\Big)$
16: end for
17: return $\Theta^{N_{burnin}+1:N_s}_X$
```

based on these user counts. We use a standard fitting function to calculate the Maximum Likelihood estimate of the regression coefficients and its covariance matrix, (steps 10-12: $\{\hat\gamma, \hat\Sigma\}$=glmfit($[N^1|X_i, N^0|X_i]$, $\forall X_i$)). This fitting strategy avoids the fitting of probit re-

Algorithm 6 User Targeting Simulator for the Campaign Effectiveness Optimization

```
Input: Targeting function $F_{targ}(X_i)$, Non-zero Effect Indicator function $F_{sig}(X_i)$, LATE$_{Ad}$ Sign function $F_{sign}^{LATE}(X_i)$, Sign Certainty weights $\mathbf{w}^{sig} = \{w^-, w^\pm, w^+\}$, User Counts $N_{Camp}^{obs}$ as defined by Eq 7.3.
//Set segment weighting function $D_w^{target}(X_i)$, based on inputs: $F_{targ}(X_i)$, $F_{sig}(X_i)$, $F_{sign}^{LATE}(X_i)$, $\boldsymbol{w}^{sig}$
Define $D_w^{target}(X_i) = \begin{cases} w^\pm \times F_{targ}(X_i) & \text{if } F_{sig}(X_i) = \textbf{false} \\ w^+ \times F_{targ}(X_i) & \text{if } F_{sig}(X_i) = \textbf{true and } F_{sign}^{LATE}(X_i) = + \\ w^- \times F_{targ}(X_i) & \text{if } F_{sig}(X_i) = \textbf{true and } F_{sign}^{LATE}(X_i) = - \end{cases}$
Set $N_{S,agg}^{new}$ to the output of Algorithm 4 with inputs:
    $F_{Targ}(X_i) = D_w^{target}(X_i), \quad N_z^{obs} = N_S^{obs} | X_i, \forall X_i$
    // Simulate Campaign Targeting, $Z_i = S$
Set $N_{Camp,agg}^{new} = \left\{ \sum_{\forall X_i} N_C^{obs} | X_i, N_{S,agg}^{new} \right\}$ // Aggregate User Counts
return $N_{Camp,agg}^{new}$
```

gressions with millions of data points. Based on these estimates, the regression parameters are sampled from multivariate normal distributions (steps 13-15: MVN$(\hat{\gamma}, \hat{\Sigma})$) by Laplace approximation method [40]. $\Theta_X^{(1:N_s)}$ samples are employed to generate credible intervals for the effect estimates conditional on user features X_i[1].

To simulate a given targeting function, we execute the Algorithm 6, which aggregates the user counts of the study group ($Z_i = S$) given: 1) a targeting function $F_{targ}(X_i)$; 2) a non-zero effect indicator function $F_{sig}(X_i)$, and 3) LATE$_{Ad}$ sign function $F_{sign}^{LATE}(X_i)$.

[1] We note that an Expectation-Maximization based point estimate can be fitted by iteratively estimating the expected missing targeting indicators (step 7 expression), and fitting the regression parameters given this expectation (steps 10-12). As a point estimator, this method would not provide parameter credible intervals.

Given the posterior distribution samples ($\Theta_X^{N_{burnin}+1:N_{burnin}+N_s}$), the user conversion probabilities are estimated based on the logit transformation of Eq 7.2 for all feature segments. We employ Eq 6.8 to estimate the median probability of the influenceable user categories of the targeted population given the user features ($P(U_i|D_i = 1, X_i)$) to determine $F_{targ}(X_i)$. We define $F_{sig}(X_i)$ to be the inclusion/non-inclusion of the zero $\text{LATE}_{Ad}|X_i$ effect in the 90% credible intervals, and $F_{sign}^{LATE}(X_i)$ to be the sign of $\text{LATE}_{Ad}|X_i$. In addition, a sign certainty weighting set, $\mathbf{w}^{sig} = \{w^-, w^\pm, w^+\}$, is fixed. These functions are combined into a segment weighting $D_w^{target}(X_i)$ (steps: 1-3). We simulate the user targeting for the users of the study group, $N_S^{obs}|X_i, \forall X_i$, by executing Algorithm 4. We use $D_w^{target}(X_i)$ as targeting function for this simulation (step: 4). We aggregate the user counts of the Control group over X_i, $N_C^{obs}|X_i$, and concatenate them to the aggregated Study user counts after targeting, $N_{S,agg}^{new}$ (step: 5). This process generates the six counts of Table 6.3 that are analyzed by the Algorithm 3 of Section 6.4.1 to estimate the local average treatment effect on the users exposed to the ad.

7.5 Results

7.5.1 Ad Effectiveness Results

We consider the user features: age, gender and income; segmented by value ranges (finite and countable). We ran the focal and placebo campaigns that generate the logged data as CPM campaigns in an exploratory phase. The campaign running time is two weeks. For $Z_i = S$, the total and targeted population sizes are 18.74 and 4.01 million. For $Z_i = P$, the total and targeted population sizes are 18.70 and 4.09 million. We consider the missing

Table 7.1: Simulator Validation. Targeting functions are trained and tested with the same data. ATE intervals, estimated based on a t-test, are shown for 0.10 significance level.

$F_{targ}(X_i)$	**ATE** (1e-6)	*lift* (%)	$F_{targ}(X_i)$	**ATE** (1e-6)	*lift* (%)
1(Random)	3.76±9.83	**7.37**	$\theta_{11}\|X_i$	2.92±10.0	5.46
ATE(X_i)	5.63±9.62	11.77	-ATE(X_i)	-1.74±10.3	-2.94
ATE$^+$(X_i)	8.74±9.53	**19.26**	-ATE$^-$(X_i)	-6.68±10.9	**-9.78**

values as a feature value (81.4% of the users have one or more feature values missing). We use the first half of the campaign as training and the second half for testing. We fit the conversion probabilities (θ_{1P}, θ_{1S}) in the training set with probit regressions as done by steps 3–4 of Algorithm 4.

We test the following targeting policies with training data: uniform, $F(X_i) = 1$; conversion optimization, $\theta_{11}|X_i$; and maximization/minimization of ATE, {ATE(X_i), $-$ATE(X_i)}. We also test a variant of the ATE maximization, where we set the segments with negative ATE to the minimum positive ATE (ATE$^+$(X_i)). Likewise, we test the minimization of ATE ($-$ATE$^-$(X_i)). Table 7.1 shows the results. As expected, maximizing ATE shows the best performance, and minimizing ATE the worst (lift= 19.29% for ATE$^+$(X_i), and lift= -9.78% for $-$ATE$^-$(X_i)). Both estimations are far from the uniform targeting (lift= 7.37%) validating the simulator.

Table 7.2 shows the testing results. We estimate that the performance of the user conversion optimization ($\theta_{11}/(1-\theta_{11})^2$) is similar to that of a uniform targeting (10.91% versus 11.01%). The best performance is provided by optimizing the lift and setting the

[2] We note that more sophisticated hierarchical classifiers are available for this prediction problem [82]

Table 7.2: Targeting Policy Testing Results. ATE intervals, estimated based on a t-test, are shown for 0.10 significance level.

	All Users		No Missing Features	
$F_{targ}(X_i)$	**ATE**(1e-5)	*lift*(%)	**ATE**(1e-5)	*lift*(%)
1 (Uniform)	1.35±1.74	11.01	2.21±4.26	14.06
$\theta_{11}/(1-\theta_{11})\|X_i$	1.38±1.77	10.91	1.98±3.85	12.25
ATE(X_i)	1.45±1.73	12.00	2.45±4.39	16.25
ATE$^+$(X_i)	1.69±1.76	**13.72**	2.92±3.55	**19.92**
lift$^+$(X_i)	**1.78±1.76**	**14.47**	3.00±3.42	**20.87**

negative segments to the minimum positive lift (lift$^+(X_i)$ with 14.47%), which is the only significant effect at 0.10 statistical level (1.78±1.76e-5). We show the effect results estimated for users with no missing features, which depict the same directional results with larger intervals.

7.5.2 Overall Campaign Effectiveness Results

We leverage demographic user features to optimize the user targeting mid-flight, *i.e.* in the middle of the campaign. For the visiting users of the CPM campaign discussed in Section 7.5.1, we know the gender, age, and income. These features are segmented by ranges, to make them finite and countable. For the current analysis, we consider the control/study treatment arms ($Z_i \in \{C, S\}$). We partition the campaign data in duration by half and train the model of Eq 7.2 based on Algorithm 5 for the first half. We perform standard dummy variable feature coding required to run the probit regression based distributions for categorical features. We test different user targeting policies, based on the user response

Table 7.3: Averaged campaign effect results for different targeting functions based on Algorithm 6 using the first half of campaign 3 as training and testing in the second half. Targeting policies: (a) Per^{+} vs Per^{-}, (b) Per^{+} vs $\{$Per$^{-}\cup$ AB$\}$, (c) Per^{+} vs $\neg$Per^{+}, (d) Y=1 vs Y=0. Second half campaign duration: 7 days.

	$\mathbf{w}^{sig} = \{1, 1, 1\}$		$\mathbf{w}^{sig} = \{0.6, 1, 1.1\}$	
Targeting Function	**LATE**$_{Ad}$	**lift**$_{Ad}$	**LATE**$_{Ad}$	**lift**$_{Ad}$
$F_{targ}(X_i)$	(1e-5)	(%)	(1e-5)	(%)
(a) $\frac{P(U_i = \text{Per}^{+} \mid D_i = 1, X_i)}{P(U_i = \text{Per}^{-} \mid D_i = 1, X_i)}$	1.27	9.86	1.18	8.53
(b) $\frac{P(U_i = \text{Per}^{+} \mid D_i = 1, X_i)}{P(U_i = \text{Per}^{-} \cup \text{AB} \mid D_i = 1, X_i)}$	1.51	11.93	1.47	10.80
(c) $\frac{P(U_i = \text{Per}^{+} \mid D_i = 1, X_i)}{1 - P(U_i = \text{Per}^{+} \mid D_i = 1, X_i)}$	1.78	**14.28**	1.80	13.40
(d) $\frac{P(Y_i = 1 \mid D_i = 1, Z_i = S, X_i)}{P(Y_i = 0 \mid D_i = 1, Z_i = S, X_i)}$	1.60	**11.72**	-	-
	$\mathbf{w}^{sig} = \{0.8, 1, 1.2\}$		$\mathbf{w}^{sig} = \{0.8, 1, 1.1\}$	
(a) $\frac{P(U_i = \text{Per}^{+} \mid D_i = 1, X_i)}{P(U_i = \text{Per}^{-} \mid D_i = 1, X_i)}$	1.51	11.49	1.85	14.42
(b) $\frac{P(U_i = \text{Per}^{+} \mid D_i = 1, X_i)}{P(U_i = \text{Per}^{-} \cup \text{AB} \mid D_i = 1, X_i)}$	1.63	12.52	1.65	12.63
(c) $\frac{P(U_i = \text{Per}^{+} \mid D_i = 1, X_i)}{1 - P(U_i = \text{Per}^{+} \mid D_i = 1, X_i)}$	1.92	14.74	1.93	**14.89**
(d) $\frac{P(Y_i = 1 \mid D_i = 1, Z_i = S, X_i)}{P(Y_i = 0 \mid D_i = 1, Z_i = S, X_i)}$	-	-	-	-

categories of Eq 6.8 in Section 6.4.3 for each user segment X_i, on the second half of the campaign.

Table 7.3 shows the results where four targeting algorithms are tested. We use the optimization of conversion probability ((d) Y=1 vs Y=0, $\mathbf{w}^{sig} = \{1, 1, 1\}$) as baseline because this targeting policy is the standard practice given observational data. We estimate that this practice is reasonably effective when compared with other targeting algorithms based on causal effects ((d)11.72% versus (b)11.93% or (a)9.86% average lift_{Ad})[3]. However, optimizing (c) Per^+ vs $\neg\text{Per}^+$ shows the highest performance given $\mathbf{w}^{sig} = \{1, 1, 1\}$ (14.28% lift_{Ad}). We note that the user targeting of the observational data is exploratory (CPM campaign); consequently the selection effect here is significantly smaller than those of CPA campaigns typically used. As a result, the performance of the standard practice (d) is likely to be inferior in reality.

We test three weighting frameworks based on the inclusion/non-inclusion of the zero campaign effect in the 90% credible intervals of the user demographic segments in training. Intuition suggests that avoiding segments with negative-only intervals and boosting segments with positive-only intervals greatly would increase the performance dramatically. However, these parameters require tuning, and a modest decrease of the segments with negative-only intervals and an increase of the segments with positive-only intervals show to be more effective.

We find that $\mathbf{w}^{sig} = \{0.8, 1, 1.1\}$ show the highest performance of the weighting frameworks we test ((c)14.89% average lift_{Ad}). This analysis shows the value of the experimental design and the estimation to optimize the user targeting in as illustrated by Fig 5.1. We caution the reader of the limitations of this study including the quality of the user features (cookie-based features), and the percentage of users with missing features estimated to be at 75%.

[3] Credible intervals are in the range of ±20% for all targeting functions evaluated. The short evaluation time, seven days, and the observed budget, which is kept constant in the simulation, are among the reasons.

7.6 Impact and Limitations

We have proposed a user targeting simulator that uses data from standard ad effectiveness causal estimation. We have found evidence that the standard practice of optimizing the conversion probability does not optimize the causal effect of the ad. We have shown that the user targeting makes a difference in the ad evaluation even when the randomized design displays placebo ads. This result contradicts the standard evaluation practice of measuring the effect with a non-optimized campaign, which is assumed to hold for future optimized exposures. We have shown the value of continuing evaluation and also the leverage of user features to improve the user targeting. In a measurement-optimization cycle (mid-flight optimization), the use of randomized experiments potentially enables the transfer of learning from attribution to user targeting and ex-ante optimization.

In the current analysis, we find the demographic user features that correlate with incremental conversions, as opposed to observed conversions. These targeting rules are often campaign-specific. However, behavioral user responses that correlate with incremental conversions are more generally valid across campaigns. This external validity of behavioral features has been demonstrated in the targeting of converting users before [66]. Thus, identifying those features provides an opportunity to optimize incremental conversions when continuing evaluation is not affordable. Characterizing user search activity in document retrieval [19],or interactively [33] potentially correlates with causally generated conversions. Similarly, user concept extraction characterized by product categories represent other highly informative behavioral signal [81]. In this context, we study the value of user clicks as a correlated signal with incremental conversions in Chapter 8.

Chapter 8

Are User Clicks Valuable to Causal Targeting?: A Potential Outcomes Approach

8.1 Introduction and Problem Context

Recent developments in online campaign attribution and evaluation have demonstrated the effectiveness of display advertising on user conversion and search keywords probabilities [57, 55, 12]. These findings have motivated advertisers and ad networks to measure the effectiveness of campaigns in metrics other than user clicks. The belief that user clicks are not informative to measure the success of a campaign is increasingly gaining acceptance in the research community and industry. Dalessandro *et al.* (2012) concluded that user clicks do not correlate with user conversions and that user targeting based on clicks is statistically indistinguishable from random guessing [28]. These conclusions are drawn based on the power of user clicks to predict conversions in observational data. However, a large percentage of these conversions are likely to be unrelated to, and not caused by, the campaign, as it is standard in online advertising attribution analysis [55].

A more accurate approach is to measure the campaign effect on the conversion probability of the users who click on the ad (clickers) with a randomized experiment. Based on this effect, we can determine the importance of the click in the user targeting optimization. However, to design such experiment one would need to randomize the users into control/study groups after finding the clickers. This randomized design is not feasible because the online ad must be displayed to the users of the study and the control groups to observe the user selection introduced by the click event.

8.2 Chapter Contribution

We propose to estimate the local average campaign effect on the clicker conversions based on the standard campaign evaluation randomized design. We compare the effect on the conversion probability of the clickers and the non-clickers to determine if the click event provides relevant information to separate users with higher or lower campaign impact. To the best of our knowledge, the proposed method is the first approach in the ad effectiveness measurement literature that estimates this effect based on randomized experiments.

We approach the problem in two phases: the randomized design, and the causal modeling given this design. For the randomized design, we discuss the issues that prevent us from designing an experiment focused solely on the clicking users. We illustrate the randomized design we employed, which is focused on the measurement of the ad exposure effectiveness.

In the casual modeling phase, we propose a method in the Potential Outcomes causal model to estimate the campaign effect on the clicker conversions based on the randomized experiment to measure the ad exposure effect. We use Principal Stratification [36] to condition the campaign effect on the user click event. This framework allows us to model the treatment causal effect conditional on post-treatment variables, which are affected by the treatment and consequently are non-ignorable [75]. The proposed approach closes the gap between a purely observational analysis of the effect on the clicker conversions and the analysis of this effect with focused randomized experiments. As the problem is different from previously addressed problems by Principal Stratification, we solve the identifiability problem, typical of these problems, with mild and reasonable assumptions in online advertising. The uncertainty of the estimations is modeled in a Bayesian framework and a Gibbs-sampling based inference approach.

We analyze the effectiveness of the estimation method with simulated data, and discuss the results for two large-scale randomized experiments in detailed. Finally, we discuss the impact and benefits of the estimation approach in the ad effectiveness literature, as well as the impact of the results for user targeting and attribution.

8.3 Chapter Assumptions

In this section, we state the main assumptions used in the current chapter. The results and conclusions of this chapter are valid to the extent that these assumptions are applicable.

1. We refer to tracking cookies as users in the experimental design and estimation. We consider stable user cookies born before the campaign starts and that are active in entire ad network.

2. The treatment assignment is assumed to be independent of the treatment effect, i.e. random assignment. This independence condition is a requirement for the methodology developed in this Chapter.

3. Stable Unit Treatment Value Assumption (SUTVA). We assume that the treatment status of any user does not affect the potential outcomes of the other users, i.e. no interference between users is assumed.

4. We model the user conversions and use clicks as a random events, with predetermined probabilities for each treatment arm. Thus, the converting users are conditionally independent of each other given a predetermined probability, and the ad-exposed users are conditionally independent of each other given a predetermined probabilities.

5. The user click is assumed to be binary (clicker or non-clicker) without considering the number clicks or ad exposures. Similarly, the converting user indicator does not consider the multiple instances of this event for a given user.

6. We consider the ad exposure effect and the local effect on the clickers to be positive. Thus, the ad exposure must have a positive effect on the user conversion probability of the targeted population for the analysis of this Chapter to be valid.

7. Model parameters are assumed to be random, with standard Jeffrey's conjugate prior distribution. Indicator random variables are assumed to be Bernoulli distributed with prior distribution: Beta(0.5,0.5). Positive-effect constraints are imposed on the prior distribution of the probability of conversion.

8. The above model components represent the structure we assumed in this Chapter. Eq 8.2 illustrates the joint distribution between the random variables of this structure.

8.4 Randomized Design

The current practice to estimate the campaign causal effect is to run a randomized experiment assuming the ad creative is the *treatment* to evaluate. In this context, the online visiting users are randomly assigned to the control or the study groups before the campaign starts. These users are maintained in the assigned group during the entire duration of the campaign. For those assigned to the study group, the campaign ad is displayed. While a placebo ad (assumed to be entirely unrelated to the advertiser running the campaign) is shown to the users of the control group [55, 87]. Then, the online users are tracked, based on tracking cookies or e-mail sign-ups, to observe if they convert in the advertiser website or not. In practice, the placebo ads are displayed by running a placebo campaign, which replicates the user selection (or targeting) performed by the advertising campaign.

Following a similar logic, to design a randomized experiment to estimate the campaign effect on the clicker conversions one can run a placebo campaign to replicate the clicking user selection. Then, a placebo ad would be displayed to the users in the control group once this selection is observed to the placebo campaign. Unfortunately, running such

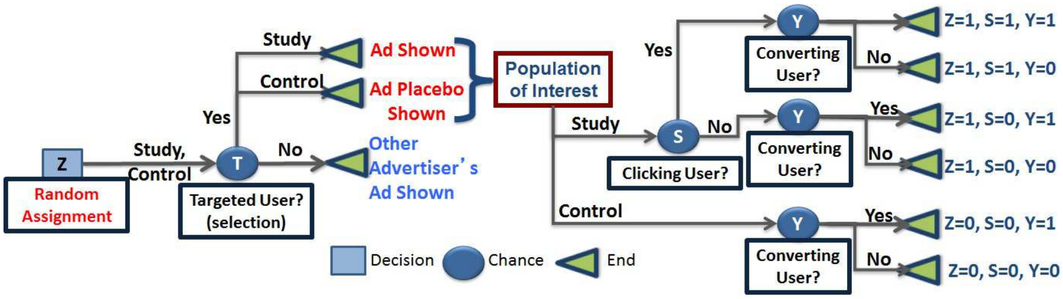


Figure 8.1: Randomized Design. The user clicks and conversions are collected for the user population of interest.

a design is not feasible because the clicking user population segment cannot be observed without showing the campaign ad. This prevents us from running a randomized experiment focused on the sub-population of clicking users.

To avoid relying on fully observational data, whose effectiveness to estimate the causal attribution has been severely questioned in Online Advertising [55], we take advantage of the randomized design used in standard campaign evaluation. Thus, we randomly assign the users to control and study groups and focus on those selected by the ad-network targeting engine. This user population represents the universe of users for the effects of the current Chapter. Fig 8.1 illustrates the randomized design. As a consequence of this design, the user selection introduced by the user clicks becomes a post-treatment variable, or a random variable that is affected by the campaign [36]. In the Potential Outcomes causal model, this variable is non-ignorable. Also this variable must be modeled to estimate causal campaign effect on these user sub-populations [75][1].

[1] In Econometric causality, the user clicking indicator would be endogenous because this variable is not controllable by the experimenter [48].

8.5 Campaign Causal Estimation

8.5.1 Potential Outcomes and Principal Stratification

Potential Outcomes Causal Model, also known as Rubin Causal Model (RCM)[75], is based on the analysis of units, treatments, and potential outcomes. Fundamentally, RCM analyzes the unit potential outcomes to each of the treatments. For two treatment arms, control and study, this framework implies that half of the data is missing because we can never observe the response of a unit in both arms. Thus, the causal inference problem is addressed as a missing value inference problem. This problem is commonly approached with a Bayesian parametric model to estimate the mean posterior distribution. RCM incorporates the treatment assignment mechanism to offer a clear distinction between randomized experiments and observational studies. If the treatment assignment is independent of the treatment effect (i.e. random assignment), then the causal estimates are unbiased. Standard notation in RCM is to consider the variable of a user i for a given treatment arm Z_i as $Y_i(Z_i)$. In spite of the ability to model the treatment effect on post-treatment variables S, typically the primary interest is to estimate the treatment effect on Y conditional on S. However, this is not straightforward because $S_i(0) \neq S_i(1)$, and consequently S_i is not ignorable. Therefore, conditioning the effect estimates of the observed values of S introduces a post-treatment bias.

Principal Stratification modeling provides a framework to determine unbiased treatment effect conditional on post-treatment variables [36]. The key element of this method is the identification of user classes, or strata, S_i^P with equal treatment effects and probability of treatment assignment. Thus, the probability of S_i^P must be independent

Table 8.1: User counts based on the user potential outcomes. N^y_{cz}, where $C_i = c$, $Z_i = z$, $Y_i = y$, are user counts for the given values of Y, Z, C. Missing values are presented as *.

User	**Potential Outcomes**				**Treatment**	**Principal**	
Counts	Control		Study		**Assignment**	**Stratum**	
N^y_{cz}	$S_i(0)$	$Y_i(0)$	$S_i(1)$	$Y_i(1)$	Z_i	$(S_i(0), S_i(1))$	C_i
$N^0_{\{0,1\}0}$	0	0	*	*	0	(0,*)	*
$N^1_{\{0,1\}0}$	0	1	*	*	0	(0,*)	*
N^0_{01}	0	*	0	0	1	(0,0)	0
N^1_{01}	0	*	0	1	1	(0,0)	0
N^0_{11}	0	*	1	0	1	(0,1)	1
N^1_{11}	0	*	1	1	1	(0,1)	1

(or ignorable) of the treatment assignment Z_i, to enforce that no treatment effect on the strata is allowed in the inference process.

8.5.2 Campaign Causal Effect on the Clicker Conversions

The randomized design of Fig 8.1 allows us to record the user clicks for both treatment groups. However, a user click on the campaign ad is not comparable to a click on a placebo ad. As a result, the user selection made by the clicker indicator in the study group is missing in the control group.

We define the following indicator random variables for each user i: Z_i for control/study group user assignments $\{0, 1\}$, S_i for non-clicker/clicker users $\{0, 1\}$, Y_i for non-converting/converting users $\{0, 1\}$. Given that the user must be in the study group to click the ad, the users of the control group never click the ad and consequently $\mathrm{S}_i = 0$ for these users. Therefore, we define the principal strata S_i^P, or user classes C$_i$ as follows:

$$S_i^P = \left\{ \begin{pmatrix} S_i(0) \\ S_i(1) \end{pmatrix} \right\} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, \qquad C_i = \begin{cases} 0 & \text{if } S_i^P = (0,0)' \\ 1 & \text{if } S_i^P = (0,1)' \end{cases} . \tag{8.1}$$

Table 8.1 illustrates the observed and missed data in the RCM notation. $C_i = 1$ are the users who click on the ad when they are assigned to the study group (clicker-if-assigned). $C_i = 0$ are the users who do not click on the ad regardless of the treatment group they are assigned to (never-clickers). Based on these definitions, we let C_i to be Bernoulli distributed with parameter π. Y_i is Bernoulli distributed with parameters θ_{cz} for the 4 combinations $C_i = c$, $Z_i = z$. Assuming a Bayesian approach to the parameter estimation, we define $\Theta = \{\theta_{cz}, \pi\}$ as random variables.

Similar to the case of randomized experiments with noncompliance [50, 4], this model is not identifiable if no further constraints are imposed. To estimate the campaign effect on the clicker conversions, we observe the stratum indicator C_i and the conversion indicator Y_i for the users in the study group $Z_i = 1$. These observed indicators allow us to estimate the user conversion probability for both strata in the study arm without constraints. By randomization, we know that the probability of observing this user selection π is the same in both treatment groups, which follows from the definition of the principal strata [36]. However, the user conversion probability for both principal strata users in the control group $\{\theta_{10}, \theta_{00}\}$ are not identifiable. Thus, to guarantee the model is identifiable we assume positive campaign effect. This assumption translates into $\theta_{c1} \geq \theta_{c0}$ for $c = \{0, 1\}$. Therefore, letting the indicator function be $I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ otherwise.

Assuming a prior distribution $P(\Theta)$, we have the joint distribution:

$$\begin{aligned} P(Y,Z,D,\Theta) = {} & P(\Theta) I_{[0,\theta_{01})}(\theta_{00}) I_{[0,\theta_{11})}(\theta_{10}) \\ & \times \prod_{\forall i} P(C_i = c|\pi) P(Y_i|C_i = c, Z_i = z, \theta_{cz}) P(Z_i = z). \end{aligned} \tag{8.2}$$

We assume standard conjugate Beta prior distributions for the Bernoulli distributed random variables $\Theta = \{\theta_{cz}, \pi\}$ for $c = \{0,1\}$ and $z = \{0,1\}$. For numerical stability, we use the Jeffreys prior distribution, Beta(0.5.0.5), which assumes a prior sample size of 1. We experiment with different prior probability but with the same sample size of 1. Given the number of users employed to estimate the conversion probabilities θ_{cz}, the effect of these prior probabilities becomes negligible.

We note that Balke and Pearl (1997) have reported in the context of imperfect compliance that estimating these effects is not feasible with no constraints [4]. They provide a set of bounds based on a method of moments assuming a large sample of individuals. The model of Eq 8.2 is fitted without relying on those bounds. As a consequence of the positive effect constraint and the full observance of the potential outcomes for the users of the study group detailed above.

8.5.3 Model Estimation

The inference objective of the joint distribution of Eq 8.2 is to estimate the posterior distribution of the parameters Θ given the observed data from Table 8.1. We denote the set of observed counts as D_{obs}. We solve this inference problem using Gibbs sampling by sampling from the conditional posterior distributions. Given an initial guess for Θ^0 and similar to standard mixture methods, we sample the missing user clicking indicator for the users in the control group and estimate the counts N_{c0}^{y} for $c = \{0,1\}, y = \{0,1\}$. We perform

Algorithm 7 Gibbs Sampling Algorithm based on the joint distribution of Eq. 8.2

```
**Input:** $D_{obs} = \left\{ N_{c1}^{y}, N_{\{0,1\}0}^{y} \right\}$ for $c = \{0,1\}, y = \{0,1\}$ from Table 8.1
Define $D_{samp} = \{N_{c0}^{y}\}$ for $c = \{0,1\}, y = \{0,1\}$
Set $\alpha_0 = 0.5$
Initial guess $\Theta^0 = \{\theta_{cz}, \pi\}^0, \quad \text{for } c = \{0,1\}, z = \{0,1\}$
**for** $s \leftarrow 1$ to $N_{burnin} + N_{samples}$ **do**
    Set $P(C_{i0}^{y} = 1|\Theta, D_{obs}) = \frac{\pi(\theta_{10})^{y}(1-\theta_{10})^{(1-y)}}{\pi(\theta_{10})^{y}(1-\theta_{10})^{(1-y)} + (1-\pi)(\theta_{00})^{y}(1-\theta_{00})^{(1-y)}}, \quad y = \{0,1\}$
    Draw $N_{10}^{y}|\Theta, D_{obs} \sim \text{Binomial}\Big(N_{\{0,1\}0}^{y}, P(C_{i0}^{y} = 1|\Theta, D_{obs})\Big), \quad y = \{0,1\}$
    Set $N_{00}^{y} = N_{\{0,1\}0}^{y} - N_{10}^{y}, \quad y = \{0,1\}$
    Draw $\theta_{c1}^{s}|\Theta_{-\theta_{c1}}, D_{samp}, D_{obs} \sim \text{Beta}\big(\alpha_0 + N_{c1}^{1}, \alpha_0 + N_{c1}^{0}\big), \quad c = \{0,1\}$
    Draw $\theta_{c0}^{s}|\Theta_{-\theta_{c0}}, D_{samp}, D_{obs} \sim \text{Beta}\big(\alpha_0 + N_{c1}^{1}, \alpha_0 + N_{c1}^{0}\big)\, I_{[0,\theta_{c1})}(\theta_{c0}), \quad c = \{0,1\}$
    Draw $\pi^{s}|\Theta_{-\pi}, D_{samp}, D_{obs} \sim \text{Beta}\Big(\alpha_0 + \sum_{z,y} N_{1z}^{y}, \alpha_0 + \sum_{z,y} N_{0z}^{y}\Big)$
**end for**
**return** $\Theta^{N_{burnin}+1:N_{samples}}$
```

this sampling step based on the probability of user clicking assignment, C_{i0}^{y}. We denote these sampled counts as $D_{samp} = \{N_{c0}^{y}\}$ for $c = \{0,1\}, y = \{0,1\}$. Given the augmented user counts, $\{D_{obs}, D_{samp}\}$, we sample the parameters Θ. The sampling distributions of the user conversion probabilities for the study group and the probability of a clicking user, $\{\theta_{c1}, \pi\}, c = \{0,1\}$ are Beta distributions. For the constrained parameters, $\{\theta_{c0}\}, c = \{0,1\}$, the conditional posterior distributions become Beta distributions truncated to be non-zero at the range $[0, \theta_{c1})$ for $c = \{0,1\}$. We sample from a truncated Beta distribution using the method detailed at [65]. This sampling process is repeated for $N_{burnin} + N_{samples}$. After discarding a set of burn-in samples, N_{burnin}, a set of random samples of the posterior distribution is obtained, $\Theta^{1:Nsamples}$. Algorithm 7 illustrates this sampling process and the posterior distribution expressions.

This inference procedure allows us to estimate the variability (or heterogeneity) of the local campaign effect from the posterior random set of samples $\Theta^{1:N_{samples}}$. Thus, the local average campaign effect on the clicker, LATE_{Click}, and non-clickers, $\text{LATE}_{NoClick}$, conversions are estimated based on these posterior distribution samples as follows:

$$\begin{aligned} \text{LATE}_{Click} &= \text{E}[Y_i|C_i = 1, Z_i = 1, \theta_{11}] - \text{E}[Y_i|C_i = 1, Z_i = 0, \theta_{10}] = \theta_{11} - \theta_{10}, \\ \text{LATE}_{NoClick} &= \text{E}[Y_i|C_i = 0, Z_i = 1, \theta_{01}] - \text{E}[Y_i|C_i = 0, Z_i = 0, \theta_{00}] = \theta_{01} - \theta_{00}. \end{aligned} \tag{8.3}$$

Therefore, LATE_{Click} and $\text{LATE}_{NoClick}$ become random variables allowing us to determine their posterior credible intervals. Also, we estimate the lifts by calculating the ad exposure effect with respect to the conversion rate in the control group for both populations.

We estimate the proportion of attributed converting users for these subpopulations with respect to the converting users in the study group (ATRB_{Click}, $\text{ATRB}_{NoClick}$) as follows:

$$\begin{aligned} \text{ATRB}_{Click} &= \text{LATE}_{Click} \times \left(N_{11}^{0} + N_{11}^{1}\right) / \left(N_{01}^{1} + N_{11}^{1}\right), \\ \text{ATRB}_{NoClick} &= \text{LATE}_{NoClick} \times \left(N_{01}^{0} + N_{01}^{1}\right) / \left(N_{01}^{1} + N_{11}^{1}\right). \end{aligned} \tag{8.4}$$

These metrics provide the campaign value in terms of attributed converting users, based on the campaign effect per user and the size of the population.

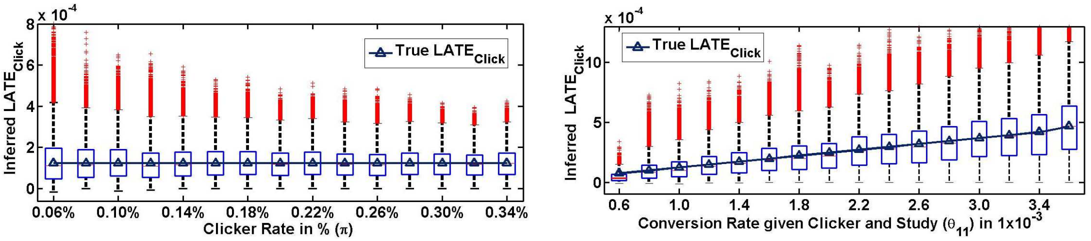
8.6 Results

8.6.1 Validation

One of the main challenges to analyzing the campaign effect on the clicker conversions is the small probability of clickers. Lewis *et al.* (2011) reported a clicker rate of 0.254% in a large-scale online experiment for more than 35 million users [55]. Even sparser is the probability of clicker and converter. In an exploratory campaign, where the user targeting is not optimized, we collect only eight clickers and converters out of more than 11 million users in the study group. This gives a 7.1e-7 joint probability of clickers and converters. Therefore, data sparsity prevents us from using large-sample approximations such as those in [4, 88] or those based on Normal approximations [48, 55], and consequently large posterior credible intervals are expected.

To analyze the power of the method to detect a given local campaign lift in the clickers, we assume a set of parameters $\{\Theta, P(Z_i), N_T\}$ (N_T is the total number of users) and simulate the data counts of Table 8.1. For each parameter set, we randomly generate 100 sampled data count sets. Then for each set, we run the inference Algorithm 7, where $N_{burnin} = 200$, $N_{samples} = 3,000$. Finally, we concatenate the posterior samples to obtain $\Theta^{1:100 \times N_{samples}}$ and calculate $\{\text{LATE}_{Click}\}^{1:100 \times N_{samples}}$ from Eq 8.3. Fig 8.2 shows the simulation results and the parameter values. These are assumed as a function of (a) the probability of a clicking user π, (b) the probability of a converting user in the study group θ_{11}. We expect that a successful campaign in optimizing clicks would increase these two parameters.

Results from these simulated data experiments show that even when we do not



(a) (b)

Figure 8.2: Boxplot of the posterior distribution of LATE$_{Click}$ based on simulated experiments as a function of: (a) the clicker rate, (b) conversion rate of the clickers in the study group. Assumed parameter values for: (a) $\{N_T$ =23e6, $P(Z)$ =0.5, θ_{11} =1e-3, θ_{01} =9.82e-5, lift LATE$_{Click}$ =0.14, lift LATE$_{NoClick}$ =0.14$\}$, (b) $\{N_T$ =23e6, $P(Z)$ =0.5, π =6.8e-4, θ_{01} =9.82e-5, lift LATE$_{Click}$ =0.14, lift LATE$_{NoClick}$ =0.14$\}$

observe the clicking user selection of the study group in the control group, we can infer the campaign effect in this sub-population without any bias. The only necessary assumption is to consider campaign positive effects, in spite of the low clicker rate. Both Fig 8.2(a)-(b) show a skew distribution[2]. However, as we increase the clicker rate π in Fig 8.2(a), the posterior distribution of LATE$_{Click}$ concentrates more at the true LATE$_{Click}$. This analysis shows that for a reasonable clicker rate of $\pi = 0.20\%$ or higher the effect distribution shows an increasingly well-defined peak. Fig 8.2 (b) shows that as the conversion rate of the clickers increases, the effect LATE$_{Click}$ increases too assuming a fixed lift. We also observe that the skewness level decreases, even though the clicker rate is low $\pi = 0.068\%$. However, the credible interval is large due to the low-clicker rate assumed. Overall, the clicker rate parameter shows to have a higher impact on the estimator power than the conversion rate of the clickers.

We note that the lifts for LATE$_{Click}$ and LATE$_{NoClick}$ are relative measures to

[2] The zero effect appears to be in the intervals because the boxplot function obtains them based on a Normal approximation. Clearly the zero effect is not in the distribution as this is a constraint of Eq 8.2.

Table 8.2: Campaign data based on notation of Table 8.1. Campaign active duration for, Campaign 1: 16 days, Campaign 2: 28 days. $\{0,1\}$ represents unobserved clicking user indicator.

Count	**Campaign 1**	**Campaign 2**
$N^0_{\{0,1\}0}$	3,621,409	11,431,495
$N^1_{\{0,1\}0}$	314	961
N^0_{01}	3,535,571	11,328,649
N^1_{01}	347	1,014
N^0_{11}	2,414	9,799
N^1_{11}	2	8

the base conversion rates, which are different between clickers and non-clickers. In terms of campaign attribution, the relevant measurements are LATE_{Click} and $\text{LATE}_{NoClick}$. Therefore, although the lifts are equivalent for the experiments of Fig 8.2, different values of LATE_{Click} and $\text{LATE}_{NoClick}$ are tested assuming different campaign attribution for these user populations.

8.6.2 Randomized Experiment Data Description

We ran two large-scale randomized experiments at the *Advertising.com* ad network collaboratively with one advertiser in the financial information services sector. We randomly assigned the users to control/study groups based on the timestamp the tracking cookie was born. To avoid user contamination, we focused on the users whose cookie was born before the campaign started. Then, for each user visit to the set of publishers' websites, the targeting engine selected those users eligible to see the ad. After this selection was made, the campaign ad was displayed to the users in the study group, and a charity ad (placebo ad) was displayed to the users in the control group. Then, the users were tracked, based on their unique cookie, to observe the user clicks on the ad and the user conversion at the advertiser's website. Fig 8.1 illustrates the randomized design, and Table 8.2 shows the aggregated user counts collected for these experiments based on the notation of Table 8.1. Although the advertiser might run the placebo and ad campaigns independently with the same targeting setup, the user randomization needs to be performed by the ad network. This process guarantees that no campaign or placebo ad is displayed to the incorrect group.

Both campaigns were run on a cost-per-thousand (CPM) business model. Thus, the user targeting was not as optimized as in the case of conversion based attribution campaigns. The goal is to explore campaign effectiveness inexpensively before the campaign is fully deployed. This practice is standard in campaign budget allocation [27]. The campaigns were run during different time periods. The same advertiser ran them, but they were otherwise unrelated. For privacy reasons, we are not allowed to disclose the ad content or the advertiser identity.

8.6.3 Campaign Evaluation Results

For comparison purposes, we estimate the ad click effect assuming we do not observe the control group of users, $\mathrm{ATE}^{obs}_{Click}$. We also estimate the ad click effect with post-treatment bias, $\mathrm{ATE}^{post}_{Click}$. Both estimations are defined as follows:

$$\begin{aligned}\mathrm{ATE}^{obs}_{Click} &= \mathrm{E}[Y_i|S_i(1)=1, Z_i=1] - \mathrm{E}[Y_i|S_i(1)=0, Z_i=1],\\ \mathrm{ATE}^{post}_{Click} &= \mathrm{E}[Y_i|S_i(1)=1, Z_i=1] - \mathrm{E}[Y_i|S_i(0)=0, Z_i=0].\end{aligned} \tag{8.5}$$

Table 8.3: Average campaign effect and attributed conversion percentage respect to the number of converting users in the study group. Results obtained from the data of Table 8.2. {*Low, Med, High*} are the {0.05, 0.5, 0.95} quantiles. We estimate C2C ATRB as the number of users who click and convert over the total converting users of the study group: $N^1_{11}/(N^1_{01}+N^1_{11})$.

	Campaign 1				**Campaign 2**			
Measure		Low	Med	High		Low	Med	High
Clicker Rate, π	(%)	0.067	0.068	0.070	(%)	0.085	0.087	0.088
ATE^{obs}_{Click} (naive)	(1e-4)	-2.33	7.30	16.92	(1e-4)	2.54	7.26	12.00
lift ATE^{obs}_{Click}	(%)	-237	**743**	1720	(%)	282	**811**	1340
$\text{ATE}^{post}_{Click}$ (biased)	(1e-4)	-2.21	7.41	17.04	(1e-4)	2.54	7.26	12.00
lift $\text{ATE}^{post}_{Click}$	(%)	-255	**855**	1960	(%)	306	**870**	**1400**
ATE_{Camp}	(1e-5)	0.37	1.20	1.99	(1e-6)	-0.33	6.13	12.50
lift ATE_{Camp}	(%)	4.06	13.94	24.02	(%)	-0.38	7.30	15.43
ATRB_{Camp}	(%)	3.74	**12.17**	20.20	(%)	-0.37	**6.80**	13.87
$\text{LATE}_{NoClick}$	(1e-5)	0.34	1.16	**2.00**	(1e-6)	0.89	5.82	**12.20**
lift $\text{LATE}_{NoClick}$	(%)	3.81	13.46	24.21	(%)	1.04	6.97	25.13
$\text{ATRB}_{NoClick}$	(%)	3.48	11.78	20.22	(%)	0.99	6.45	13.53
LATE_{Click}	(1e-4)	**0.35**	4.61	13.72	(1e-4)	**0.43**	4.65	11.04
lift LATE_{Click}	(%)	7.28	150.77	874.20	(%)	7.21	145.85	813.12
ATRB_{Click}	(%)	0.02	**0.32**	0.95	(%)	0.04	**0.45**	1.06
C2C ATRB	(%)	-	**0.57**	-	(%)	-	**0.78**	-

ATE^{obs}_{Click} provides the conversion probability change of the clickers versus the non-clickers, which is an intuitive measurement of the value of the click indicator. In a pure prediction optimization framework, this measure would represent the importance of the user click indicator feature. $\text{ATE}_{Click}^{post}$ represents the campaign effect conditional on the user clicking indicator without correcting for post-treatment bias introduced by this indicator, or its endogeneity.

Table 8.3 shows the effect results for the overall campaign, ATE_{Camp} obtained based on the user counts aggregated for the entire campaign, and the disaggregate effects $\text{LATE}_{NoClick}$, LATE_{Click} as defined by Eq 8.3. We set $N_{burnin} = 2,000$, $N_{samples} = 20,000$ for the Gibbs sampling of Algorithm 7. In addition, the non-causal estimates, ATE_{Click}^{obs} and $\text{ATE}_{Click}^{post}$, are depicted.

Table 8.2 shows that the conversion rate for the clickers in the study group is close to 10 times higher than for the non-clickers $N_{11}^{1}/(N_{11}^{1} + N_{11}^{0})$=8.27e-4 versus 9.81e-5 for campaign 1, and 8.16e-4 versus 8.95e-5 for campaign 2. ATE_{Click}^{obs} shows this (naive) effect estimation, based on the two-sample t-test with different variances[3]. We observe that the upper interval bound of lift ATE_{Click}^{obs} for both campaigns larger than 1,000%. Likewise, the lower interval bound of lift ATE_{Click}^{obs} for campaign 2 is over 300%. These results show the over-estimation of the value of user clicks when the objective is to optimize conversion prediction. Similarly, neglecting to correct the post-treatment induced bias ($\text{ATE}_{Click}^{post}$) over-estimates the campaign effect severely, which is close in magnitude to the naive estimation ATE_{Click}^{obs} (in average: lift ATE_{Click}^{obs}=743% vs lift $\text{ATE}_{Click}^{post}$=855% for campaign 1, and lift ATE_{Click}^{obs}=811% vs lift $\text{ATE}_{Click}^{post}$=870% for campaign 2). Therefore, not correcting the post-treatment bias eliminates most of the power of the randomized experiment to estimate

[3] We consider the two-sample t-test to handle different control and study populations sizes. Thus, $\bar{Y}_1 = \hat{p}_1$, $s_{\bar{Y}_1}^2 = \hat{p}_1 \times (1-\hat{p}_1)$, where $\hat{p}_1 = \sum Y_i^1/N_1$. The t-statistic is computed as: $t = (\bar{Y}_1 - \bar{Y}_2)/\sqrt{s_{\bar{Y}_1}^2/N_1 + s_{\bar{Y}_2}^2/N_2}$.

the campaign effect on the clicker conversions.

We observe a clicker rate of less than 0.1%, which is a consequence of non-optimized campaigns. As a result, the average campaign effect, ATE_{Camp}, and the average local impact in the non-clickers, $\text{LATE}_{NoClick}$, are similar because the vast majority of the users are non-clickers $C_i = 0$. We observe a larger effect for the clickers than for the non-clickers. As we discussed in section 8.6.1, we expect a skewed posterior distribution given the observed clicker rate. This skewness is evident in the lift percentage estimation where the right-hand tail is in the order of hundreds. In spite of the wide credible interval, we observe larger campaign effect in users who click on the ad. We estimate this effect by analyzing the lower quantile of LATE_{Click} and the upper quantile of $\text{LATE}_{NoClick}$ for both campaigns. Therefore, a pessimistic scenario for the campaign effect on the clicker conversions shows an increase of **75%** (3.50e-5 - 2.00e-5 with respect to 2.00e-5) for campaign 1, and **252%** (4.30e-5 - 1.22e-5 with respect to 1.22e-5) for campaign 2, with respect to the campaign effect on the non-clicker conversions. This analysis shows that, as intuition suggests, user click probability is a measure of campaign success, and the user clicks on ads are **not** random events as the previous literature suggests [28].

In terms of the overall campaign attribution, we note that a significant amount of conversions attributed to the campaign is obtained from the non-clickers due to the user volume of this sub-population. Only 2.63% of the campaign 1 attribution (ATRB_{Click}=0.32 with respect to ATRB_{Camp}=12.17) and 6.62% of campaign 2 attribution (ATRB_{Click}=0.45 with respect to ATRB_{Camp}=6.80) are associated to the clickers. Even when the effect on the clicker conversion probability is 252% higher than for the non-clickers, the volume of non-clickers is more than 900 times greater ($\pi/(1-\pi)$ =0.999/1e-3). Click-to-conversion attribution framework (C2C ATRB) is a popular industry practice that assigns the conversion credit to the campaign of the last user click. As a result, C2C does not assign any credit to the conversions of non-clickers. We calculate the C2C attribution (C2C ATRB) percentage as % $N^1_{11}/(N^1_{01} + N^1_{11})$. We note that the median campaign attribution is significantly larger than C2C attribution for both campaigns (campaign 1: ATRB$_{Camp}$=**12.17%** versus C2C ATRB=**0.57%**, campaign 2: ATRB$_{Camp}$=**6.8%** versus C2C ATRB=**0.78%**). This comparison shows that the C2C attribution practice underestimates the causal attributed value of the campaign.

8.7 Impact and Limitations

We have proposed a method to estimate the ad exposure effect on the clicker conversion probability using randomized experiments. We have shown that the ad effect evaluation is as biased as the naive observational estimation if the click-induced post-treatment bias is not addressed. The crucial limitation of our approach is that we consider ad positive effects only.

Contrary to the general belief that clicks do not measure campaign success, we find that the ad exposure effect is higher for users who click the ad. In spite of the large credible interval, a pessimistic analysis shows a substantial increase in the conversion probability for the clickers when compared to the non-clickers. These results are consistent with the most used business models based on clicks. However, campaigns also appear to increase conversions among users who do not click, so attribution methods based solely on clicks are likely to be biased against campaigns. This contradicts the general belief in the advertising industry that C2C conversion attribution models over-estimate the value of campaigns [3, 78]. We conclude that the population of clicking users is more valuable than the non-clicking population. There is a correlation between user clicks and causal effect on user conversions. Consequently, optimizing ads to increase user clicks, which are more frequent than user conversions, may increase the effectiveness of ads. However, the targeting policy should not optimize user clicks only, as a large percentage of users affected by the ad do not click on it. A combined policy to target clickers and non-clickers should be considered.

The proposed method applies more generally to study the connection between clicks and conversions using randomized experiments, including interventions designed to increase conversions by increasing clicks. Similarly, recent evidence suggests different ad exposure effects between conversion-optimized and CPM campaigns [7]. The reason users click matters: the results would be quite different for clicks that result from ads that confuse users.

There are many instances where ad exposure is randomized and click data are available, but the is not used because advertisers assume that the clicks do not reveal user intent. We have shown that better methods can use such data to advantage.

Part V

Closing Remarks

Chapter 9

Conclusion and Further Research

We have approached the evaluation of Online Display Advertising from the analysis of observational conversion and impression time series data, to the detailed randomized experimental design in a real targeting advertising system. In this chapter, we discuss further research paths based on the Dissertation contributions.

The time series method we propose in Chapter 3 requires aggregated data without the need for user tracking and features (often fragmented or incomplete). This method relies heavily on the power of the time series model to predict the daily number of conversions in the counterfactual case that the campaign is not run. Further developments of this approach include the use of campaign metadata aiming at improving this counterfactual prediction. The integration of propensity scores based methods for users with observable features and the time-series-based attribution potentially improves the accuracy of both attribution paradigms.

We have identified some drawbacks of the current industry practice to evaluate campaigns using randomized experiments in Chapter 5. One of those drawbacks includes the disregard of the effect of the campaign presence in the marketplace (or the ad slot value). Another limitation is the overlook of the user targeting confounding effect in the external validity of the estimated results. We have proposed a causal inference method in the Potential Outcomes causal framework to estimate the campaign effect on the targeted users, and to identify the targeting capability to show ads to positively influenced users. We have found evidence that performance-based CPA campaigns incentivize the targeting of users who convert regardless of the campaign ad exposure. As a consequence of this result, we have analyzed the impact of different user targeting policies to increment user conversion. One fundamental assumption of randomized design is the Stable Unit Treatment Value Assumption (SUTVA), which implies that the treatment status of any user does not affect the potential outcomes of the others. That is, no interference between users is assumed. There are some situations where this assumption is potentially violated. Particularly, the case of viral marketing oriented campaigns where the objective is to target the most influential users in a social network. Although the recent literature has focused on the feasibility of SUTVA and has proposed methods to account for network interference between individuals in the evaluation of ranking feed algorithms [46], SUTVA is largely assumed in online advertising evaluation. Besides, these methods require observable interacting features to infer the potential interference as in the case of network link information. Developing simple and accurate approximations to account for SUTVA relaxations in the context of online advertising evaluation is an open research area.

We have proposed an offline evaluation framework based on the user response prediction of the targeted and non-targeted groups of logged data of randomized experiments. Also, a methodology to improve the campaign performance mid-flight (mid-campaign before it ends) has been developed. Our results show that the standard targeting practice of serving ads to the users who are most likely to convert is similar to a uniform targeting, and significantly inferior to the optimization of the campaign causal effect. The proposed framework and results demonstrate the potential benefits and value of continuous randomized experimentation during the campaign duration, as opposed to the standard initial evaluation of the ad offer and design based on a low-budget CPM campaign. The deployment of a causally conversion generating objective function in a real targeted campaign needs to be tested. Similarly, the effective integration of logged observational ad exposure data and machine learning aiming at optimizing causally-generated conversions, as opposed to predicting conversions, is a research challenge. Also, theoretical economic studies that consider different incentives of the current business models to show the benefits of targeting with the objective of optimizing causal attribution require more research to be developed.

We have analyzed the user click behavioral feature as a potential user response that correlate with causally generated conversions. We have estimated the ad exposure effect on the conversion probability of the users who click on the ad, and compared this effect with the effect on the conversion probability of the non-clickers. We have shown that designing a focused randomized experiment to measure this effect on the clickers is not feasible and proposed to use the randomized design employed for ad exposure evaluation. We have developed a causal inference method, based on Potential Outcomes causal framework, to estimate the local ad exposure effect on the clickers. Results indicate that the ad exposure effect on the conversion probability of the clickers is significantly larger than the effect on the non-clickers based on a pessimistic analysis. As a result, optimizing user clicks maximizes the causally generated conversions by the ad exposure. The method we have proposed opens a path for more studies of the user clicks to validate further the conclusion of the current analysis. Separating the campaign presence effect and the ad effect on the clickers need to be modeled by combining click analysis approach of Chapter 8 and the targeting effect estimation of Chapter 6. Similarly, by expanding the user click effect estimate with user visits to the advertiser website, we potentially relate the impact on the clickers with those who arrive at the advertiser website. Those users might visit the advertiser page by different means (for instance by online search). Understanding what motivates a click, and why many users who are affected by the campaign do not click on the ad, is an open research problem.

We note that the campaign effect results we have estimated are mostly average effects. A different approach to conversion attribution is to consider the probability of causation for converting users given their complete user history and features, i.e. user heterogeneity [38]. In this framework, the causal inference problem is approached at the individual level, the converting user, and the goal is to determine if the campaign has caused the user conversion. The fundamental challenge is that converting users are likely to be unique, in terms of their demographic and behavioral features. Consequently, detecting the right control users becomes troublesome without further assumptions, even when the user randomization is assumed to be perfect [39]. This approach represents a further line of research, particularly when the analysis of campaign average effects might hide critical advertising practices.

To conclude, the online display advertising evaluation research in economics has focused primarily on the effect estimation with the objective of showing if a given campaign is adding value. The results are often used in a budget allocation framework at the marketing channel level. On the other hand, the user targeting research has been developed mostly in a Machine Learning framework and based on the use of large amounts of ad exposure data (big data) by predictive analytic techniques. These two problems have been addressed by two distinct research communities. As a result, the advertising effect estimation in a measurement and user targeting optimization cycle has not been discussed thoroughly. In the current Dissertation, we have approached this gap between these two research communities with the objective of improving the measurement-optimization cycle. Overall, a significant amount of research remains to be developed. Including the detailed modeling of behavioral user features in the search of user signals and responses that correlate with causally generated conversions, to effectively integrate a value generating user targeting.

Bibliography

[1] M. Aly, A. Hatch, V. Josifovski, and V. K. Narayanan. Web-scale user modeling for targeting. In *Proceedings of WWW Conference 2012*, pages 3–12. ACM, 2012.

[2] J. D. Angrist and J.-S. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 1 edition, Dec. 2008.

[3] Atlas. Engagement mapping: A new measurement standard is emerging for advertisers, 2008.

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

[5] J. Barajas, R. Akella, A. Flores, and M. Holtan. Estimating ad impact on clicker conversions for causal attribution: A potential outcomes approach. In *15th SIAM International Conference on Data Mining 2015*, pages 640–648, 2015.

[6] J. Barajas, R. Akella, and M. Holtan. Evaluating user targeting policies: Simulation based on randomized experiment data. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 11–12, Republic and Canton

of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.

[7] J. Barajas, R. Akella, M. Holtan, and A. Flores. Experimental designs and estimation for online display advertising attribution in marketplaces. *Working Paper*, 2015.

[8] J. Barajas, R. Akella, M. Holtan, J. Kwon, A. Flores, and V. Andrei. Dynamic effects of ad impressions on commercial actions in display advertising. In *Proceedings of 21st ACM CIKM*, pages 1747–1751, 2012.

[9] J. Barajas, R. Akella, M. Holtan, J. Kwon, A. Flores, and V. Andrei. Impact of ad impressions on dynamic commercial actions: value attribution in marketing campaigns. In *Proceedings of the 21st international conference companion on World Wide Web*, WWW '12 Companion, pages 459–460, New York, NY, USA, 2012. ACM.

[10] J. Barajas, R. Akella, M. Holtan, J. Kwon, A. Flores, and V. Andrei. Dynamic evaluation of online display advertising with randomized experiments: An aggregated approach. In *Proceedings of WWW Conference*, Companion, pages 115–116, 2013.

[11] J. Barajas, R. Akella, M. Holtan, J. Kwon, and B. Null. Measuring the effectiveness of display advertising: a time series approach. In *WWW (Companion Volume)*, pages 7–8, 2011.

[12] J. Barajas, J. Kwon, R. Akella, A. Flores, M. Holtan, and V. Andrei. Marketing campaign evaluation in targeted display advertising. In *ADKDD '12: Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, pages 1–7. ACM, 2012.

[13] J. Barajas, J. Kwon, R. Akella, A. Flores, M. Holtan, and V. Andrei. Measuring dynamic effects of display advertising in the absence of user tracking information. In *Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy*, ADKDD '12, pages 8:1–8:9, New York, NY, USA, 2012. ACM.

[14] R. Berman. Beyond the last touch: Attribution in online advertising, September 2013. Working Paper.

[15] P. Berthon, L. F. Pitt, and R. T. Watson. The world wide web as an advertising medium: Toward an understanding of conversion efficiency. *Journal of advertising research*, 36(1):43–54, 1996.

[16] T. Blake and D. Coey. Why marketplace experimentation is harder than it seems: The role of test-control interference. In *Proceedings of the 15th ACM EC*, pages 567–582. ACM, 2014.

[17] T. Blake, C. Nosko, and S. Tadelis. Consumer heterogeneity and paid search effectiveness: A large scale field experiment. Technical report, NBER, 2014.

[18] A. Broder and V. Josifovski. Introduction to computational advertising. Stanford University, 2011.

[19] K. L. Caballero, J. Barajas, and R. Akella. The generalized dirichlet distribution in enhanced topic detection. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pages 773–782. ACM, 2012.

[20] C. K. Carter and R. Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3):541–553, 1994.

[21] G. Casella and R. Berger. *Statistical Inference*. Duxbury Resource Center, June 2001.

[22] 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.

[23] D. Chan, Y. Yuan, J. Koehler, and D. Kumar. Incremental clicks impact of search advertising. Technical report, Google, Inc., 2011.

[24] 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 2011*, pages 1307–1315, 2011.

[25] Y. Chen, D. Pavlov, and J. F. Canny. Large-scale behavioral targeting. In *Proceedings of the 15th ACM SIGKDD*, KDD '09, pages 209–218, New York, NY, USA, 2009. ACM.

[26] D. M. Chickering and D. Heckerman. A decision theoretic approach to targeted advertising. In *Proceedings of the UAI'00*, pages 82–88, 2000.

[27] A. Chittilappilly. Using experiment design to build confidence in your attribution model. Online Metrics Insider, July 2012.

[28] B. Dalessandro, R. Hook, C. Perlich, and F. Provost. Evaluating and optimizing online advertising: Forget the click, but there are good proxies. *Big Data*, 3(2):90–102, June 2015.

[29] M. G. Dekimpe and D. M. Hanssens. Sustained spending and persistent response: A

new look at long-term marketing profitability. *Journal of Marketing Research*, 36:397–412, November 1999.

[30] L. Devroye. *Non-Uniform Random Variate Generation.* Springer-Verlag, 1986.

[31] Digiday and Google. Real-time display advertising state of the industry, February 2011.

[32] Econsultancy and Google. Marketing attribution: Valuing the customer journey, 2012.

[33] K. L. C. Espinosa and R. Akella. Incorporating statistical topic information in relevance feedback. In *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012*, pages 1093–1094, 2012.

[34] A. Farahat and M. C. Bailey. How effective is targeted advertising? In *Proceedings of the 21st WWW 2012 Conference*, pages 111–120, 2012.

[35] J. Faraway and C. Chatfield. Time series forecasting with neural networks: A comparative study using the airline data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 47:231–250, 1998.

[36] C. E. Frangakis and D. B. Rubin. Principal stratification in causal inference. *Biometrics*, 58(1):21–29, 2002.

[37] C. E. Frangakis, D. B. Rubin, M.-W. An, and E. MacKenzie. Principal stratification designs to estimate input data missing due to death. *Biometrics*, 63(3):641–649, 2007.

[38] D. A. Freedman and P. B. Stark. The swine flu vaccine and guillain-barré syndrome

a case study in relative risk and specific causation. *Evaluation Review*, 23(6):619–647, 1999.

[39] D. A. Freedman and P. B. Stark. The swine flu vaccine and guillain-barré syndrome: A case study in relative risk and specific causation. *Law and Contemporary Problems*, 64:49–62, 2001.

[40] S. Geisser, J. Hodges, S. Press, and A. ZeUner. The validity of posterior expansions based on laplace's method. *Bayesian and likelihood methods in statistics and econometrics*, 7:473, 1990.

[41] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis, Second Edition*. Chapman and Hall/CRC, 2 edition, Jul 2003.

[42] Z. Ghahramani and G. E. Hinton. Parameter estimation for linear dynamical systems. Technical report, Technical Report CRG-TR-96-2, University of Totronto, Dept. of Computer Science, 1996.

[43] A. Ghosh, R. P. McAfee, K. Papineni, and S. Vassilvitskii. Bidding for representative allocations for display advertising. *CoRR*, abs/0910.0880, 2009.

[44] A. Goldfarb. What is different about online advertising? *Review of Industrial Organization*, 44(2):115–129, 2014.

[45] A. Goldfarb and C. Tucker. Online display advertising: Targeting and obtrusiveness. *Marketing Science*, 30(3):389–404, 05-06 2011.

[46] H. Gui, Y. Xu, A. Bhasin, and J. Han. Network a/b testing: From sampling to estima-

tion. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 399–409, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.

[47] L. Ha. Online advertising research in advertising journals: A review. *Journal of Current Issues and Research in Advertising*, 30(1):31–48, 2008.

[48] J. J. Heckman. Econometric causality. Working Paper 13934, NBER, April 2008.

[49] IAB. Best practices for conducting online ad effectiveness research, June 2011.

[50] 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.

[51] H. Jin and D. B. Rubin. Principal stratification for causal inference with extended partial compliance. *Journal of the American Statistical Association*, 103(481):101–111, 2008.

[52] P. Kireyev, K. Pauwels, and S. Gupta. Do display ads influence search?: Attribution and dynamics in online advertising, 2013. Working paper (Harvard Business School).

[53] J. Kwon, R. Akella, J. Barajas, A. Flores, V. Andrei, and M. Holtan. Inferring ad influence under segmentation bias. In *JSM Proceedings, Statistics in Marketing Section*, 2012.

[54] A. Lambrecht and C. Tucker. When does retargeting work? information specificity in online advertising. *Journal of Marketing Research*, 50(5):561–576, 2013.

[55] R. Lewis, J. Rao, and D. Reiley. Here, there, and everywhere: correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of 20th ACM WWW*, pages 157–166, 2011.

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

[57] R. Lewis and D. Reiley. Online ads and offline sales: Measuring the effects of retail advertising via a controlled experiment on Yahoo!, August 2013. Working Paper.

[58] R. A. Lewis and J. M. Rao. On the near imposibility of measuring advertising effectiveness. *Working paper*, 2012.

[59] H. Li and P. Kannan. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, 51(1):40–56, 2014.

[60] S. Lysen. Incremental clicks impact of mobile search advertising. Technical report, Google, Inc., 2013.

[61] M. Mahdian and K. Tomak. Pay-per-action model for on-line advertising. *Int. J. Electron. Commerce*, 13:113–128, December 2008.

[62] P. Manchanda, J. P. Dube, K. Y. Goh, and P. K. Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43:98–108, 2006.

[63] H. Migon, D. Gamerman, H. Lopes, and M. Ferreira. Dynamic models. In *Bayesian Thinking Modeling and Computation*, pages 553 – 588. Elsevier, 2005.

[64] W. Morrison and R. Coolbirth. IAB marketplace: Networks and xchanges. Event Recap, March 2008.

[65] S. Nadarajah and S. Kotz. R programs for computing truncated distributions. *Journal of Statistical Software*, 16(2), 2006.

[66] S. Pandey, M. Aly, A. Bagherjeiran, A. Hatch, P. Ciccolo, A. Ratnaparkhi, and M. Zinkevich. Learning to target: What works for behavioral targeting. In *CIKM '11*, pages 1805–1814. ACM, 2011.

[67] K. Pauwels, I. Currim, M. Dekimpe, E. Ghysels, D. M. Hanssens, N. Mizik, and P. Naik. Modeling marketing dynamics by time series econometrics. Open access publications from katholieke universiteit leuven, Katholieke Universiteit Leuven, 2005.

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

[69] J. Pearl. Direct and indirect effects. In *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence*, UAI '01, pages 411–420, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[70] G. Petris, S. Petrone, and P. Campagnoli. *Dynamic Linear Models with R.* Springer-Verlag, 2009.

[71] R. Prado and M. West. *Time Series: Modeling, Computation, and Inference.* CRC Texts in Statistical Science. Chapman & Hall, 2010.

[72] Z. Rodgers. Pure play wins the day as forrester ranks attribution vendors. AdExchanger.com, April 2012.

[73] G. Rodriguez-Yam, R. A. Davis, and L. L. Scharf. Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. *Unpublished manuscript*, 2004.

[74] D. B. Rubin. Direct and indirect causal effects via potential outcomes*. *Scandinavian Journal of Statistics*, 31(2):161–170, 2004.

[75] D. B. Rubin. Causal inference using potential outcomes. *Journal of the American Statistical Association*, 100(469):322–331, 2005.

[76] D. B. Rubin et al. Causal inference through potential outcomes and principal stratification: application to studies with censoring due to death. *Statistical Science*, 21(3):299–309, 2006.

[77] N. Sahni, D. Zou, and P. K. Chintagunta. Effects of targeted promotions: Evidence from field experiments. *Stanford University Graduate School of Business Research Paper No. 15-4*, 2014.

[78] X. Shao and L. Li. Data-driven multi-touch attribution models. In *Proceedings of the 17th ACM SIGKDD*, pages 258–264, 2011.

[79] S. Spencer, J. O'Connell, and M. Greene. The arrival of real-time bidding. IAB, Google, Forrester, 2011.

[80] C. Tucker. The implications of improved attribution and measurability for online advertising markets, November 2012.

[81] C. Wang and R. Akella. Concept-based relevance models for medical and semantic information retrieval. In *Proceedings of the 24th ACM international conference on information and knowledge management (CIKM)*, 2015.

[82] C. Wang, R. Akella, and S. Ramachandran. Hierarchical service analytics for improving productivity in an enterprise service center. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010, Toronto, Ontario, Canada, October 26-30, 2010*, pages 1209–1218, 2010.

[83] L. Wang, G. Libert, and P. Manneback. Kalman filter algorithm based on singular value decomposition. In *Proceedings of the Decision and Control Conference 1992*, pages 1224 –1229 vol.1, 1992.

[84] P. Wang, W. Sun, D. Yin, J. Yang, and Y. Chang. Robust tree-based causal inference for complex ad effectiveness analysis. In *Proceedings of the 8th ACM WSDM*, pages 67–76. ACM, 2015.

[85] M. West and J. Harrison. *Bayesian forecasting and dynamic models (2nd ed.)*. Springer-Verlag, 1997.

[86] C. workbench team. PROMO: Simple causal effects in time series, 08 2008.

[87] T. Yildiz and S. Narayanan. Star digital: Assessing the effectiveness of display advertising, March 2013. Harvard Business Review: Case Study.

[88] J. L. Zhang and D. B. Rubin. Estimation of causal effects via principal stratification when some outcomes are truncated by death. *Journal of Educational and Behavioral Statistics*, 28(4):353–368, 2003.

[89] Y. Zhang and X. Li. Fixed-interval smoothing algorithm based on singular value decomposition. In *Proceedings of the Control Applications, 1996.*, pages 916 –921, sep 1996.