Online Tracking and Publishers' Revenues: An Empirical Analysis

Veronica Marotta[1], Vibhanshu Abhishek[2], and Alessandro Acquisti[3]

[1]Carlson School of Management, University of Minnesota, vmarotta@umn.edu

[2]Paul Merage School of Business, University California Irvine, vibs@uci.edu

[3]Heinz College, Carnegie Mellon University, acquisti@andrew.cmu.edu

PRELIMINARY DRAFT - MAY 2019*

Abstract

While the impact of targeted advertising on advertisers' campaign effectiveness has been vastly documented, much less is known about the value generated by online tracking and targeting technologies for publishers – the websites that sell ad spaces. In fact, the conventional wisdom that publishers benefit too from behaviorally targeted advertising has rarely been scrutinized in academic studies. We investigate how the (un)availability of users' cookies, which affects the ability of advertisers to perform behavioral targeting, impacts publishers' revenues. We leverage a rich dataset of millions of advertising transactions completed across multiple websites owned by a large media company. We implement an augmented version of the inverse probability weighting, a double-robust estimator that allows us to estimate effects even in the presence of potential confounding. We find that when a user's cookie is available publisher's revenue increases by only about 4%. This corresponds to an average increase of $0.00008 per advertisement. The results contribute to the current debate over online behavioral advertising, and how benefits accrued from tracking and targeting online consumers may be differentially allocated to various stakeholders in the advertising ecosystem.

*Please contact the authors for the most recent version. Acquisti gratefully acknowledges support from the Alfred P. Sloan Foundation. For a list of Acquisti's other funding sources, please visit http://www.heinz.cmu.edu/~acquisti/cv.htm.

1 Introduction

The online advertising market presents us with a puzzle. On the one hand, the market for online ads has been growing at impressive rates for several years. Much of that growth has been fuelled by the data industry's ability to target ads based on consumers' interests and preferences. A report released by the IAB estimates that advertising revenues in the US reached $88 billion in 2017, with a growth rate of 21.4% relative to 2016 (IAB, 2018). This measure of revenues is an aggregate: it captures revenues for any entity involved in the selling process. Therefore, it includes revenues for publishers (the websites that sell advertisements on their web-pages), as well as ad exchanges (which run real-time auctions for the advertisement allocation). And yet, on the other hand, reports that focus exclusively on digital publishers (the final sellers of the ad) suggest that for about 40% of them revenues are stagnant or shrinking (Econsultancy, 2015). The potential implication is that the explosive growth in revenues from online ads may not being experienced in the same manner by all stakeholders in the advertising ecosystem. Our investigation tries to provide a new piece in that puzzle. We use a unique dataset to investigate how publishers' revenues change when the ads they display can, or cannot, be behaviorally targeted to visitors of their sites.

The online advertising market has experienced a technological revolution, shifting from traditional business practices to automated, data-driven technologies for the allocation of ads. Traditionally, publishers would directly negotiate contracts with advertising firms interested in marketing their products to specific segments of consumers. Additionally, advertisers would mostly target their ads to the relevant consumers using demographic information (targeting on gender, for example) or content-based information (targeting on the content the user is consuming in a given moment, for example fitness magazines (GarlicMediaGroup, 2017)). Nowadays, almost 74% of the US digital display ad spending originates from "programmatic" advertising (eMarketer, 2017) – the "machine based buying and selling of digital media including auction based methods like Real-Time Bidding" (IAB, 2015), in which merchants (or their buying agents) bid to show ads for their products to Internet users. The introduction of increasingly personalized targeting practices has been fostered by the ability to collect consumers' information online through various tracking technologies. Behavioral advertising, for example, is the practice of monitoring people's online behaviors (websites visited, articles read, videos watched) and using the collected information to show people individually targeted advertisements (Boerman et al., 2017). Often, that tracking happens via cookies placed on Internet visitors' computers by advertising networks and websites.

These practices have been under the scrutiny of privacy advocates, and various proposals have been offered to restrict the collection and use of consumers' personal information online. In May 2018, the EU introduced the General Data Privacy Regulation (GDPR), with the objective of restoring consumers' control over their personal data. The advertising industry has raised concerns over regulatory interventions such as GDPR. By and large, the industry claims that overly stringent protection of personal information hurts Internet ad revenues and, through that, reduces the availability of free content and free services online (ITIF, 2010). Furthermore, personalized targeting—the practice of tailoring advertising based on increasingly detailed information about consumers—may become much less relevant if less consumer data is available to advertisers. As a result, consumers may be actually worse off when their privacy is protected (ITIF, 2010).

In reality, while impacts of targeted advertising on advertisers campaign effectiveness, click-through rates, and sales, have been documented, relatively little attention has been paid to the value generated by tracking and targeted practices for publishers — that is, the websites and content generators that sell advertisements spaces on their web-pages. As such, the claim that restrictions on the ability to track and target consumers would hurt the industry as a whole has not been empirically scrutinized in a comprehensive manner.

From a theoretical perspective, the ability to behaviorally target consumers (for instance, while participating in programmatic real-time auctions) may actually produce an array of diverse (even contradictory) effects on the welfare of different stakeholders. In principle, behavioral advertising gives advertisers the improved opportunity to deliver timely, relevant, and potentially highly performing advertisements (IAB, 2015). This more valuable form of advertising should (and, in fact does) command higher bids on ad exchanges by merchants: it has been theoretically shown that the ability to behaviorally target specific audiences (or segments) of consumers, in real-time, tends to increase advertisers willingness to pay for the advertisement (Chen and Stallaert, 2014). But if advertisers are paying more, publishers – the final seller of the ads – should consequently experience an increase in revenue. And yet, it has also been shown that when advertisers can highly personalize ads, they are, in fact, reaching narrower consumers' segments where the competition may be drastically reduced (Levin and Milgrom, 2010). If, in the auction, the degree of competition sufficiently decreases, the (clearing) price for the ad may decrease, potentially leading to a reduction in the publisher's revenue.

While theoretical predictions on the impact of targeted ads on publishers' revenues are mixed, empirical evaluations seem to be lacking altogether. In this paper, we investigate how the ability to track users through cookies and use their information to potentially behaviorally target them, affects publishers' revenue. We leverage a rich and novel proprietary dataset from a large media company that comprises numerous websites. The data consist of millions of "transactions" (ads purchased by advertisers and displayed to Internet visitors) completed in a week across all online outlets owned by the company. The data contains detailed information about the type of ad transaction (that is, whether the ad was sold through real-time auctions or other selling mechanisms), the features of the ad, the website where the ad was displayed, the revenue received by the site for each transaction, and whether the visitor's cookie ID is available. This last piece of information is of particular relevance, because the behavioral targeting capabilities of programmatic advertising are heavily tied to the availability of tracking cookies (AdAge, 2015).

Cookies are text files that can be stored on a user's browser and that are used to track a user's activity online (PioneerMedia, 2017). When a publisher decides to sell its inventory (the amount of ad space a publisher has available to sell on its web-pages) through programmatic auctions, he joins the advertising network of an ad exchange – an intermediary platform that (typically) holds second-price auctions for the advertisement allocation. The ad exchange is then able to set tracking cookies on the browsers of visitors to the publisher's website(s). By so doing, the ad exchange is able to identify a given user on any website that is part of the ad exchange's network, and, therefore, is able to track the user across multiple websites and collect information about her online behavior. In turn, this allows the ad exchange to make inferences about the user's interests and preferences and to classify the user as belonging to specific "audiences" or segments. For example, a user who frequently browses websites about automobiles will be classified as belonging to the consumer audience of automobile lovers. Consequently, when an auction is being held to show an ad to the user, the ad exchange can retrieve the cookie associated to the user and use the associated information to allow advertisers to behaviorally target advertisements to that user. If, however, the ad exchange is not able to set cookies on the user (because, for example, the user's browser does not allow this), the audiences to which the user supposedly belongs to may not be identified, and that information cannot be retrieved by the ad exchange during the auction. Thus, in our dataset, the ad displayed to that visitor will not be behaviorally targeted. Note that tracking cookies are not the only way to track users' behavior online. Nevertheless, in our dataset, we observe that for the ad transactions concluded without the user's cookie ID being available, the information about which audiences (segments) the user belongs to is not available. Consequently, the behavioral targeting data field is zero, indicating that the advertisement sold during the auction was *not* targeted to consumer audiences.

We exploit the presence or absence of cookies to estimate how limits to the ability to engage in behavioral advertising affects publishers' revenues. When there is no cookie associated to a given user, advertisers can still bid to display their ads to that visitor, but neither they, or the ad exchange, can behaviorally target their ads. Note that the ads can still be targeted based on other features, such as content (contextual targeting). The publisher receives as revenue some portion of those bids from the ad exchanges. We investigate how publishers' revenue changes when tracking cookies are available versus when tracking cookies are not available.

The presence or absence of cookies on a visitor's computer is affected by various factors, including the settings of the browser used by the visitor or her deliberate use of tracking-avoidance strategies such as cookie management apps. Selectivity concerns arise from the fact that, if the absence of cookies on a visitor's browser is an explicit choice by the visitor herself, her decision may be correlated with other traits which, in turn, may affect (and explain) merchants' bidding behavior. We use augmented inverse probability weighting (AIPW) to correct for those concerns (Scharfstein et al., 1999). AIPW is a double-robust estimator that allows the estimation of treatment effects in the presence of confounding factors. If visitors who choose to impede cookies being set on their browsers are systematically different from users who do not do so, and if we cannot control for such differences, traditional estimated effects will be biased. Under defined circumstances, AIPW can produce unbiased estimates (Glynn and Quinn, 2010), as we describe in detail in the empirical approach section. We additionally implement a number of robustness checks to confirm our results.

Our analysis finds that, after accounting for other factors (including those that may be used for non-behavioral forms of targeting, such as visitors device information or geolocation), when the user's cookie is available publisher's revenue increases by about 4%. The increase is significant from a statistical perspective. Nevertheless, from an economic perspective, the increase corresponds to an average increment of just \$ 0.00008 per advertisement. In Section 5.2, we contextualize such increment vis-à-vis other factors, such as the costs publishers incur to track visitors' information while staying compliant with data regulations, the privacy concerns raised by tracking technologies across website visitors, and the benefits accrued from targeting by other stakeholders in the advertising ecosystem.

While our results can be interpreted as the increase in value generated for publishers by the presence of a cookie (relative to non-cookie, controlling for other factors), they should not be assumed to capture the possible value generated by behavioral advertising as a whole (for publishers or other stakeholders). For instance, while when the cookie is not present behavioral targeting cannot be implemented in the context of our data, when the cookie *is* present advertisers may still endogenously decide whether to engage (or not) in behavioral advertising. Thus, some of the revenues associated with cookies could in theory originate from winning bids by merchants who opted not to behaviorally target their ads. In practice, the overwhelming majority of merchants' bids associated with cookies in our database did in fact use cookies (when present) for behavioral targeting.

Despite that caveat, our results contribute a novel and possibly unexpected angle to the current debate on tracking and targeting technologies, casting some light over how the purported benefits from tracking and behavioral targeting may be differentially allocated to different stakeholders in the advertising ecosystem. Both anecdotal evidence and industry reports indicate that advertising merchants can pay substantial premiums in online auctions to behaviorally target ads for their products. For instance, a recent article published in The American Prospect claimed that "an online advertisement without a third-party cookie sells for just 2 percent of the cost of the same ad with the cookie."[1] While the precise parameters of reference in the quoted article are not directly comparable to those in our dataset, if cookie-targeted ads tend to be notably more expensive for advertising merchants to buy than cookie-less ones, while publishers' revenues for displaying cookie-targeted ads increased by small margins (4%, in our dataset) over cookie-less ones (see Sections 4 and 5), research questions would arise concerning the comparative value added of various forms of targeting and the allocation of value from behaviorally targeted advertising. In turn, the online advertising industry has suggested that data regulations limiting the ability of ad companies to track users would ultimately produce nefarious implications for consumers' ability to access free online content. If publishers received, in fact, but a small fraction of the larger premium paid by merchants to behaviorally target ads, legitimate questions would also arise over the precise contours of those predictions.

[1] Laura Bassett, "Digital Media Is Suffocating - and Its Facebook and Googles Fault," The American Prospect, May 6, 2019.

2 Related Works

The literature on online advertising is ample and has focused on different themes. The great majority of existing works across different fields – including information systems, marketing, and economics – have analyzed the effectiveness of online advertising from the perspective of the advertisers: the companies that buy online advertisements to market their products to audiences of consumers.

The impact of online advertising has been measured by looking at metrics such as purchase probabilities (Manchanda et al., 2006; Sahni, 2015), click-through rates (Farahat and Bailey, 2012; Bleier and Eisenbeiss, 2015), sales (Lewis and Reiley, 2014), page views and visits (Rutz and Bucklin, 2012; Johnson et al., 2017), and online searches (Ghose and Todri-Adamopoulos, 2016; Fong, 2016). Despite measurement and attribution challenges (Lewis and Rao, 2015), many studies seem to concur that targeted advertising is beneficial and effective for advertising firms: Manchanda et al. (2006) show how the number of banner ad exposures has a positive impact on purchase probabilities, and the returns from ads that are targeted are the highest; Rutz and Bucklin (2012) show the positive impact that banner ads have on brand-specific page views; Lewis and Reiley (2014) run a large-scale, randomized experiment and find a positive effect of online ads on offline sales; and Ghose and Todri-Adamopoulos (2016) show how exposure to display ads increases users' propensity to conduct both active as well as passive searches on the advertiser's web-page.

Existing works have investigated the effectiveness of different types of targeting and analyzed how the effect of advertising varies with the degree of personalization of the ad, its timing, and other placement factors. Bleier and Eisenbeiss (2015) show that ad personalization increases click-through rates (relative to untargeted ads), especially at an early information state of the consumer's purchase decision process; Sahni (2015) investigates the impact of temporal spacing between ad exposures on the likelihood of purchase; Lambrecht and Tucker (2013) find that re-targeting may be less effective than generic brand ads, depending on whether the consumer's preference is already well-formed or not; and Fong (2016) shows how targeted offers can result in a decrease in the consumers' search activity on the retailer's website.

While the above works have focused on so-called display ads, a stream of literature has investigated search advertising and keywords bidding (Ghose and Yang, 2009; Yang and Ghose, 2010; Chen et al., 2009; Zhang and Feng, 2011; Liu et al., 2010). Finally, works have also explored the different pricing mechanisms that can be used by platforms to price online ads (Asdemir et al., 2012; Hu et al., 2015).

In our specific context, the publisher is selling display ads on its web-pages using a CPM (cost per thousand) model, where advertisements are sold (and bought) on the basis of impression. In contrast to numerous efforts at investigating the impact of online advertising on advertisers' metrics, very little work has focused on the supply-side of the market – that is, on the publishers that sell advertising inventory on their websites. From a theoretical perspective, a few papers have investigated the incentives of publishers to engage in behavioral advertising (Chen and Stallaert, 2014) or publishers' incentives to perform cookie matching (Ghosh et al., 2015). Chen and Stallaert (2014) show how, under certain circumstances, the publisher's revenue may increase (almost double) when behavioral advertising is used; in others, however, publisher's revenue may actually decrease, due to a decrease in the price paid by the advertisers. Relatedly, Ghosh et al. (2015) develops a model to investigate the incentive of publishers to perform cookie matching and share information about the websites a user has visited, making it possible to target advertisements based on prior browsing history. These works contextualize and extend results from the auction theory literature (Fu et al. (2012), Board (2009), Hummel and McAfee (2016)) that have highlighted the nuanced impact the ability to target – that is, the ability of bidders to leverage more information to form bids – can have on the outcome of the auction, in terms of the price paid by the winner and, therefore, revenue for the seller of the item being auctioned.

In summary, theoretical work in this area has shown that the ability of advertisers (the bidders in the auction) to target ads may either increase or decrease revenues for the seller (the publisher that sells ads space). To the best of our knowledge, works aimed at empirically estimating to what extent online targeted advertising (or specific types of targeting) generates value for publishers are lacking. One reason could be the difficulty in obtaining the necessary data, with the required level of granularity, in the absence of field experimentation. In this paper, we aim at filling the gap. We investigate how the presence or absence of a user's tracking cookie – which affects the ability of firms to behaviorally target ads to that user – impacts publisher's revenue from displaying the ad.

3 Institutional Details

The programmatic ecosystem is complex. It consists of multiple players, including advertisers (companies that buy ads), publishers (websites that sell ads), ad exchanges, and other advertising intermediaries that act between the advertiser and the publisher (for instance, platforms that facilitate the buying and selling of advertisements by running online auctions). Traditionally, the selling and buying of advertisements would happen through one-to-one contractual agreements between a publisher (the seller) and an advertiser (the buyer), in direct communication and negotiation with each other. Nowadays, most of the transactions are instead concluded through online platforms that work either on the buyer or the supplier side, and ad exchanges that put into communication advertisers and publishers. The buyer side comprises agency trading desks (ATDs) and demand side platforms (DSPs), technology platforms that manage advertisers' campaigns and can buy advertisements on the advertisers behalf. The supply side is comprised by supply side platforms (SSPs) that manage the publishers inventory, as well as ad exchanges and networks. Ad exchanges constitute a central component of the ecosystem. They are auction-based marketplaces that facilitate the buying and selling of ad inventory and through which the targeting of advertisements is implemented. In a lot of cases, ad exchange, DSPs, and SSPs can be integrated together, as in the case of the Google ecosystem.

The focus of this paper is on publishers. Publishers are media companies that own online outlets (such as news websites, magazines, and so forth) through which they provide content. They sell advertisement spaces available on such outlets. The objective of the publisher is to sell the available advertisement inventory to interested advertisers that would like to show ads to online users. The advent of programmatic advertising has given publishers the opportunity to manage and sell their inventory through automated online platforms. The media company that provided us with the data manages its inventory through a supply side platform that allows the publisher to sell advertisement spaces through different contractual mechanisms. These include:

1. Open Auctions: Advertisements are sold through an open auction run by the platform's ad exchange. Any interested advertiser (that is part of the same ad network) can participate and submit a bid for the impression. The ad exchange that runs the auction determines the auction closing price as the greater of the second-highest net bid in the auction or the seller's reserve price applied to that impression. The publisher is paid the closing price, net of the ad exchange revenue share.

2. Private Auctions: The publisher can decide to hold an auction, through the ad exchange, with a group of selected buyers. Differently from open auctions, private auctions give a selected group of buyers priority to inventory before it becomes available in the open marketplace. A private auction can be thought of as an auction with limited, premium advertisers. Apart from this difference, the auction mechanism works as for the open auctions.

3. Preferred Deals: This contractual option does not rely on auctions. Preferred deals take place when the publisher makes a fixed price deal with a specific buyer or advertiser. Impressions are not guaranteed using preferred deals, but the buyer gets priority and exclusive access to inventory without any commitment to purchase.

4. Programmatic Guaranteed: Similar to traditional transactions, with guaranteed deals

buyers can reserve a fixed number of impressions at a fixed price. In this case, the intermediary platform simply provides the publisher with a managing system that allows the publisher to manage all its inventory in one consolidated platform.

Private Auctions, Preferred Deals, and Programmatic Guaranteed are also referred to as Programmatic Direct. Open Auction transactions, by contrast, are concluded through Real-Time Bidding.

Our analysis focuses on Open Auctions for a number of reasons. First, Open Auctions and Real-Time Bidding (RTB) are becoming the most prominent type of programmatic advertising, accounting for almost half of the programmatic ad spending in the US (Adweek, 2017) and almost 70% of the programmatic ad spending in China (eMarketer, 2016). Second, in Open Auctions, publishers do not control the final price for the ads, as inventory is allocated through Real-Time Bidding (Research and Markets, 2016). This is important, as it allows us to investigate how the clearing price for the auction (and, therefore, the publisher's revenues, which cannot be endogenously decided by the publisher itself) changes when the user's cookie is not available in the auction for the ad. Finally, Open Auctions allow real-time targeting of ads based on users' features that are tracked through the use of cookies. While tracking cookies are not the only existing way to track users' behavior online, the behavioral targeting capabilities of programmatic advertising are still heavily tied to the availability of tracking cookies (PioneerMedia, 2017). For transactions in our dataset that were concluded through Open Auctions, when the user's cookie ID is not available, the information about which audiences (segments) the user belongs to is also not available; consequently, the behavioral targeting data field is zero, indicating that the advertisement sold during the auction was not targeted on consumer's audience data.

4 Data

We leverage an extensive data set obtained from a large media company that owns numerous newspapers and magazines, along with the related online websites. The data consist of millions of advertisement transactions completed on the online outlets owned by the company during a week in May 2016. The data include comprehensive information about: date and time of the transaction, geo-location of the visitor, type of transaction (Open Auction or other, as described above), advertisement features (such as ad type and size), seller revenue for the impression, name of the advertiser that won the impression, website where the ad was shown (with specific URL to the article where the ad was shown), whether the ad was bought from an anonymous website, and whether the user's cookie ID was available. More than 90% of the transactions in our data are concluded through Open Auctions. They constitute the focus of our analysis.

Table 1 shows descriptive statistics for the Open Auction transactions. Among those, about 91% have a cookie associated with the transaction. The average CPM (cost per thousand, paid by the advertising firms) across the sample is \$1.14; transactions with a cookie have an average CPM of \$1.18, while transactions without a cookie have an average CPM of \$0.74. Importantly, however, this difference cannot be directly interpreted as the premium that ads targeted on cookies command over ads non targeted on cookies, since these means do not account for other differences between the transactions, such as browsers used by the visitor, geo-location, and so on. We need to account for those factors in order to estimate, specifically, the increase in revenue associated to the presence of a cookie. Additionally, we need to correct for the selectivity issue that arises by the fact that visitors who do not allow cookies may be systematically different from those who do allow them. More than 70% of transactions are concluded from US IP addresses (where the visitor is browsing from); about 15% from European IP addresses; and fewer than 5%, from Australian IP addresses. The proportion of transactions without a cookie is about 8% in the US, 9% in Europe and 13% in Oceania. In terms of users' devices, most ad transactions in our data happen on desktops

(56%), followed by mobile phones (31%) and the remainder on tablets. Ad transactions that happen on mobile phones are less likely to have an associated cookie (12%), while transactions on desktops without cookie constitute about 4%. This difference may be due to the fact that cookies' abilities are limited on mobile devices and do not work consistently across mobile browsers (MarTech, 2015). Chrome and Safari are the two most prevalent browsers in our data, with Chrome being associated to about 43% of the ad transactions and Safari to about 38%. About 73% of the ads shown on a Safari browser do not have a cookie associated, whereas on Chrome this is the case about 17% of the time. The difference is probably due to different default tracking settings across the two browsers, with Safari impeding, by default, third-party tracking cookies being set on the user's machine (the user has to explicitly allow the usage of third-party cookies) (TheVerge, 2013).

All the advertisement transactions are concluded on one of the websites owned by the company. Our dataset covers 60 distinct websites, or "domains." Table 2 shows statistics for the websites included in our sample. The domains range from news websites (19%) to magazines for lifestyle and design (40%), fashion (13%), and entertainment (6%) to more specialized magazines for young women (11%), man-enthusiasts (6%), and motors (2%). Column (3) in Table 2 reports the total traffic (total number of unique visitors to the websites, in million) generated by the websites in each category since the beginning of the year and up to the week the data refer to. As of May 2016, websites in the fashion-luxury category had generated the most traffic, with about 37 million, followed by news (30 million), young women (26 million), men-enthusiast (14 million), entertainment (about 12 million), and the remaining category with about 7-8 million. The fourth column of the table reports the daily page views for each website category. The final column shows the percentage of transactions concluded without an associated cookie, in each category.

Our sample includes about 3,800 different advertisers – the companies that participate in the auctions and buy the publisher's inventory to show ads for their products. Table 3 shows descriptive statistics for the business category advertisers belong to. About 35%

of the winning advertisers are included in the retail trade category, which includes, for example, electronic shopping, department stores, and so forth. About 15% are classified as information business, including telecommunication companies, cable services, data processing services, and related companies. About 11% are classified as finance and insurance, including insurance agencies, consumer loans, and investment advisors. Among the other business categories, manufacturing follows with 10%; accommodation and food services (4%); transportation (2%) and arts, entertainment and recreation (1%). In buying publishers' inventory, advertisers can rely on different demand-side platforms (DSP) – something we also observe in our dataset. We track 108 different DSPs. Over 83% of the transactions are completed through the top five DSPs, with the largest DSP accounting for about 60% of the transactions alone.

TABLE 1: *Data Descriptives*

	Cookie	**No Cookie**	**CPM (All Sample)**
%	91%	9%	100%
Average CPM	1.18$ [1.14]	0.74$ [0.84]	1.04$ [1.12]
US (72%)	92%	8%	1.28$
EU (15%)	90%	10%	0.96$
Oceania (4.2%)	87%	13%	0.91$
Desktop (56%)	96%	4%	1.26$
Mobile (30%)	88%	12%	1.08$
Chrome (43%)	97%	3%	1.26$
Safari (38%)	84%	16%	0.97$
Explorer (8%)	97%	3%	1.29$

Table 2: *Website Descriptives*

Website Type	**%**	**Traffic**	**Daily Pageviews**	**Cookie = 0**
Lifestyle-Design	41%	8.4 M	1.2 M	7.5%
News	19%	30 M	2.2 M	9%
Fashion-Luxury	13%	37 M	1.4 M	7%
Young Women	12%	26 M	1.4 M	9.5%
Man-Enthusiast	6%	14.3 M	0.9 M	9%
Entertainment	6%	11,8 M	2.6 M	7%
Motors	1.5%	6,2 M	0.5 M	7%
Health	1%	6.6 M	0.1 M	13%
Other	0.5%	0.9 M	0.07 M	8.5%

Table 3: *Advertiser Descriptives*

Business Type	**%**	Cookie = 0
Retail Trade	35%	7%
Electronic Shopping and Mail-Order	*28%*	
Department Stores	*11%*	
Electronics Stores	*7%*	
Information	15%	5%
Telecommunications, Cable and Others	*40%*	
Software	*18%*	
Data Processing/Hosting Services	*25%*	
Finance and Insurance	11%	5%
Insurance Agencies	*35%*	
Investment Advice	*26%*	
Consumer Lending	*10%*	
Manufacturing	10%	6%
Automobiles	*39%*	
Small Electrical Appliances	*20%*	
Accommodation and Food Services	4%	5%
Transportation and Warehousing	2%	6%
Arts, Entertainment, Recreation	1%	5%

5 Analysis

5.1 Empirical Approach

Our data consist of advertising transactions concluded with the user's cookie being available or not. If the user's cookie is not available, we observe that the ads sold are not behaviorally targeted; advertisers can still bid to show their ads to visitors of a certain site or of certain content (contextual advertising) or using other observable features (such as country of origin, device type, and so forth). On the other hand, if a cookie is available then we may have transactions that are behaviorally targeted, as well as transactions that are not (in our dataset, the overwhelming majority of transactions associated with cookies did use the latter for behavioral targeting).

Variation in cookie availability allows us to separate transactions in two groups: 1) a group of transactions for which the user cookie ID is available; and 2) a group of transactions for which the user cookie ID is not available.

Even though our data are observational, we can exploit the fact that whether a user's browser allows cookies or not, it is exogenous to the publisher, ad exchange, and the advertisers. When a user is browsing a given website, her browser can allow tracking cookies to be set, or not. If cookies are allowed, the ad exchange is able to retrieve the cookie and, therefore, any information associated to the user (and tracked to the cookie), such as which behavioral audiences have been associated to the user based on his browsing behavior. If, instead, the browser does not allow cookies to be set, the ad exchange is not able to retrieve the information associated to it and behavioral targeting is not implemented. In other words, when a user is browsing a given website, the publisher sends a request to the ad exchange to run an auction to show an ad to that user. The auction proceeds regardless of whether the user has a cookie associated or not; as such, whether a transaction is concluded with or without a cookie is not a choice made by the publisher, nor the ad exchange, nor the advertisers that participate in the auction.

While the availability of cookies is exogenous to the publisher, the ad exchange, and the advertisers, it may be endogenous to the user. For instance, the visitor may be using an app to manage cookies (specifically deleting or rejecting tracking cookies); or, the visitor may have chosen a particular browser specifically for its privacy features (such as blocking cookies by default). If the decision to reject or allow cookies is made by the user, a different problem arises: users who do not allow cookies to be stored on their browser may be systematically different from users who do allow cookies. Therefore, self-selection needs to be corrected for estimates to be unbiased.

To deal with the potential selectivity issue, we implement an *augmented* version of the inverse probability weighting estimator, AIPW. Similarly to the traditional inverse probability weighting (Imbens, 2000), the AIPW involves a multi-step approach:

1. Estimate the treatment model: Estimate, for each observation, the probability of receiving the treatment. In our case, this translates into estimating the probability that a user has an associated cookie ID. The treatment model is usually estimated through a Logit.

   $$Prob_i(Cookie) = F(\beta_1 Demographics_i + \beta_2 Device_i + \beta_3 Location_i + \beta_4 X_i)$$

   where X is vector of any other included explanatory variables and F is the function used to estimated the probability model, for example, the Logit function.

2. Weight each observation with the respective probability of receiving the treatment: Let us call p, the probability of receiving the treatment, as estimated by using the model above. The observations that did receive the treatment (in our context, that have an associated cookie) are then weighted by $1/p$, whereas those that did not receive the treatment (that do not have an associated cookie) are weighted by $1/(1-p)$. Stated differently, each observation is given as weight the inverse of the probability of the treatment it received. In this way, treatment cases that resemble the control are given

more weight, and control cases that resemble the treatment group are also given more weight.

3. Estimate the outcome model: Estimate the final outcome model by OLS, using the weighted observations, and obtain the average treatment effect. More specifically, and differently from the traditional IPW, two outcome models are estimated, one for transactions without an associated cookie, and one for transactions with an associated cookie. Additionally, the two models can contain different covariates. In our context, the outcome variable is the revenue received by the publisher for a given advertisement transaction.

$$Y_i(o) = \beta_0 + \alpha Ad\_feat_i + \theta Website\_feat_i + \gamma User\_feat_i + \delta Advertisers\_feat_i + \eta X_i + \epsilon_i, o = (0,1)$$

where Y_i is the publisher's revenue for transaction (advertisement) i; Ad_feat is a vector of ad-level features; $Website\_feat$ is a vector of website-level features; $User\_feat$ is a vector of user-level features; $Advertisers\_feat$ is a vector of advertiser-specific features; and X, is a vector of any additional available covariate.

Recall that AIPW estimates two outcome models (so the model specified above is estimated twice), one for the transactions that have an associated cookie ($o = 1$) and one for the transactions that do not have an associated cookie ($o = 0$). The fundamental novelty of AIPW is that such estimator uses an augmentation term in the outcome model to correct the estimate in case the treatment model (that is, the Logit model, or any other model specified to estimate the probability that a cookie is available) is mis-specified. This implies that it is sufficient that only one of the two models (treatment or outcome) is correctly specified (that is, the specification controls for all the relevant covariates) for the estimator to be consistent – a property called *double-robustness*. In our context, it is sufficient that we correctly specify either the Logit model with which we predict the probability of the treatment, or the regression model with which we estimate the outcome model (Scharfstein et al., 1999). Additionally, AIPW has been shown to be unbiased under a true linear model outcome, and its performance is superior to traditional weighting estimators, particularly under severe confounding (Glynn and Quinn, 2010).

Our outcome of interest – the sellers revenue from a transaction – is the result of a programmatic, deterministic process (the auction) where advertisers enter their bids on the basis of a defined and finite set of information, which is the same information that we observe in our dataset. By leveraging the rich dataset at our disposal, we can control for the relevant covariates that affect our outcome of interest.

5.2 Results

The results for the AIPW estimation are shown in Table 4. The estimated potential outcome mean for transactions without cookies is \$0.93, while the potential outcome mean for transactions with cookies is \$1.02, a statistically significant difference. We run the model using a logarithmic transformation of the variable, and the results suggest that when a cookie is available, publisher's revenue increases by about 4%.

The probability of a cookie being associated to the user seems to significantly relate to technical aspects having to do with the type and model of the device employed by the user (Table 5). For example, users that use Chrome, Explorer, and Firefox are significantly more likely to have a cookie associated, compared to those who use Safari. Users who are browsing using a mobile device (either mobile phone or tablet) are less likely to have a cookie associated than users browsing using a desktop or laptop. This may reflect the fact that the cookie tracking abilities are limited in mobile environments, specifically within apps (IAB, 2013). The geographical location matters as well, as it can capture differences in laws and regulations; for example, European IP addresses are less likely to have a cookie associated, relative to US IP addresses. Similarly, Canadian, Asian, and Australian IP addresses are less likely to have a cookie associated, relative to US IP addresses. Instead, features more specific to the users, such as the topic browsed (sports, motors, fashion, etc.), do not seem to be associated with the likelihood of having a cookie or not. The treatment model also includes date and time fixed effects, website fixed effects and other features (such as browser version, operating system version, mobile phone type, mobile phone carrier, and topics browsed by the user).

TABLE 4: *Augmented Inverse Probability Weighting Estimation Results*

	AIPW			
	Coeff.(Cookie)	*Std. Errors*	*P>—z—*	*[95% Conf. Interval]*
Seller_Revenue	0.0857	0.0009	0.000	[0.0837 - 0.0876]
$E(SellerRevenue \mid cookie = 0)$	0.9341	0.0033	0.000	[0.9276 - 0.9406]
$E(SellerRevenue \mid cookie = 1)$	1.0198	0.0034	0.000	[1.0129 - 1.0266]

Table 5: *Treatment Model, AIPW*

	(1) Prob. Cookie
Mobile	-0.439***
	[0.0793]
Tablet	-0.941***
	[0.0525]
Chrome	1.035***
	[0.0438]
Explorer	1.040***
	[0.0665]
Windows	0.104*
	[0.0462]
Canada	-0.280***
	[0.0650]
Europe	-0.129***
	[0.0283]
Asia	-0.237***
	[0.0446]
Oceania	-0.415***
	[0.0382]
Sport	0.146
	[0.320]
Motors	0.0007
	[0.317]
Fashion	-0.0302
	[0.309]
Health_fitness	-0.144
	[0.310]
Travel_adventure	-0.285
	[0.324]
Entertainment	0.231
	[0.309]
Samsung	-0.127
	[0.0812]
iphone	-0.565***
	[0.0692]
Date Fixed Effects	Y
Time Fixed Effects	Y
Website Browsed Fixed Effects	Y
Other Features	Y
Constant	2.344***
	[0.307]
Observations	199999

Standard errors in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Robustness Checks

6.1 Boosted Regression Trees

In the previous section, the treatment model was estimated using a Logit. Despite the widespread use of logistic models for the estimation of propensity scores, in recent years different and more flexible methods have started receiving attention. The reason is that logistic regression models assume the covariates to be linear and additive on the log-odds scale. Interactions or non-linear terms need to be selected and included by the researcher. Nevertheless, machine learning offers more flexible models that can be used in order to capture any possible non-linearity in the probability of treatment. Among those, generalized boosted models (GBM) are becoming popular for propensity scores estimation.

GBM is a general, data-adaptive algorithm that can estimate non-linear relationship between an outcome of interest and a large number of covariates. While there exist different variants of boosting, such as the original AdaBoost (Freund and Shapire, 1997) or gradient boosting (Friedman, 2001), in our context we opt for GBMs, as they are known to produce models with well-calibrated probability estimates (Ridgeway, 1999). GBMs add together a number of simple functions with the objective of estimating a smooth function of a large number of covariates. Since we implement boosted regression trees, in our analysis each function is represented by a tree with a given depth. In other words, the use of boosted regression trees is an ensemble method that combines the strengths of two algorithms, boosting and decision trees. GBM iteratively forms a collection of regression tree models to add together to estimate the propensity score. After the propensity scores are estimated, we proceed with our analysis as before. When implementing boosted regression trees, a number of parameters need to be specified, including the number of trees and the depth (number of interactions). Following suggested best practices, we set the number of interactions to four. The algorithm reports the optimal number of trees to be around 11,000. The optimal number of trees is determined by the algorithm as the number of trees that minimizes the absolute standard

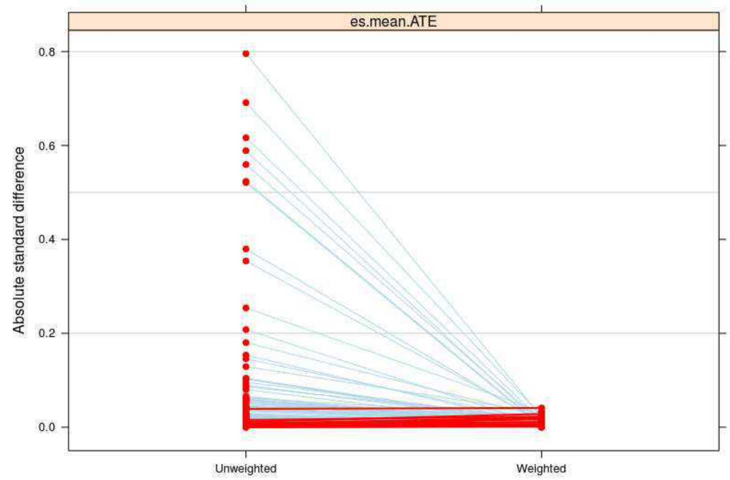

Figure 1: Absolute standard difference of unweighted vs. weighted covariates

difference between the covariates in the two groups (the group of observations with cookie = 0 and the group of observations with cookie = 1).

Figure 1 shows the absolute standard differences between the two groups before and after weighting. Before weighting, the two groups present significant differences in terms of covariates; those differences no longer seem statistically different after weighting. This is also confirmed by looking at the rank of p-values for the covariates, before and after weighting (Figure 2). If there are no significant differences in the mean values of the covariates between the two groups, then the p-values should be uniformly distributed, and follow the 45 degree line. Table 6 shows the results, which are consistent with our previous analysis.

TABLE 6: *AIPW with Boosted Regression*

	AIPW			
	Coeff.(Cookie)	*Std. Errors*	*P>—z—*	*[95% Conf. Interval]*
Seller_Revenue	0.0859	0.0009	0.000	[0.0840 - 0.0878]
$E(SellerRevenue \mid cookie = 0)$	0.9361	0.0033	0.000	[0.9296 - 0.9426]
$E(SellerRevenue \mid cookie = 1)$	1.0221	0.0035	0.000	[1.0152 - 1.0289]

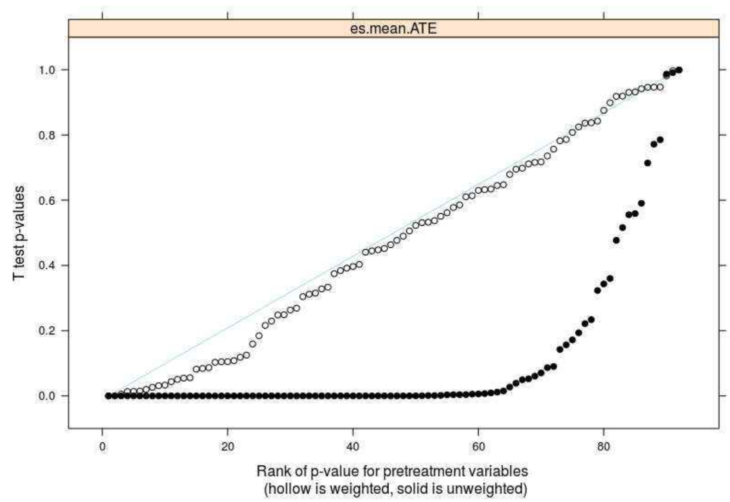

Figure 2: Rank of p-values for unweighted vs weighted variables

6.2 Including Income Data

While our dataset is rich and detailed, we do not have information about visitors' income. Nevertheless, for the transactions that happen in the United States, we know the zip-code associated to the visitor's location. Therefore, we can collect zip-code level income information and match it to our dataset. [2] The median income in our dataset is around \$61,815; the median income for transactions with cookies is \$61,612 while the median income for transactions without cookies is \$64,325. We repeat our AIPW analysis and find that results are consistent with our previous analysis.

[2] We obtain zip-code income level information from the Michigan Population Studies Center.

7 Discussion

The impressive growth of the online advertising market has been catalyzed by the introduction of tracking and targeting technologies that leverage consumer data to personalize marketing messages in real time. Targeted advertising is expected to create value for advertisers (the ad buyers), allowing them to target a preferred audience and increase advertising campaign effectiveness; as well as for publishers (the ad sellers), whose profits are expected to increase due to advertisers willingness to bid higher prices to reach specialized audiences. While the impact of targeted advertisement on advertisers campaign effectiveness has been documented, little attention has been paid to the value generated by tracking technologies for publishers. In this paper, we make a step forward by empirically investigating how a publisher's revenue from an ad changes, when the transaction is concluded without the user's cookie being available and, therefore, without the advertisers being able to perform behavioral targeting.

We leveraged a rich dataset shared by a media company that owns a number of websites and sells advertisements through online auctions. We used the augmented inverse probability weighting estimator, a double-robust estimator that allows us to estimate the impact of cookie availability on publisher's revenue, even in the presence of severe confounding. We further show that the result is robust to a number of robustness checks.

Our findings so far suggest that when the user's cookie is available, publisher's revenue increases – but the increase is just about 4%. This corresponds to an average increase of $ 0.00008 per advertisement. While the increase is significant from a statistical perspective, it may be useful to consider its economic significance. Consider a website with 500,000 sessions per day. Average number of page-views per session is 2 and average number of ads shown on each page is 4. In total, the website shows (sells) an average of 4,000,000 ads per day. If the website had to decide to not set tracking cookies on the visitors, according to our estimate, it would forfeit about $320 in revenue per day, or almost $10,000 per month. If setting tracking cookies on visitors was cost free, the website would definitely be losing money. However, the widespread use of tracking cookies – and, more broadly, the practice of tracking users online – has been raising privacy concerns that have led to the adoption of stringent regulations, in particular in the European Union. Privacy regulations put into place specific requirements and provisions pertaining to the processing of personally identifiable information (personal data) of individuals; as such, they require websites and, more generally, companies that collect, store, and manage consumers' information, to put in place appropriate technical and organizational measures that ensure the compliance with such requirements. Nevertheless, such measures come at a cost and sometimes a prohibitive one.

As an example, the International Association of Privacy Professionals has estimated that Fortune's Global 500 companies will spend roughly $7.8 billion in order to ensure they are compliant with GDPR – the recently introduced EU General Data Protection Regulation (IAPP, 2018). The regulation is affecting websites as well, and not only European ones but also US websites that have European readership. In fact, there have been cases such as the Los Angeles Times or Chicago Tribune, that have shut down their websites for European users. Or, websites such as USA Today which is offering a "European Union Experience" to European readers. In other words, the website redirects European readers to a version of the site that does not set cookies or collect any personal information about the user. By so doing, the website is unable to show personalized, targeted ads to the users and, therefore, it is forfeiting the potential extra revenue from behavioral targeting.

Important to clarify, our result can be interpreted as the value generated for publishers by the presence of a cookie, but cannot be interpreted as the value generated by behavioral advertising, even though there is a clear correlation between the two. The reason is that our empirical strategy and econometric approach helps us mitigate potential endogeneity related to the presence or not of a cookie, but cannot mitigate endogeneity related to the ex-ante choice made by advertising companies to use behavioral targeting or not. Stated differently, while when the cookie is absent, in the context of our data behavioral targeting is not implemented, when the cookie is present the winning advertiser may or may not behaviorally target its ad. That said, the overwhelming majority of merchants' bids associated with cookies in our database did in fact use cookies, when present, to behaviorally target ads.

Additionally, our data, despite being very rich and detailed, pertain to the websites of one big media company. While we believe the results to be applicable to publishers of the same size and type, they may not generalize to the entire universe of existing websites. Finally, in our dataset we do not have information on the age of a cookie; as such, we cannot explore how the value generated by the presence of a cookie changes with its age.

Despite this, our results can inform the debate on the benefits of tracking technologies and can offer insights into the allocation of benefits from behavioral advertising across different stakeholders, and into the potential implications of data regulations limiting the ability of ad companies to track users.

We plan on expanding our main result in a number of ways. First, while we cannot explore how the effect of a cookie varies with age, we explore how the effect of a cookie varies with the amount of information tied to it. The idea is that the cookie of a user who is not very active will be less informative than the cookie of a user who is more active online. Initial findings suggest that the impact of the presence of a cookie on the publisher's revenue is increasing with the amount of information tied to the cookie, but at a decreasing rate: in other words, information seem to be very valuable (from the publisher's perspective) when we compare cookies with very little information to cookies with some information; after a certain point, adding more information to a cookie does not seem to create additional value for the publisher. Second, we plan on investigating how the composition and type of advertisers change in auctions with and without user cookies. This may allow us to provide further insights on the potential mechanisms at play by investigating how the (un)availability of a cookie changes the competition in the auction.

8 Outcome Models, AIPW

Table 7: *Outcome Models, AIPW*

	(Cookie = 1) Seller Revenue	(Cookie = 0) Seller Revenue
Reserve Price	0.950***	0.933***
	[0.00539]	[0.0343]
Mobile	0.103***	0.0326**
	[0.0133]	[0.0166]
Tablet	0.0784***	0.0264
	[0.0140]	[0.0200]
Btf	-0.0200***	0.00449
	[0.00414]	[0.00518]
Firefox	-0.0676***	0.0196
	[0.0126]	[0.0244]
Chrome	-0.0515***	-0.00477
	[0.0109]	[0.0162]
Windows	-0.0358***	-0.00303
	[0.00714]	[0.0139]
South_America	-0.145***	-0.0861***
	[0.00860]	[0.0272]
Europe	0.0203***	-0.0148
	[0.00782]	[0.0115]
Asia	-0.0697***	-0.0210
	[0.0152]	[0.0185]
Female	0.0304***	0.00530
	[0.00671]	[0.00947]
Parent	-0.0351***	0.0136
	[0.0136]	[0.0119]
Image_ad	0.00290	0.0442***
	[0.00647]	[0.00979]
Date Fixed Effects	Y	Y
Time (Hour) Fixed Effects	Y	Y
DSP Fixed Effects	Y	Y
Website Fixed Effects	Y	Y
Advertiser Fixed Effects	Y	Y
User's Audiences Fixed Effects	Y	
Other Device Features	Y	Y
Other Ad Features	Y	Y
Other Content Features	Y	Y
Constant	0.404***	0.201***
	[0.0282]	[0.0649]
Observations	1836000	163640

Standard errors in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: *Balance of treatment and control*

	E(Y1|t=1)	E(Y0|t=1)	E(y0|t=0)
ddate1	0.20	0.20	0.23
ddate2	0.20	0.20	0.19
ddate3	0.20	0.20	0.20
ddate4	0.20	0.20	0.19
ddate5	0.20	0.20	0.19
morning	0.32	0.32	0.29
afternoon	0.34	0.34	0.34
evening	0.21	0.21	0.23
night	0.14	0.14	0.14
mobile	0.30	0.29	0.45
tablet	0.14	0.14	0.25
desktop	0.56	0.57	0.30
chrome_b	0.43	0.42	0.18
explorer_b	0.08	0.08	0.03
firefox_b	0.07	0.07	0.04
other_b	0.04	0.03	0.03
safari_b	0.38	0.39	0.73
apple	0.45	0.46	0.77
android	0.14	0.13	0.06
windows	0.40	0.40	0.16
other_os	0.01	0.01	0.01
vodafone	0.01	0.01	0.01
verizon	0.02	0.02	0.03
tmobile	0.01	0.01	0.01
sprint	0.01	0.01	0.01
ATTmobile	0.01	0.01	0.02
all_US	0.72	0.74	0.67
north_america	0.01	0.01	0.02
south_central_america	0.01	0.01	0.01
all_EU	0.15	0.14	0.18
all_Asia	0.04	0.04	0.04
africa	0.02	0.01	0.01
oceania	0.04	0.05	0.07
crafts	0.01	0.01	0.01
misc_man	0.00	0.00	0.01
misc_woman	0.01	0.01	0.01
miscellaneous	0.01	0.01	0.01
sport	0.01	0.01	0.01
business_career	0.00	0.00	0.00
motors	0.02	0.02	0.02
beauty	0.09	0.09	0.08
food_drinks	0.12	0.12	0.12
fashion	0.06	0.05	0.04
health_fitness	0.05	0.05	0.06
home_design	0.12	0.12	0.12
outdoor_garden	0.02	0.02	0.02
real_estate	0.00	0.00	0.01
lifestyle	0.11	0.12	0.11
society	0.00	0.00	0.00
love_relationships	0.03	0.03	0.04
travel_adventure	0.01	0.01	0.01
entertainment	0.10	0.09	0.08
celebrities	0.05	0.04	0.05
news_topic	0.17	0.18	0.18
tech_science	0.01	0.01	0.01
art_culture	0.00	0.00	0.00
comics_games	0.00	0.00	0.00
website1	0.01	0.01	0.00
website2	0.04	0.04	0.03
website3	0.09	0.09	0.10
website4	0.12	0.12	0.11
website5	0.01	0.01	0.01
website6	0.07	0.07	0.07
website7	0.06	0.05	0.05
website8	0.05	0.05	0.05
website9	0.01	0.01	0.02
website10	0.03	0.03	0.03
website11	0.09	0.09	0.06
website12	0.05	0.05	0.04
website13	0.07	0.07	0.07
website14	0.03	0.03	0.03
website15	0.01	0.01	0.01
website16	0.01	0.01	0.02
website17	0.03	0.03	0.04
website18	0.02	0.02	0.02
website19	0.01	0.01	0.01
website20	0.01	0.01	0.01
website21	0.02	0.02	0.04
website22	0.09	0.09	0.11
website23	0.02	0.02	0.02
website24	0.03	0.03	0.04

References

AdAge. Things you need to know now about programmatic buying. Technical report, 2015.

Adweek. Programmatic digital display ads now account for nearly 80% of us display spending. Technical report, 2017.

K. Asdemir, N. Kumar, and V. S. Jacob. Pricing models for online advertising: CPM vs. CPC. *Information Systems Research*, 23(3-part-1):804–822, 2012.

A. Bleier and M. Eisenbeiss. Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science*, 34(5):669–688, 2015.

S. Board. Revealing information in auctions: The allocation effect. *Economic Theory*, 38 (1):125–135, 2009.

S. C. Boerman, S. Kruikemeier, and F. J. Zuiderveen Borgesius. Online behavioral advertising: A literature review and research agenda. *Journal of Advertising*, 46(3):363–376, 2017.

J. Chen and J. Stallaert. An economic analysis of online advertising using behavioral targeting. *Management Information Systems Quarterly*, 38(2):429–449, 2014.

J. Chen, D. Liu, and A. B. Whinston. Auctioning keywords in online search. *Journal of Marketing*, 73(4):125–141, 2009.

Econsultancy. Digital publishing: Increasing advertiser value through data and identity. Technical report, 2015.

eMarketer. Real-time bidding vs. programmatic direct share of mobile programmatic display ad spending in China. Technical report, 2016.

eMarketer. US programmatic digital display ad spending 2015-2019. Technical report, 2017.

A. Farahat and M. C. Bailey. How effective is targeted advertising? In *Proceedings of the 21st International Conference on World Wide Web*, pages 111–120. ACM, 2012.

N. M. Fong. How targeting affects customer search: A field experiment. *Management Science*, 63(7):2353–2364, 2016.

H. Fu, P. Jordan, M. Mahdian, U. Nadav, I. Talgam-Cohen, and S. Vassilvitskii. Ad auctions with data. In *Algorithmic Game Theory*, pages 168–179. Springer, 2012.

GarlicMediaGroup. Behavioral targeting in. Demographic targeting out. Technical report, 2017.

A. Ghose and V. Todri-Adamopoulos. Toward a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly*, 40(4):889–910, 2016.

A. Ghose and S. Yang. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Science*, 55(10):1605–1622, 2009.

A. Ghosh, M. Mahdian, R. P. McAfee, and S. Vassilvitskii. To match or not to match: Economics of cookie matching in online advertising. *ACM Transactions on Economics and Computation (TEAC)*, 3(2):12, 2015.

A. N. Glynn and K. M. Quinn. An introduction to the augmented inverse propensity weighted estimator. *Political Analysis*, 18(1):36–56, 2010.

Y. Hu, J. Shin, and Z. Tang. Incentive problems in performance-based online advertising pricing: Cost per click vs. cost per action. *Management Science*, 62(7):2022–2038, 2015.

P. Hummel and R. P. McAfee. When does improved targeting increase revenue? *ACM Transactions on Economics and Computation (TEAC)*, 5(1):4, 2016.

IAB. Cookies on mobile 101. Technical report, 2013.

IAB. Internet advertising report. Technical report, 2015.

IAB. Internet advertising report. Technical report, 2018.

G. W. Imbens. The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710, 2000.

ITIF. Stricter privacy regulations for online advertising will harm the free internet. Technical report, 2010.

G. A. Johnson, R. A. Lewis, and E. I. Nubbemeyer. Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, 54(6):867–884, 2017.

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

R. A. Lewis and J. M. Rao. The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, 130(4):1941–1973, 2015.

R. A. Lewis and D. H. Reiley. Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on yahoo! *Quantitative Marketing and Economics*, 12(3):235–266, 2014.

D. Liu, J. Chen, and A. B. Whinston. Ex ante information and the design of keyword auctions. *Information Systems Research*, 21(1):133–153, 2010.

P. Manchanda, J.-P. Dubé, K. Y. Goh, and P. K. Chintagunta. The effect of banner advertising on internet purchasing. *Journal of Marketing Research*, 43(1):98–108, 2006.

MarTech. How the cookies crumble in a mobile-first world. Technical report, 2015.

PioneerMedia. Programmatic advertising targeting. Technical report, 2017.

Research and Markets. Real-time bidding market in the us 2016-2020. Technical report, 2016.

O. J. Rutz and R. E. Bucklin. Does banner advertising affect browsing for brands? Clickstream choice model says yes, for some. *Quantitative Marketing and Economics*, 10(2): 231–257, 2012.

N. S. Sahni. Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Quantitative Marketing and Economics*, 13(3):203–247, 2015.

D. O. Scharfstein, A. Rotnitzky, and J. M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94 (448):1096–1120, 1999.

TheVerge. Firefox to follow safari, start blocking cookies from third-party advertisers. Technical report, 2013.

S. Yang and A. Ghose. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science*, 29(4):602–623, 2010.

X. Zhang and J. Feng. Cyclical bid adjustments in search-engine advertising. *Management Science*, 57(9):1703–1719, 2011.