# The Impact of Privacy Protection on Online Advertising Markets*

Miguel Alcobendas, Yahoo Research
Shunto J. Kobayashi, Caltech
Ke Shi, Caltech
Matthew Shum, Caltech

October 6, 2023

**Abstract**

Online privacy protection has gained momentum in recent years and spurred both government regulations and private-sector initiatives. A centerpiece of this movement is the removal of third-party cookies, which are widely employed to track online user behavior and implement targeted ads, from web browsers. Using banner ad auction data from Yahoo, we study the effect of a third-party cookie ban on the online advertising market. We first document stylized facts about the value of third-party cookies to advertisers. Adopting a structural approach to recover advertisers' valuations from their bids in these auctions, we simulate a few counterfactual scenarios to quantify the impact of Google's plan to phase out third-party cookies from Chrome, its market-leading browser. Our counterfactual analysis suggests that an outright ban would reduce publisher revenue by 54% and advertiser surplus by 40%. The introduction of alternative tracking technologies under Google's Privacy Sandbox initiative would recoup part of the loss. In either case, we find that big tech firms can leverage their informational advantage over their competitors and gain a larger surplus from the ban.

*We thank Panle Jia Barwick, Steven Berry, Yunmi Kong, Leon Musolff, Michael Ostrovsky, Harry J. Paarsch, Xuan Teng, Christine Zulehner and seminar participants at Arizona State, Berkeley Haas, Bristol, Caltech, CMU, Iowa State, UCLA, Yahoo, Cowles Models and Measurement Conference 2021, Louvain Economics of Digitization webinar, Paris Digitization Conference 2021, International Industrial Organization Conference 2022, the Twenty-Fourth ACM Conference on Economics and Computation (EC'23), and TSE-Yale Regulating the Digital Economy Conference for comments.

# 1 Introduction

Privacy protection is a key topic in the current policy discussions in the digital landscape. Much of the debate surrounds the use of third-party cookies, a device long employed by internet companies to track user behavior across the web, collect user information, and target them with highly personalized ads. However, heightened concerns surrounding digital privacy have spurred policy debates and initiatives to curb the pervasive use of third-party cookies. A wave of data privacy legislation has been introduced or proposed in the European Union and across the United States to limit the use of third-party cookies.[1] In the private sector, Apple's Safari and Mozilla Firefox, two popular web browsers in the market, have disabled third-party cookies by default. Google has planned to follow suit and phase out third-party cookies in Chrome, currently the market-leading web browser. Dubbed "Cookiepocalypse" in the industry, the plan met widespread outcry and pushback and has been postponed several times because it strikes at the foundation of the online advertising market. Moreover, removing third-party cookies–a decentralized protocol–could lead to industry concentration in the online ad supply chain, triggering antitrust sirens from legislators and government agencies.[2]

In this paper, we investigate the welfare consequences of Google's plan to remove third-party cookies and introduce alternative tracking technologies under its "Privacy Sandbox" initiative.[3] Our key contribution is to quantify the unequal distributional effects on the demand side of the online advertising market, which encompasses advertisers and their intermediaries who purchase advertising opportunities and match them with advertisers. Although potentially beneficial to consumer privacy, the proposed plans could have negative spillovers in terms of information monopoly and anti-competitive practices of large companies. Removing third-party cookies will un-

[1]See the General Data Protection Regulation (GDPR) of the European Union, the California Consumer Privacy Act of 2018 (CCPA), the Colorado Privacy Act (CPA), and the Virginia Consumer Data Protection Act (VCDPA).

[2]The EU has launched an antitrust probe into Google's plan to ban third-party cookies in Chrome. In the United States, federal lawmakers have also voiced antitrust concerns over the plan in a 2020 report by the US House Subcommittee on Antitrust.

[3]For the purpose of evaluating the distributive effects of Chrome's blocking third-party cookies on various parties in the online advertising market and whether its advertising network constitutes a monopoly, this article focuses on the publishers and advertisers who are *direct* participants in the market. The welfare impact of the user side is nuanced and involves consideration of their preference for privacy, a topic subject to much debate. See Barth and de Jong (2017) for a discussion of the privacy paradox.

dermine firms' ability to target consumers and reduce the surplus of advertisers and their intermediaries. Notably, certain intermediaries, such as major tech companies like Google, can directly obtain users' behavioral information from its widely popular online products (the Google search engine, Gmail, YouTube, etc.), while other smaller intermediaries have no such recourse. Although the proposed new technology might partially offset the loss, we demonstrate that this is insufficient to diminish the information advantage enjoyed by large players.

To this end, we analyze a large sample of detailed bid-level data of online banner ad auctions from Yahoo, a prominent online news and media publisher. Online ads are sold via auctions: online publishers offer ad spaces when users access their websites, and advertisers bid to determine whose ad is shown. To streamline the process, advertisers use *demand-side platforms (DSPs)* to participate in auctions and bid on their behalf. Third-party cookies enter the process by allowing DSPs to retrieve information associated with the user and more accurately evaluate the ad opportunity. Our first set of empirical results confirms the value of third-party cookies to advertisers. We find that bidders are more likely to submit a bid and bid a higher amount in auctions with third-party cookies. Comparing DSPs' bidding decisions for users with third-party cookies to those without, we find that third-party cookies increase DSPs' bids by around 30% on average. The highest bid, which translates into the publisher's revenue, increases by as much as 80% on average.

Our primary interest is the revenue and welfare effects after Google blocks third-party cookies on Chrome and introduces alternative tracking technologies on the browser. Because the plan is yet to transpire and the bidders' underlying valuations are not observed, we adopt a structural approach to recover valuations and compute the counterfactual revenue and welfare for players in the market. Our empirical model is a first-price auction model with asymmetric bidders. We enrich the model with two essential features of the advertising market: bidder heterogeneity and auction heterogeneity. We characterize the equilibrium as a system of differential equations and adopt a numerical approach to approximate the bidding functions. The recovered valuation distributions and bidding strategies are consistent with the intuition that bidders value impressions with cookies more and bid for those more aggressively.

We then simulate the effect of "Cookiepocalypse," a third-party cookie ban on Chrome without any alternative means to track users. We consider two counterfactual specifications: a baseline *symmetric* ban in which all bidders are affected by the cookie ban and no longer receive cookie information, and an *asymmetric* ban in which one privileged bidder continues to observe cookie information for Chrome users. The second scenario emulates the information advantage enjoyed by a "Big Tech" player in the market. In the absence of third-party cookies, large firms like Google still have first-party access to user information inaccessible to other online advertising businesses. For each simulation, we solve the auction model under the counterfactual valuation distributions without third-party cookies.

We find a large negative effect worthy of the name Cookiepocalypse: in the baseline symmetric specification, such a ban would reduce the publisher's revenue by 54% and advertiser surplus by 40%. The asymmetric specification illustrates the egregiously unequal welfare distribution and anti-competitive impact of the cookie ban. The privileged bidder with exclusive access to Chrome users' data wins auctions twice as often and earns even more surplus compared to the no-ban status quo. Our results confirm and justify the antitrust concerns raised by Google's plan.

Our second counterfactual builds upon the first and introduces an alternative tracking technology that provides limited behavioral information on Chrome users. Google is developing a set of tools under the "Privacy Sandbox" initiative to replace third-party cookies. The spirit of its proposed technologies is to generate groups of users with similar interests, giving advertisers a way of targeting them without exposing details on individual users. We find that such a more privacy-friendly tracking technology would indeed soften the impact of "Cookiepocalypse" in terms of both welfare and concentration.[4] The revenue loss decreases to 13% from 54% in the first counterfactual and that advertiser surplus falls from 40% to 8%. Furthermore, although the informationally advantageous bidder still gains more surplus compared to the status quo, other bidders' performance is only mildly impacted. Our results demonstrate the importance and benefits of providing advertisers with an alternative means to target users in order to mitigate the revenue and competitive impacts of the ban.

[4]There are additional antitrust implications over Google's becoming the dominant data vendor for its Privacy Sandbox product. For instance, these concerns led to antitrust investigations by the UK and EU regulatory authorities (`https://www.wsj.com/articles/google-chrome-privacy-plan-faces-u-k-competition-probe-11610119589`). These implications, while interesting, are outside the scope of the present paper.

## 1.1 Related literature

Our paper contributes to several existing strands of literature. First, our article contributes to the literature on targeting in advertising.[5] Many empirical studies find positive effects of targeting for advertisers and publishers (Rutz and Bucklin (2012), Lewis and Reiley (2014), Ghose and Todri-Adamopoulos (2016)). Our first set of empirical results is consistent with this strand of literature. Levin and Milgrom (2010), on the other hand, discuss trade-offs in narrower versus broader (or "conflated") targeting and argue that the former thins out markets and reduces competition and prices. Rafieian and Yoganarasimhan (2021) empirically confirm this prediction and show that the optimal level of targeting is not necessarily the finest level. Our results suggest that third-party cookies do not suffer from the problem of market-thinning.

Methodologically, our empirical approach connects with the structural empirical literature on auctions.[6] We model ad auctions via a first-price auction model with a binding reserve price, and we incorporate observed heterogeneity as well as unobserved heterogeneity (Krasnokutskaya, 2011; Hu, McAdams, and Shum, 2013; Haile and Kitamura, 2019). In addition, similarly to Athey, Levin, and Seira (2011), Krasnokutskaya and Seim (2011), and Kong (2020), we allow the valuation distributions to differ across bidders to capture the observed difference in their bidding behaviors.[7] To overcome the complexities introduced by auction and bidder heterogeneity, for both estimation and counterfactual analysis, we employ Mathematical Programs with Equilibrium Constraints (MPEC) developed by Hubbard and Paarsch (2009); Hubbard, Kirkegaard, and Paarsch (2013); Hubbard and Paarsch (2014) to obtain equilibrium bidding strategies numerically.

Our work also contributes to the growing literature on the economics of privacy and data protection policies.[8] Several papers study the effect of restricting third-

[5]See Goldfarb (2014) and section 6 of Goldfarb and Tucker (2019) for reviews of this literature on targeting in online advertising.

[6]There are a number of surveys of this literature, including Hong and Paarsch (2006), Athey and Haile (2007), and Perrigne and Vuong (2019).

[7]While our study takes the existing auction format (first-price) as given, in the particular context of online ad auctions, there is a strand of theoretical literature studying auction design (Celis, Lewis, Mobius, and Nazerzadeh (2014), Abraham, Athey, Babaioff, and Grubb (2020)).

[8]See Acquisti, Taylor, and Wagman (2016) and Brown (2016) for reviews of the economics of privacy and Goldfarb and Que (2023) for a review of the economics of digital privacy. Several authors (Goldberg, Johnson, and Shriver, 2019; Aridor, Che, and Salz, 2020) study the impact of the European Union's General Data Protection Regulation (GDPR) on web traffic and ad revenue. See Johnson (2022) for a survey of studies on the economic consequences of GDPR.

party cookies in online advertising and find a loss ranging from 4 percent to 66 percent (Beales and Eisenach, 2014; Marotta, Abhishek, and Acquisti, 2019; Johnson, Shriver, and Du, 2020). The industry estimate is closer to the upper end, where a study by Google finds that disabling third-party cookies results in an average loss of 52% (Ravichandran and Korula, 2019). While most of these papers are retrospective studies using historical data, our paper provides a counterfactual scenario of the much-discussed Chrome cookie ban which, while planned, has yet to take place.

Finally, this article also connects with the emerging literature on the anti-competitive practices of big tech firms, particularly through the channel of data collection and privacy policy. Consent requirements may favor large firms (Campbell, Goldfarb, and Tucker (2015), Goldberg, Johnson, and Shriver (2019), Kesler, Kummer, and Schulte (2019)). Johnson, Shriver, and Goldberg (2022) and Peukert, Bechtold, Batikas, and Kretschmer (2022) show that the GDPR has led to a greater market concentration in the media tech industry, with Google emerging as a clear winner from the policy. Our article is the first to structurally evaluate the impact of Chrome's plan to remove third-party cookies from an antitrust point of view, connecting privacy policy with competition and demonstrating the skewed distribution of profits due to information monopoly.

# 2 Market background

## 2.1 Online ad auctions

Our analysis focuses on real-time auctions of banner ad space shown to users when they browse web pages. Banner ads are displayed in rectangular boxes between or on the side of the main text. In industry parlance, the ad space for sale is called an *impression*–each time an ad is displayed on the user's screen, it is counted as one impression. The seller is the *publisher* whose web page is browsed by the user and who has an ad space for offer (Yahoo, in our case). The bidders are *advertisers* who compete for the ad space to impress the user. The auctions are mediated through an *ad exchange*, the "auction house" for ad spaces. Auctions at the Yahoo ad exchange, which are the focus of this paper, are in the *first-price sealed-bid* format.

The process of online ad auctions can involve many parties interacting automatically in real time. The auction is triggered when the user opens the web page through her browser. The publisher packages the offer of an ad space along with information about the user and sends it to the ad exchange.[9] The ad exchange then sends out a bid request to potential bidders (DSPs), inviting them to submit a bid. Given the large volume of auctions and the complexity of online bidding, advertisers do not participate directly in these auctions, but rather via *demand-side platforms (DSPs)*, which bid on behalf of their advertiser clients.[10] Using information about the user ready to view the ad, the DSP selects the most suitable advertiser for that impression and calculates the optimal bid for the ad space, considering competition from other DSPs. In any auction, DSPs typically submit only one bid on behalf of one of their advertiser clients.[11] In what follows, we use the terms advertisers and DSPs interchangeably and abstract away from the distinction between the DSPs and their advertiser clients.

DSPs are heterogeneous based on their purpose, specialty, and scope, and in this paper, we highlight that such heterogeneity is reflected in their bidding behavior. DSPs fall into three categories: general-purpose DSPs, rebroadcasters, and specialized DSPs. *General-purpose DSPs* provide a wide range of targeting options and optimization tools to help advertisers reach their target audience. They are typically used by large and medium-sized advertisers with sizable budgets and broad campaign objectives. *Rebroadcasters*, as the name implies, rebroadcast advertising opportunities to their own ad exchanges and consolidate bids from multiple DSPs participating in them, acting as intermediaries that increase market thickness. Rebroadcasters often provide additional services to help other DSPs target users. *Specialized DSPs* focus on reaching potential customers who have indicated specialized interests or previously interacted with a brand or website. They are particularly valuable for e-commerce advertisers looking to re-engage potential customers as well as subscription-based services to retain existing subscribers.

[9] The offer is usually made through a supply-side platform server that acts on behalf of the publisher. This step is not relevant to our purpose. A data management platform could also be involved to retrieve stored information of the user that may be of interest to the advertisers. The supply-side platform packages the ad space offer with all relevant information and sends it to the ad exchange.

[10] Many major internet companies, e.g., Amazon, Facebook, and Google, own DSP services. These DSPs bid for ad spaces on their own companies' and other publishers' websites. Yahoo also maintains its own DSP.

[11] Decarolis, Goldmanis, and Penta (2020) and Decarolis and Rovigatti (2021) study the potential anti-competitive effects of the delegation between the advertisers and the DSPs.

Anticipating our empirical implementation, we further categorize general-purpose DSPs and specialized DSPs by their size as either large or small. The size of the DSP captures the budget, experience, and sophistication of the DSPs. These aspects are relevant to their valuation distributions of impressions as well as bidding strategies, which are crucial in our empirical exercise below.

## 2.2 Cookies and behavioral targeting

To make their ads more effective, advertisers use *cookies* to track user activities and implement behavioral targeting. Cookies are small files of data created by a web server and stored on the user's device when a user browses a website. Cookies contain user-associated IDs that point to entries stored in databases containing information about the user. For example, if a user visits a news website for the first time and selects English as her preferred language, the website stores this information in its server and saves a cookie file on the user's device. The next time the user visits the website, it will read the local cookie file, identify the user with the information on the database, and automatically select English as the preferred language. This type of cookie is accessible only by this specific news website and is known as a *first-party cookie* because it is hosted and used exclusively by the website. First-party cookies are generally not controversial because they improve user experience using stored information such as login credentials, settings and preferences, and items in the shopping cart.

*Third-party cookies*, on the contrary, are the subjects of intense scrutiny because of their role in user activity tracking and behavioral targeting. As the name suggests, they are cookies created by third-party entities linking to their respective databases. To continue the example above, in addition to its own content, the news website also contains bits of websites embedded by third-party servers, such as banner ads or share buttons linking to social media. These servers could also store cookies of their own to identify the user and track her activity on the news website. A distinguishing feature of third-party cookies is that they can be used to track the user's activities across a range of websites. If the user visits a retail website that hosts the same cookie and browses, say, headphones, the third-party server would store this information and match it with the same user who visited the news website earlier. This allows cross-website ad targeting as it enables an ad for the headphones she browsed on the retail website to be shown to this user during her next visit to the news website. Third-party cookies, therefore, play a critical role in behavioral targeting in online advertising and as such, are often considered an infringement on consumer privacy.

## 2.3 Privacy protection

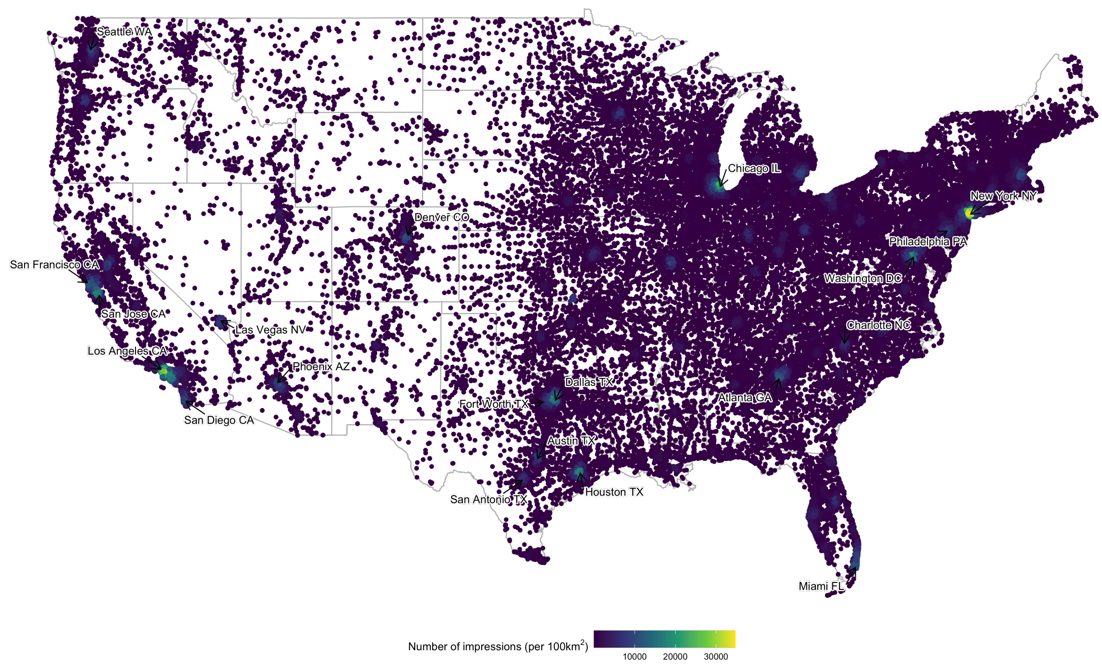
Given the controversial nature of third-party cookies and the growing concern over privacy breaches, many internet entities have either eliminated or curtailed third-party cookies in recent years. Web browsers have been at the forefront of this move. Safari and Firefox (which we refer to as the *blocked* browsers) have already blocked third-party cookies for their users and effectively shut down behavioral targeting by blocking the execution of scripts embedded by third-party servers. Third-party cookies are mostly unavailable for users of blocked browsers. On the other hand, as of 2022, Chrome, together with a few other browsers including Microsoft Edge (the *allowed* browsers), still enables third-party cookies by default. Third-party cookies are generally available on these browsers but could still be absent for a host of reasons.[12]

In addition to private-sector initiatives, the CCPA and other similar privacy regulations require large websites like Yahoo to implement a "Do Not Sell My Personal Information" link that enables users to opt out of the sale of their personal information. Under such an opt-out arrangement, publishers are not allowed to monetize the user's personal information (cookie, IP address, or precise geo-location data) by sharing it with third parties.[13] When cookies are no longer employed, DSPs have significantly less information about users and cannot engage in accurate behavioral ad targeting. In our empirical analysis below, we will exploit the variation in third-party cookie availability to evaluate the effect of behavioral ad targeting.

[12]For example, third-party cookies could be unavailable if the user chooses to block third-party cookies in their browser settings, or browses in private (incognito) mode, or has recently cleared cookies in her browser.

[13]Internet companies can still use broad geographical location (e.g., city) and contextual information of ad opportunities coming from these users for targeted ads at a broader stroke.

Figure 1: Geographical distribution of impressions




*Note*: Each dot represents the number of impressions originating within the 10 by 10 km$^2$ area around the dot during the week.

# 3 Data and Descriptive Statistics

We employ bidding data from banner ad auctions on sixteen websites of Yahoo, including Homepage, News, Finance, etc. We focus on a specific display ad format known as medium rectangular (MREC) units, which has the dimension 300×250 and is displayed to the right of the main content. This is one of the most popular ad formats, and the fixed size and position help us eliminate potential heterogeneity arising from these aspects. We consider a sample of user impressions from the United States during one week in May 2022. Figure 1 shows the geographical distribution of our sample, which roughly coincides with the population density of the US. The dataset consists of over 5.5 million bids from about 740,000 auctions.

Table 1 presents summary statistics of key variables in the dataset. The variable *bid* is the submitted bid price of an individual DSP. For reasons of confidentiality, we

Table 1: Summary statistics

| Variable | No. observations | Pct. missing | Mean | Std. Dev. | Min | Median | Max |
|---|---|---|---|---|---|---|---|
| *Auction:* | | | | | | | |
| Bid | 5,529,489 | 0.000 | 1.000 | 1.692 | 0.064 | 0.589 | 275.760 |
| No. bidders | 736,745 | 0.000 | 7.505 | 4.745 | 1.000 | 7.000 | 26.000 |
| Winning (highest) bid | 736,745 | 0.000 | 2.052 | 3.206 | 0.064 | 1.211 | 275.760 |
| *Cookie availability:* | | | | | | | |
| Pct. cookie matched | 736,745 | 0.000 | 0.577 | 0.404 | 0.000 | 0.800 | 1.000 |
| Cookie matched | 736,745 | 0.000 | 0.689 | 0.463 | 0.000 | 1.000 | 1.000 |
| *Privacy:* | | | | | | | |
| Opt-out | 736,745 | 0.000 | 0.089 | 0.284 | 0.000 | 0.000 | 1.000 |
| Blocked | 736,745 | 0.000 | 0.215 | 0.400 | 0.000 | 0.000 | 1.000 |
| *Device:* | | | | | | | |
| Computer | 736,745 | 0.000 | 0.968 | 0.177 | 0.000 | 1.000 | 1.000 |
| *Demographics:* | | | | | | | |
| Female | 736,745 | 0.000 | 0.125 | 0.331 | 0.000 | 0.000 | 1.000 |
| Male | 736,745 | 0.000 | 0.146 | 0.353 | 0.000 | 0.000 | 1.000 |
| Gender unknown | 736,745 | 0.000 | 0.729 | 0.444 | 0.000 | 1.000 | 1.000 |
| Age 24 and below | 736,745 | 0.000 | 0.001 | 0.031 | 0.000 | 0.000 | 1.000 |
| Age 25 to 44 | 736,745 | 0.000 | 0.053 | 0.225 | 0.000 | 0.000 | 1.000 |
| Age 45 to 64 | 736,745 | 0.000 | 0.120 | 0.325 | 0.000 | 0.000 | 1.000 |
| Age 65 and above | 736,745 | 0.000 | 0.064 | 0.245 | 0.000 | 0.000 | 1.000 |
| Age unknown | 736,745 | 0.000 | 0.761 | 0.426 | 0.000 | 1.000 | 1.000 |
| *Proxies for user information:* | | | | | | | |
| Interest segments (10,000s) | 736,745 | 0.581 | 2.558 | 1.100 | 0.000 | 2.551 | 8.741 |
| Months monetized | 736,745 | 0.580 | 29.118 | 24.225 | 0.000 | 32.000 | 55.000 |
| Total revenue (normalized) | 736,745 | 0.580 | 0.000 | 1.000 | -0.685 | -0.370 | 94.183 |
| Average revenue (normalized) | 736,745 | 0.580 | 0.003 | 0.063 | -26.035 | 0.000 | 1.712 |
| Days in database (10,000s) | 736,745 | 0.725 | 1.742 | 0.510 | 0.000 | 1.912 | 1.912 |

normalize the submitted bids to have a sample mean equal to 1. For every auction, we observe the *number of bidders* (out of a total of 33 DSPs) who entered the auction and submitted a bid, as well as the *winning (highest) bid*. There is substantial variation in the number of actual bidders for each auction, with a mean of 7.5 bidders and a standard deviation of 4.7. Our empirical model will factor in this important behavioral pattern and account for bidders' entry decisions.

Two key variables describe the availability of third-party cookies for each impression. The variable *percentage of cookie matched* is the percentage of DSPs in each auction who matched the user with a profile in the bidders' database constructed with third-party cookies. Small percentages of cookie matched indicate that less information is available for the user.[14] The variable *cookie matched* is a binary variable indicating whether the percentage of cookie matched is nonzero for the impression.

[14]Cookie-match information is unavailable for two of the DSPs in our sample; for that reason, we do not model cookie availability at the user-DSP level but rather construct the aggregate measure at the user level.

In other words, it indicates whether at least one bidder has a cookie identifier for the user. For ease of interpretation, our empirical analysis will primarily focus on this variable. In what follows, we refer to impressions with cookie matched = 1 as "cookie impressions" and those with cookie matched = 0 as "cookieless impressions."

The variable *opt-out* is a binary variable indicating if the user opts out of behavioral targeting. The variable *blocked* is a binary variable indicating if a browser blocks third-party cookies by default i.e. it is equal to 1 for Safari or Firefox and 0 for other browsers. About 9% of auctions are for opt-out impressions, while 20% of auctions involve impressions using browsers that block third-party cookies.

We include additional characteristic variables indicating the amount of information available on the user. Yahoo's database of user profiles (including those without Yahoo accounts) contains its best guess (based on machine learning procedures) of the user's characteristics and proxies well for the user-specific information that can be inferred from third-party cookies. These include *gender* and *age* categories. The variable *interest segments* (in 10,000s) tallies the total number of interest segments that the user belongs to, where each segment is a prediction of the user's likely interest in a particular subject (e.g., automobile, basketball, gardening, etc.) The variable *months monetized* is the number of months that the user has been monetized by Yahoo, and the *total revenue* and *average revenue* are the total and average monthly revenue derived from the user, respectively, where total revenue is normalized with mean 0 and standard deviation 1. Finally, the variable *days in database* (in 10,000s) is the number of days for which the user profile has existed in Yahoo's database. A smaller number of days may imply that less information is available for the user.[15]

[15] We note two caveats of these user-specific variables. First, while these variables quantify the user information observed by Yahoo's DSP, in our empirical analysis, we use these variables to proxy for what *any DSP* knows about these users, i.e., we assume all the DSPs observe the same information as Yahoo. Without data from other DSPs, it is impossible to validate this assumption; however, since many of the users in our dataset have registered Yahoo accounts, we believe that the information that Yahoo has on these users represents a "best case" (upper-bound) on the information that any DSP might have on these users.

Second, we observe a large incidence of missing data: about 70% of the users have unknown age and gender information. As age and gender are typically inferred indirectly from users' internet activities using machine learning algorithms, missing values for these variables typically imply that not enough tracking information is known about these users to permit reliable inference. Furthermore, the variables interest segments, months monetized, and revenue are unknown for around 60% of the analyzed users. The lack of such information is often due to users opting out or using browsers that block third-party cookies. To address this problem and as a robustness check, we have also implemented our empirical analysis on the subsample of users with a Yahoo account, for which the overall incidence of missing data is lower, and confirmed the robustness of our results.

Table 2: Summary statistics of impression characteristics by browser

| | Chrome | Edge | Safari | Firefox | Other |
|---|---|---|---|---|---|
| Proportion | 0.576 | 0.199 | 0.109 | 0.106 | 0.009 |
| Cookie matched | 0.869 | 0.847 | 0.000 | 0.000 | 0.842 |
| Opt-out | 0.088 | 0.097 | 0.056 | 0.121 | 0.033 |
| Female | 0.137 | 0.139 | 0.000 | 0.000 | 0.049 |
| Male | 0.154 | 0.170 | 0.000 | 0.000 | 0.065 |
| Gender unknown | 0.709 | 0.691 | 1.000 | 1.000 | 0.886 |
| Age 24 and below | 0.001 | 0.001 | 0.000 | 0.000 | 0.000 |
| Age 25 - 44 | 0.074 | 0.050 | 0.000 | 0.000 | 0.015 |
| Age 45 - 65 | 0.153 | 0.150 | 0.000 | 0.000 | 0.055 |
| Age 65 and above | 0.068 | 0.117 | 0.000 | 0.000 | 0.048 |
| Age unknown | 0.703 | 0.681 | 1.000 | 1.000 | 0.881 |

In addition to the user-specific characteristics, we observe variables associated with the origination of the impression. These include the *time (hour)* and the *city* of the impression, the *website* (a total of 16 including Yahoo Homepage, News, Finance, etc.) that published the impression, the device (*computer*) which indicates the user browsed with either a computer or a smartphone/tablet, and the *browser* (Safari, Firefox, Edge, Chrome, and others) with which the user accessed the web page.

Because our analysis focuses on the impact of Google's plan to terminate third-party cookies on Chrome, in Table 2, we show the mean statistic of key impression characteristics, broken down by browser. Importantly, Chrome accounts for almost 60% of the impressions in our data and dominates the browser industry, suggesting a substantial impact of Google's plan on the market. Impressions from Safari and Firefox, the two browsers that ban third-party cookies by default, account for roughly 20% of impressions. Accordingly, impressions from Safari and Firefox are missing third-party cookie information, i.e., *cookie matched* $= 0$, and gender and age information is unavailable for these impressions as well.

## 3.1 Bidding patterns

Table 3 presents a comparison of summary auction statistics for impressions with and without cookies. For either category, we calculate the averages and standard deviations (in parentheses) of the bid, the winning bid, the number of bidders, and the

Table 3: Comparison between auctions with and without third-party cookies

| Variable | Cookie impressions | Cookieless impressions |
|---|---|---|
| Bid | 1.041 | 0.764 |
| | (1.715) | (1.566) |
| Winning bid | 2.454 | 1.166 |
| | (3.487) | (2.278) |
| No. bidders | 9.283 | 3.558 |
| | (4.315) | (2.933) |
| Entry probability | 0.265 | 0.102 |
| | (0.441) | (0.302) |

*Notes:* The mean values are reported in both columns and standard deviations are in parentheses below. The bid is averaged at the bid level. The winning bid and the number of bidders are averaged at the auction level. Entry probability is calculated by first constructing a binary variable *Entry* for every auction-bidder pair. It is equal to 1 if the bidder submitted a bid in the auction.

entry probability. Impressions with third-party cookie identifiers fare better for all variables of interest. In particular, submitted bids on average are about 25% higher (1.0 versus 0.76) for cookie impressions, and winning bids for cookie impressions are over two times higher (2.5 versus 1.2) than cookieless impressions. The difference arises from both higher submitted bids and a larger number of participating bidders, with bidders more than twice as likely to enter auctions for cookie impressions. Finally, the standard deviation of bids and winning bids are higher for cookie impressions. This is expected because DSPs have the most information on these users, which increases the targeting opportunities and hence, the variation in advertisers' bids.

Figure 2a shows the empirical CDFs of submitted bids in the dataset for five categories of DSPs in the dataset by their type and size (as discussed in Section 2.1): 5 large general-purpose, 10 small general-purpose, 9 rebroadcaster, 3 large specialized, and 6 small specialized DSPs. Consistent with the results in Table 3, DSPs tend to bid higher for cookie impressions. In fact, the distribution of bids for cookie impressions first-order stochastically dominates that for cookieless impressions. Figure 2a also shows heterogeneity in submitted bid distributions among different groups of DSPs. The differences are driven by a few factors: Large DSPs generally have better access to user information, have more budget and experience, and are more sophisticated in matching advertisers with impressions. Specialized DSPs could focus on some areas of advertising, such as retailing or reconnecting with existing customers (e.g. retargeting). In terms of auction participation, Figure 2b displays the frequencies with which the five groups of DSPs participate in auctions for impressions with and without third-party cookies, and it also highlights heterogeneity in entry behavior across DSPs. The observed heterogeneity among the bidding DSPs motivates us to adopt an auction model with asymmetric bidders in the structural estimation exercise discussed below.

## 3.2 Evidence of the value of third-party cookies

Next, we present reduced-form evidence of the value of third-party cookies to advertisers. Specifically, we run regressions of the following form:

$$y_i = \beta_c \text{Cookie}_i + \mathbf{x}_i' \boldsymbol{\beta} + \alpha_i + \epsilon_i \tag{1}$$

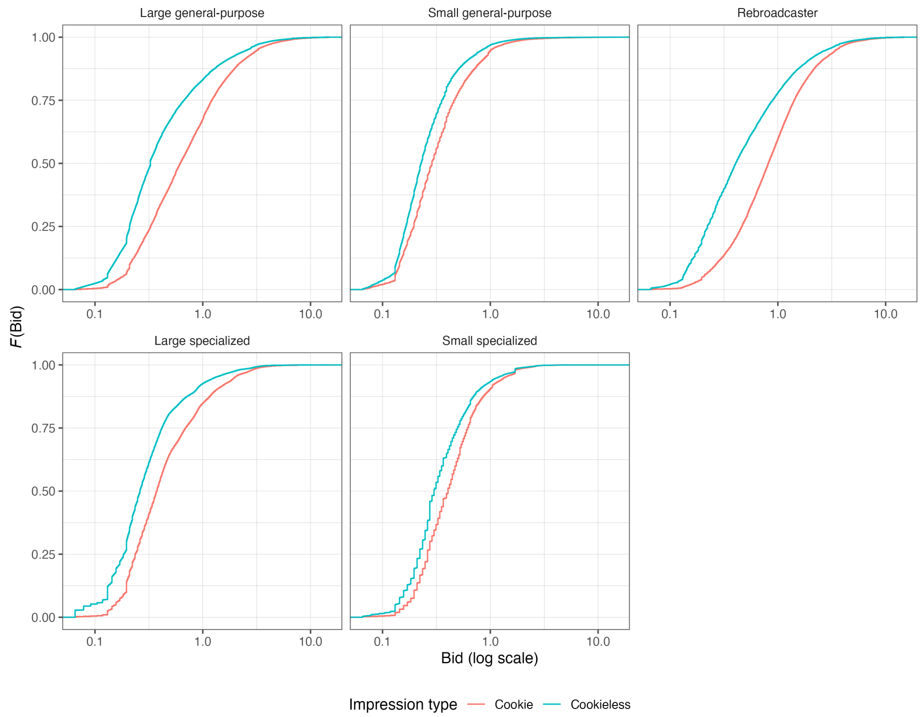
where $i$ indexes a bidder or an auction depending on the model, $y_i$ is the outcome variable to be specified later, $\text{Cookie}_i$ indicates if third-party cookies are available for the impression, $\mathbf{x}_i$ is a vector of covariates that include gender and age information as well as proxies for the amount of information available on the user, $\alpha_i$ includes fixed effects of the hour in the day, the city, the website, and the browser. For models at the bidder level, we also include a DSP fixed effect to capture bidder heterogeneity. Standard errors are clustered by the hour, the city, and the website to account for potential correlations. The variable of interest is $\text{Cookie}_i$, where a positive and significant estimate of $\beta_c$ would indicate the value of third-party cookies to the advertisers.

We first analyze the effect of cookie availability on submitted bids by taking the outcome variable $y_i = \log(\text{Bid}_i)$ for every bid $i$ in equation 1. Table 4 columns (1)-(3) report the results of three alternative specifications. Column (1) includes only cookie availability and fixed effects; column (2) adds additional covariates; column (3) further adds a DSP fixed effect to account for bidder heterogeneity. We find quantitatively similar results in these models: having third-party cookies increases submitted bids by around 30%.

Next, we take the outcome variable $y_i = \log(\text{Winning bid}_i)$ for each auction $i$ in equation 1 to examine the effect of cookie availability on the highest bid, which

Figure 2: Cookie vs. cookieless: observed bidders' behavior by DSP group

(a) Empirical CDFs of submitted bids (log scale)




(b) Average entry frequencies

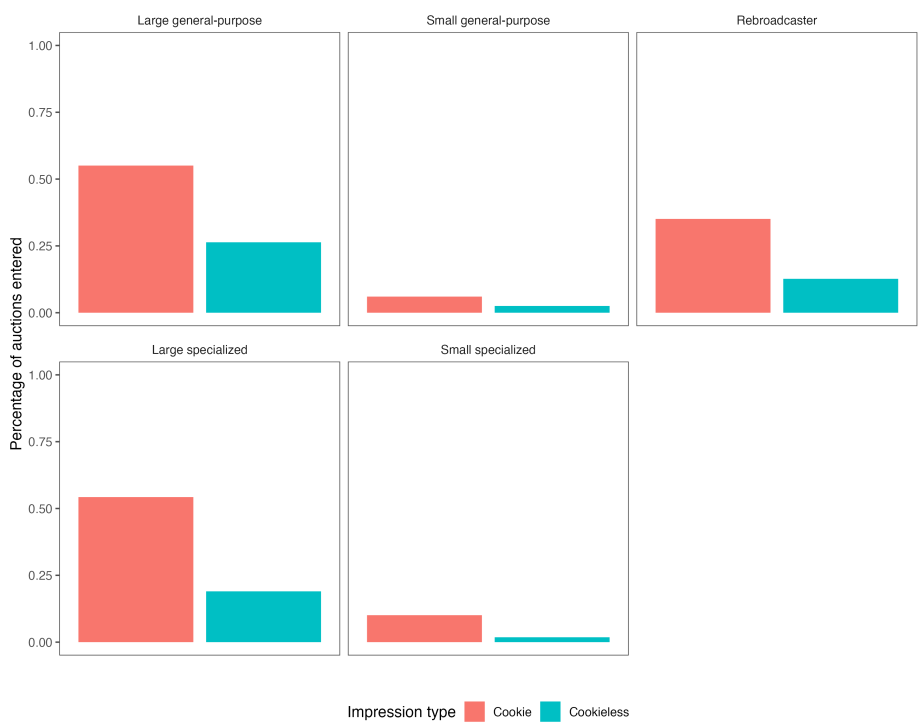


Table 4: Regression results for submitted bids and winning bids

| Dependent Variables: | | log(Bid) | | log(Winning bid) | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Cookie | 0.335*** | 0.318*** | 0.314*** | 0.887*** | 0.783*** |
| | (0.028) | (0.046) | (0.031) | (0.018) | (0.042) |
| Opt-out | | 0.013 | -0.004 | | -0.021 |
| | | (0.026) | (0.020) | | (0.024) |
| Computer | | -0.231*** | -0.177*** | | -0.397*** |
| | | (0.028) | (0.030) | | (0.055) |
| Gender female | | 0.097*** | 0.095*** | | 0.260*** |
| | | (0.012) | (0.009) | | (0.018) |
| Gender male | | 0.069*** | 0.064*** | | 0.221*** |
| | | (0.010) | (0.007) | | (0.014) |
| Age 24 and below | | 0.066*** | 0.050*** | | -0.054 |
| | | (0.008) | (0.009) | | (0.033) |
| Age 25 to 44 | | 0.015* | 0.009 | | -0.100*** |
| | | (0.008) | (0.007) | | (0.016) |
| Age 45 to 64 | | -0.010 | -0.015** | | -0.141*** |
| | | (0.006) | (0.006) | | (0.019) |
| Age 65 and above | | -0.022** | -0.031*** | | -0.178*** |
| | | (0.009) | (0.008) | | (0.016) |
| Interest segments | | -0.002 | 0.002 | | 0.044*** |
| | | (0.001) | (0.001) | | (0.005) |
| Months monetized | | 0.003*** | 0.002*** | | 0.003*** |
| | | (0.000) | (0.000) | | (0.000) |
| Total revenue (normalized) | | -0.037*** | -0.027*** | | -0.043*** |
| | | (0.002) | (0.002) | | (0.003) |
| Days in database | | -0.051*** | -0.044*** | | -0.053*** |
| | | (0.004) | (0.004) | | (0.005) |
| *Fixed-effects* | | | | | |
| Time (hour) | Yes | Yes | Yes | Yes | Yes |
| City | Yes | Yes | Yes | Yes | Yes |
| Website | Yes | Yes | Yes | Yes | Yes |
| Browser | Yes | Yes | Yes | Yes | Yes |
| DSP | | | Yes | | |
| Observations | 5,529,489 | 5,529,489 | 5,529,489 | 736,745 | 736,745 |
| Adjusted $R^2$ | 0.10623 | 0.11052 | 0.31362 | 0.24918 | 0.26361 |

*Notes:* The base levels for age and gender are both Unknown. Standard errors are clustered by the hour of the day, the city, and the website and are heteroskedasticity-robust. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively.

Table 5: Regression results for the number of bidders and entry decision

| Dependent Variables: | No. bidders | | Entry | | | |
|---|---|---|---|---|---|---|
| | (1) OLS | (2) OLS | (3) OLS | (4) OLS | (5) OLS | (6) Logit |
| Cookie | 5.796*** | 5.220*** | 0.161*** | 0.145*** | 0.145*** | 0.184*** |
| | (0.281) | (0.259) | (0.008) | (0.007) | (0.007) | (0.013) |
| Opt-out | | 0.098 | | 0.003 | 0.003 | 0.018* |
| | | (0.133) | | (0.004) | (0.004) | (0.010) |
| Computer | | -0.926*** | | -0.024*** | -0.024*** | -0.045*** |
| | | (0.103) | | (0.003) | (0.003) | (0.006) |
| Gender female | | 0.768*** | | 0.021*** | 0.021*** | 0.035*** |
| | | (0.046) | | (0.001) | (0.001) | (0.003) |
| Gender male | | 0.327*** | | 0.009*** | 0.009*** | 0.021*** |
| | | (0.054) | | (0.002) | (0.002) | (0.004) |
| Age 24 and below | | -0.182* | | -0.005* | -0.005* | -0.018*** |
| | | (0.090) | | (0.003) | (0.003) | (0.004) |
| Age 25 to 44 | | -0.498*** | | -0.014*** | -0.014*** | -0.020*** |
| | | (0.071) | | (0.002) | (0.002) | (0.004) |
| Age 45 to 64 | | -0.627*** | | -0.017*** | -0.017*** | -0.023*** |
| | | (0.094) | | (0.003) | (0.003) | (0.005) |
| Age 65 and above | | -0.805*** | | -0.022*** | -0.022*** | -0.028*** |
| | | (0.099) | | (0.003) | (0.003) | (0.005) |
| Interest segments | | 0.432*** | | 0.012*** | 0.012*** | 0.012*** |
| | | (0.059) | | (0.002) | (0.002) | (0.002) |
| Months monetized | | 0.019*** | | 0.001*** | 0.001*** | 0.000*** |
| | | (0.002) | | (0.000) | (0.000) | (0.000) |
| Total revenue (normalized) | | -0.180*** | | -0.005*** | -0.005*** | -0.006*** |
| | | (0.012) | | (0.000) | (0.000) | (0.000) |
| Days in database | | -0.256*** | | -0.007*** | -0.007*** | -0.008*** |
| | | (0.021) | | (0.001) | (0.001) | (0.001) |
| *Fixed effects* | | | | | | |
| Time (hour) | Yes | Yes | Yes | Yes | Yes | Yes |
| City | Yes | Yes | Yes | Yes | Yes | Yes |
| Website | Yes | Yes | Yes | Yes | Yes | Yes |
| Browser | Yes | Yes | Yes | Yes | Yes | Yes |
| DSP | | | | | Yes | |
| Observations | 736,745 | 736,745 | 26,522,820 | 26,522,820 | 26,522,820 | 2,652,282 |
| Adjusted $R^2$ | 0.44635 | 0.46616 | 0.04701 | 0.04908 | 0.26756 | |

*Notes:* The base levels for age and gender are both Unknown. Column (6) reports the marginal effects of the logit model at the mean or mode values of the explanatory variables using a 10% sample of the dataset. The raw estimates are reported in table 9 of the appendix. Standard errors are clustered by the hour of the day, the city, and the website and are heteroskedasticity-robust. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively.

translates to the revenue for the publisher (Yahoo). Table 4 columns (4) and (5) report the results of two alternative specifications. We find that having third-party cookies increases the highest bids, and consequently Yahoo's revenue, by a substantial 75%. Observe that this effect is more than double the effect in columns (1)-(3). The difference can be attributed to the fact that the bid regression does not account for entry; it only captures submitted bids.

An important feature of the online ad market is that bidders participate in auctions selectively. Recall that table 1 showed substantial variation in the number of bidders for different auctions, with a mean of 7.5 bidders and a standard deviation of 4.7. Therefore, we run regression 1 with the outcome variable $y_i$ as the number of bidders in each auction $i$. Table 5 columns (1) and (2) report the results of two alternative specifications. We find that, on average, an auction with third-party cookie identifiers induces about 5 more bidders (out of 33) to participate in the auction compared to an impression without. This is broadly consistent with some DSPs' strategies who simply only enter auctions with third-party cookie identifiers.

Lastly, we examine the effect of cookie availability on the entry decision of bidders in the auctions. In model 1, the outcome variable $y_i$ is $\text{Entry}_i$, a binary variable constructed for each auction-bidder pair that is equal to 1 if the bidder submitted a bid in the auction. Table 5 columns (3)-(5) report the results of three alternative specifications of such a linear probability model. We find that, on average, bidders are about 14% more likely to participate and submit a bid if the impression has third-party cookie identifiers. Assuming independence between the 33 DSPs, the increase in entry probability translates to an average increase in the number of bidders by $33 \times 0.14 \approx 5$, which is consistent with the estimation above. As a robustness check, we estimate a logit model for auction participation, $\text{Entry}_i = \mathbf{1}\{\beta_c \text{Cookie}_i + \mathbf{x}_i' \boldsymbol{\beta} + \alpha_i + \epsilon_i \geq 0\}$, where $\epsilon_i$ follows the standard logistic distribution. Table 5 column (6) reports the estimated marginal effects at the mean or mode values of the explanatory variables. The magnitude of the effect of cookies is comparable to those of the linear models. In the appendix, we report the point estimates of the logit model. The estimated coefficient on cookie availability translates into an odds ratio of $e^{1.19} = 3$; that is, the probability that a bidder participates in an auction for a cookie impression is three times higher than that for an auction for a cookieless impression.

# 4 Structural Estimation

## 4.1 Auction model and equilibrium characterization

Our empirical model is an independent private-value auction model with asymmetric bidders and binding reserve price (Krishna, 2009; Hubbard and Paarsch, 2014). We adopt the independent private-value assumption to reflect how users' impressions are horizontally differentiated; for instance, an impression from a male consumer is more valuable for male fashion brands but less valuable for female fashion brands. As our descriptive evidence (Figure 2a) shows that there is significant heterogeneity in bidding behavior across bidders, we allow for bidder heterogeneity in valuation distributions. Finally, our descriptive evidence shows that bidders enter only a fraction of auctions, and auctions in our data have reserve prices that vary across different websites.[16]

Consider an auction of an impression with a reserve price $r$ and $i = 1, 2, \cdots, N$ potential buyers. Suppose each bidder $i$ draws an independent private value $v_i$ from a distribution $F_i(v_i)$ that is differentiable with a density function $f_i(v_i)$. We suppress the dependency on auction characteristics now and will allow them to depend on both observed and unobserved auction characteristics later. Assume that all valuation distributions have a common, compact support $[0, \overline{v}]$. If no one bids above the reserve price, then the impression is not sold. Otherwise, the auction is resolved by the first-price mechanism where the bidder with the largest bid wins the auction and pays his bid $b_i$.

Suppose that all bidders are in equilibrium and use a bidding strategy $\beta_i(v_i)$ that is differentiable and monotone increasing in his valuation $v_i$. If the submitted bid $b_i$ is less than the reserve price $r$, he loses the auction and receives zero profits. Otherwise, the expected profit of bidder $i$ given his bid $b_i$ is

$$\pi_i(b_i) = (v_i - b_i) \prod_{j \neq i} F_j\big(\varphi_j(b_i)\big), \tag{2}$$

[16] An alternative approach is to introduce an entry stage where bidders endogenously decide if they would participate in an auction by comparing the expected profit to the bid preparation cost. This is not applicable in our context because the bid preparation cost in terms of computation and communication with the ad exchange is minimal compared to the reserve price.

where, for simplicity, $\varphi_j(b) = \beta_j^{-1}(b)$ denotes the inverse bid function.[17] The first-order condition of the profit maximization problem yields the following equilibrium condition:

$$\frac{1}{\varphi_i(b_i) - b_i} = \sum_{j \neq i} \frac{f_j\big(\varphi_j(b_i)\big)}{F_j\big(\varphi_j(b_i)\big)} \varphi_j'(b_i) \tag{3}$$

for $i = 1, 2, \cdots, N$. Equation (3) is a system of nonlinear ordinary differential equations in the inverse bid functions $\varphi_1, \cdots, \varphi_N$ that characterizes the Bayes-Nash equilibrium.[18]

In addition to the characterization above, we require two additional boundary conditions in order to solve the system. The lower boundary condition requires that any bidder who draws the reserve price $r$ would bid the reserve price. That is, for $i = 1, 2, \cdots, N$,

$$\varphi_i(r) = r. \tag{4}$$

The upper boundary condition requires that all bidders will submit the same bid $\bar{b}$ when they draw the highest valuation $\bar{v}$. In terms of the inverse bid function $\varphi_i$, we have for $i = 1, 2, \cdots, N$,

$$\varphi_i(\bar{b}) = \bar{v}. \tag{5}$$

## 4.2 Specifications and estimation procedure

In every auction, there is a constant number of $N = 33$ potential bidders who are both qualified and ready to submit a bid.[19] As explained earlier, we model auction interaction at the DSP level rather than the thousands of advertisers that the DSPs bid on behalf of. This assumption stays close to reality and also simplifies the computation. We maintain the assumption that auctions in our sample are independent of one another, abstracting away from potential dynamic considerations of the DSPs.

[17] Observe that the probability of winning is

$$\Pr(i \text{ wins}|b_i) = \prod_{j \neq i} \Pr\big(b_i > \beta_j(v_j)\big) = \prod_{j \neq i} \Pr\big(v_j < \beta_j^{-1}(b_i)\big) = \prod_{j \neq i} F_j\big(\beta_j^{-1}(b_i)\big) = \prod_{j \neq i} F_j\big(\varphi_j(b_i)\big).$$

[18] The existence and uniqueness of such an equilibrium are generally guaranteed under mild conditions. See Appendix G of Krishna (2009) for a discussion on the existence of such an equilibrium. See Lebrun (1999) for the conditions for the uniqueness of the equilibrium.

[19] These are the DSPs that have registered and established a business relationship with Yahoo's ad exchange, and all of them were actively participating in the ad exchange during the sample period.

Consider an auction $t$. Let $x_t$ denote the observed characteristics known to all DSPs (such as the user's cookie availability, opt-in/opt-out status, browser type, and other characteristics including gender and age.) We let the valuation distribution of each bidder $i$, $F_{it}(\cdot)$, depend on both observed and unobserved auction characteristics. Specifically, we assume that the log of valuation, $\log(v_{it})$, follows a normal distribution with mean $x_t'\gamma + \alpha_i + u_t$ and variance $\sigma_i^2$, where $u_t$ is the unobserved auction heterogeneity that is distributed normally with mean 0 and variance $\sigma_u^2$.[20]

There are two features in our specification that are integral to online ad auctions. First, we account for bidder heterogeneity by allowing asymmetric bidder valuation distributions through $\alpha_i$ and $\sigma_i$. We let each bidder $i$ fall into five distinct groups according to their type and size: large general-purpose, small general-purpose, rebroadcaster, large specialized, and small specialized (see section 2.1). With slight abuse of notation, the subscript $i$ of the parameters $\alpha_i$ and $\sigma_i$ denotes the group to which the bidder belongs. As explained, different types of DSPs cater to advertisers of different budgets, objectives, and targeted consumers, which may lead to an ex-ante difference in their valuations for impressions. The size of DSPs is a key dimension that captures their experience and expertise in matching advertisers with impressions.[21]

Second, the term $u_t$ captures the unobserved heterogeneity of the auction and is assumed to take a normal distribution with mean 0 and standard deviation $\sigma_u$. It essentially has a multiplicative effect on valuations as in Krasnokutskaya (2011). This allows for bids within an auction to be correlated conditional on observable characteristics, suggesting that there are hidden characteristics commonly observed by the DSPs but not the econometrician.

We adopt a nested estimation procedure in which the inner loop solves for the inverse bidding strategies $\varphi_{it}(b)$ using the equilibrium characterization (3) and the outer loop estimates the valuation parameters using maximum likelihood.

[20] The parametric approach follows the earlier empirical studies of auctions with high-dimensional auction characteristics (Athey, Levin, and Seira, 2011; Krasnokutskaya and Seim, 2011). A nonparametric approach is not ideal in our context because of the curse of dimensionality. In addition, the binding reserve price gives rise to the truncation of valuation and unobserved heterogeneity. The method allows us to parametrically recover the valuation distributions and the distribution of unobserved heterogeneity, components important for counterfactual simulations.

[21] A fully asymmetric version of the model with a distinct valuation distribution for every bidder is not desirable in our empirical setting. This alternative information structure would require that bidders know all their competitors' exact valuation distributions–a very strong assumption. It is more realistic to assume that bidders only know their competitors' group-specific parameters.

For the outer loop, the valuation parameters are estimated parametrically with maximum likelihood. Specifically, let $s_{it}$ be an indicator variable equal to 1 if bidder $i$ submits a bid in auction $t$ and 0 otherwise. The likelihood of bidder $i$'s observed bidding behavior $s_{it}$ and $b_{it}$ in auction $t$ given $u_t$ is

$$\mathcal{L}_{it}(s_{it}, b_{it}, x_t, u_t; \gamma, \alpha_i, \sigma_i) = \big(F_{it}(r_t)\big)^{1-s_{it}} \big(f_{it}(\varphi_{it}(b_{it}))\varphi'_{it}(b_{it})\big)^{s_{it}}, \tag{6}$$

where the first component $F_{it}(r_t)$ corresponds to the probability of non-participation due to valuation below the reserve price $r_t$, and the second component $f_{it}(\varphi_{it}(b_{it}))\varphi'_{it}(b_{it})$ is the density function of bids obtained by change of variable using the inverse bidding function $\varphi_{it}$. Then the joint likelihood of all bidders in auction $t$ is given by

$$\mathcal{L}_t(\boldsymbol{s}_t, \boldsymbol{b}_t, x_t; \gamma, \boldsymbol{\alpha}, \boldsymbol{\sigma}, \sigma_u) = \int \Big(\prod_{i=1}^{N} \mathcal{L}_{it}\Big)\phi(u_t)du_t, \tag{7}$$

where the unobserved heterogeneity is integrated out with respect to its normal density function $\phi(u_t)$ with mean 0 and variance $\sigma_u^2$. We estimate the structural parameters by maximizing the sum of $\log(\mathcal{L}_t)$ over the auctions $t$ in the data.

The inner loop solves for the inverse bidding functions $\varphi_{it}(b)$ for every auction. Because the equilibrium characterization (3) admits no closed-form solutions, we adopt a numerical approach to solve the system. Following Hubbard and Paarsch (2009); Hubbard, Kirkegaard, and Paarsch (2013); Hubbard and Paarsch (2014), we use Mathematical Programs with Equilibrium Constraints (MPEC) to solve for the equilibrium of the first-price auction model with asymmetric bidders. We approximate the inverse bidding functions $\varphi_{it}(b)$ as a linear combination of the first $K$ Chebyshev polynomials scaled to the interval $[r_t, \bar{b}_t]$:

$$\varphi_{it}(b) = \sum_{k=0}^{K} c_{k,it} T_k(b), \tag{8}$$

where $T_k(b)$ is the Chebyshev polynomial of degree $k$ scaled to the interval $[r_t, \bar{b}_t]$.

Then, we use the MPEC approach to discipline the Chebyshev coefficients $\boldsymbol{c}_t$ so that the first-order conditions defining the inverse equilibrium bid functions are approximately satisfied, subject to the boundary conditions (4) and (5). In addition, we impose rationality (players must bid less than their valuation) and monotonicity (bid functions are increasing) as shape constraints on the Chebyshev approximations (Hubbard and Paarsch, 2009; Hubbard, Kirkegaard, and Paarsch, 2013). Specifically, from equation (3), we define

$$G_{it}(b; \boldsymbol{c}_t, \bar{b}_t) = 1 - (\varphi_{it}(b) - b) \sum_{j \neq i} \frac{f_{jt}\big(\varphi_{jt}(b)\big)}{F_{jt}\big(\varphi_{jt}(b)\big)} \varphi'_{jt}(b), \tag{9}$$

where $\boldsymbol{c}_t$ are the coefficients of the linear combination of Chebyshev polynomials. Let $\mathcal{B}$ be the set of Chebyshev nodes in $[r_t, \bar{b}_t]$. For every auction $t$, we solve the following constrained optimization problem to obtain $\varphi_{it}$ and $\bar{b}_t$:[22]

$$\min_{\boldsymbol{c}_t, \bar{b}_t} \quad \sum_{i=1}^{N} \sum_{b \in \mathcal{B}} \left[G_{it}(b; \boldsymbol{c}_t, \bar{b}_t)\right]^2 \tag{10}$$

$$\text{s.t. } \varphi_{it}(r_t) = r_t,\ \varphi_{it}(\bar{b}_t) = \bar{v},\ \varphi_{it}(b) \geq b,\ \varphi'_{it}(b) \geq 0,\ \text{for } i = 1, \ldots, N \text{ and } b \in \mathcal{B}.$$

### 4.3 Estimation results

Table 6 reports the estimated parameters of the bid distributions. The estimates are of the expected sign and magnitude. In particular, the estimated coefficient of cookie availability is positive and significant, confirming that third-party cookies increase bidders' valuations. An impression with third-party cookies available raises the mean valuation by as much as 129 percent compared to an impression without cookies. Given bid shading in first-price auctions, the estimate is consistent with the reduced-form estimate of the effect on submitted bids. The estimated intercepts $\alpha_i$ and standard deviation $\sigma_i$ show substantial differences in the mean and variance parameters of valuation distribution across different DSP groups, where both small general-purpose and small specialized DSPs have low valuation distributions, reflecting their resource constraints. Lastly, the estimated variance of unobserved auction heterogeneity $\sigma_u$, while smaller in comparison to group-specific variances, remains statistically significant and positive. This suggests the presence of unobserved variations in auctions that are not accounted for by group-specific differences in the data.

We next present each bidder group's bidding pattern in response to third-party

[22] In the implementation, we use the first $K = 5$ order Chebyshev polynomials and 20 Chebyshev nodes for $\mathcal{B}$ to numerically approximate the inverse bid functions. These specifications are sufficiently flexible for approximations in our setting.

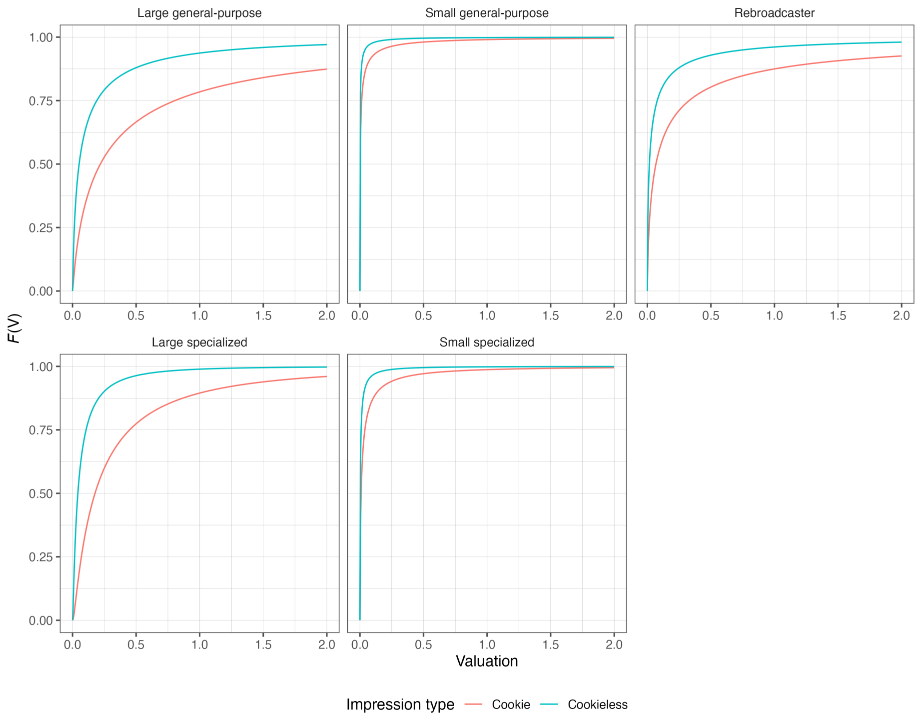
Table 6: Estimated parameters of valuation distributions

| Parameter | | Estimate |
|---|---|---|
| $\gamma$ | | |
| | Cookie | 1.288*** |
| | | (0.004) |
| | Opt-out | -0.047*** |
| | | (0.007) |
| | Gender female | -0.121*** |
| | | (0.009) |
| | Gender male | 0.003 |
| | | (0.009) |
| | Age 44 and below | 0.030*** |
| | | (0.010) |
| | Age 45 to 64 | -0.005 |
| | | (0.010) |
| | Age 65 and above | -0.037*** |
| | | (0.011) |
| | Interest segments | 0.063*** |
| | | (0.001) |
| | Months monetized | 0.004*** |
| | | (0.000) |
| | Total revenue (normalized) | 0.000 |
| | | (0.000) |
| | Days in database | 0.402*** |
| | | (0.025) |
| | Website fixed effects | Yes |
| | Browser fixed effects | Yes |
| $\alpha$ | | |
| | Large general-purpose | -2.972*** |
| | | (0.007) |
| | Small general-purpose | -7.490*** |
| | | (0.010) |
| | Rebroadcaster | -4.144*** |
| | | (0.007) |
| | Large specialized | -3.185*** |
| | | (0.007) |
| | mall specialized | -6.111*** |
| | | (0.008) |
| $\sigma$ | | |
| | Large general-purpose | 1.931*** |
| | | (0.002) |
| | Small general-purpose | 2.587*** |
| | | (0.004) |
| | Rebroadcaster | 2.342*** |
| | | (0.002) |
| | Large specialized | 1.383*** |
| | | (0.001) |
| | Small specialized | 2.089*** |
| | | (0.002) |
| $\sigma_u$ | | 0.626*** |
| | | (0.001) |

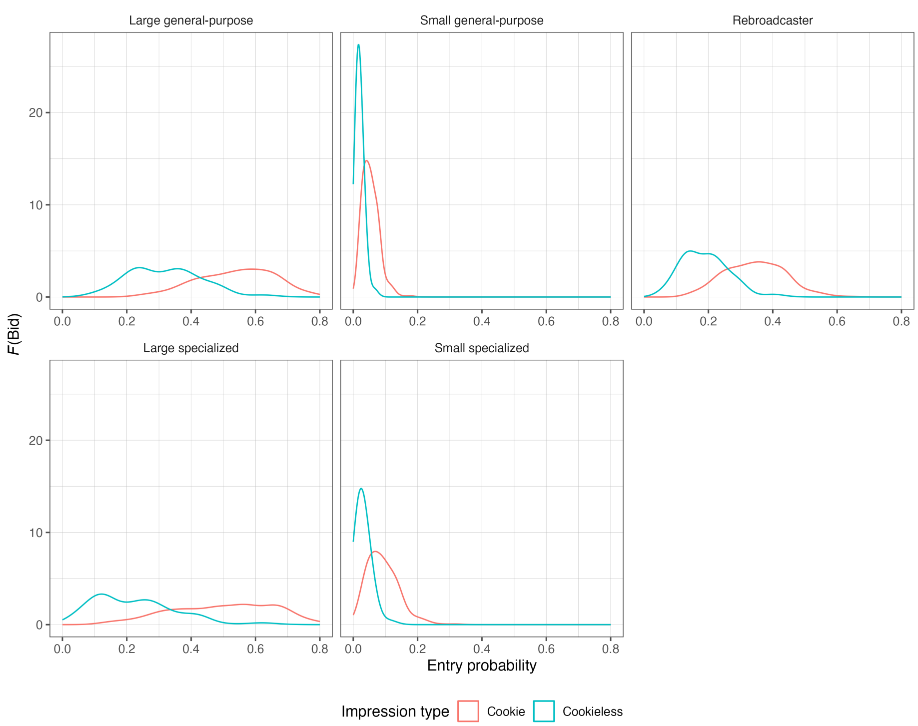
*Notes:* Parameter estimates of the log of valuation, $\log(v_{it})$, which follows a normal distribution with mean $x_t'\gamma + \alpha_i + u_t$ and variance $\sigma_i^2$, where $u_t$ is the unobserved auction heterogeneity that is distributed normally with mean 0 and variance $\sigma_u^2$. Estimates of website and browser fixed effects are not reported in the table. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively.

Figure 3: Cookie vs. cookieless: estimated bidders' behavior by DSP group

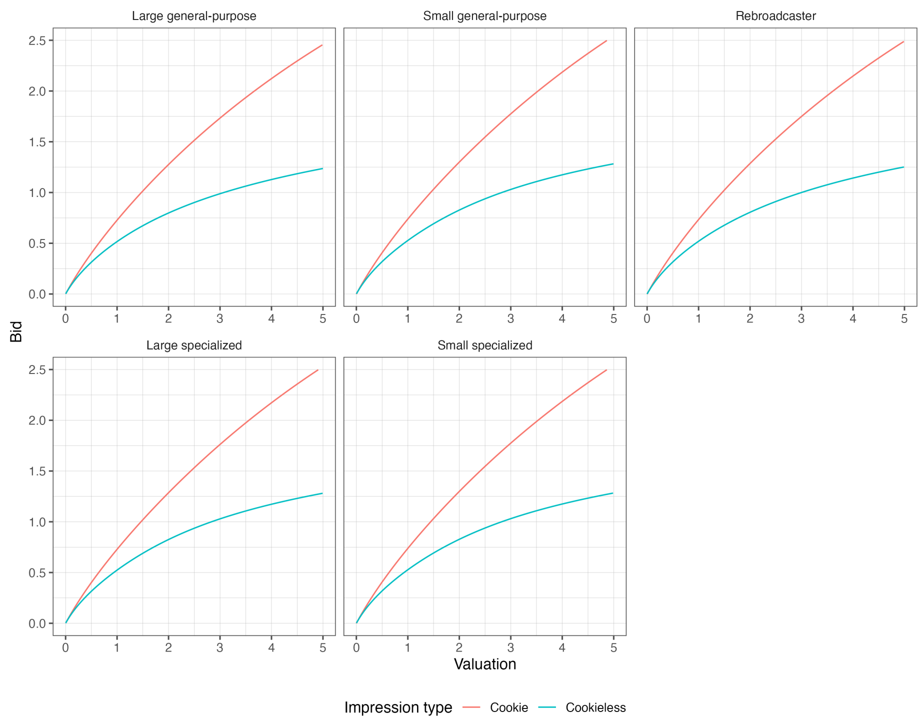
(a) CDFs of valuation distributions




(b) Density of entry probability



(c) Bid functions




*Notes:* Plots of bidder behavior by DSP groups with estimated parameters. The DSPs are grouped according to their purpose, specialty, and size. See section 2.1 for more details on the classification of DSPs. Subplot (a) shows the empirical density of entry probability, i.e. how likely the valuation exceeds the reserve price and the bidder submits a bid in an auction. (b) shows the cumulative distribution function $F_i$ of valuations at average auction characteristics. (c) shows the bid function $\beta_i$ at average auction characteristics. See figure 4 of the appendix for the bid functions of big and small general-purpose DSPs on the same plot.

cookie availability in terms of their valuation distribution, entry probability, and bidding strategy. Following the group classification outlined in section 2.1, we organize the plots by large general-purpose, small general-purpose, rebroadcaster, large specialized, and small specialized DSPs. Each figure shows the outcome variables for both impressions with and without cookies. For illustration, the valuation distribution and the bidding strategy are evaluated at the average values of the covariates.

Figure 3a shows the cumulative distribution functions (CDFs) of recovered valuation distributions. Figure 3b presents the empirical density of fitted entry probability, i.e. for all auctions in the data, the probability that the recovered valuation exceeds the reserve price. For either figure, we observe a clear dominance relationship of cookie impressions over cookieless ones across different DSP groups. Bidders are more likely to place a higher value and submit a bid in an auction with third-party cookies. There is also substantial heterogeneity across bidder groups. Notably, the effect of cookie availability is more pronounced for large DSPs.

Figure 3c presents the bidding strategy $\beta_i$ and shows that bidders bid more aggressively for cookie impressions.[23] Observe that, for the same valuation, bidders on average place bids on a cookie impression that are about twice as much as those on a cookieless impression. The difference can be attributed to the competition intensity between the two types of auctions, where fewer bidders would participate in auctions for cookieless impressions. Overall, our estimated structural results demonstrate that the difference between the average revenue from the two types of auctions comes from the difference in valuations, entry behavior, and bidding strategies.

# 5 Counterfactual Simulations

Using the structural estimates and the MPEC equilibrium solver, we simulate counterfactual scenarios to investigate the welfare redistribution of (1) Cookiepocalypse, the planned removal of third-party cookies from Chrome, and (2) Privacy Sandbox, the implementation of alternative tracking technologies. We show that the proposed changes have significant anti-competitive implications in terms of welfare distribution

[23]Given the relatively large number of bidders (33 in our data), average bidding strategies appear similar across bidder groups, though they do exhibit differences. See Figure 4 in the Appendix for a comparison. In particular, we find that smaller bidders adopt more aggressive bidding strategies to compete against larger bidders, who tend to have higher valuations.

among advertisers.

For each scenario, we consider three specifications. First, we simulate a status quo scenario as the benchmark (a less noisy version of the status quo in the data), to which we will compare the counterfactual scenarios. We will see that the results from the status quo are comparable to the summary statistics from the actual data. Second, we simulate a *symmetric* ban in which the cookie ban applies to all bidders, and none of them observe the cookie information. Third, we simulate an *asymmetric* ban by designating one bidder from the large general-purpose DSP group as the "Big Tech" DSP who retains access to Chrome users' third-party cookie information, but none of the other bidders observe any cookie information for Chrome users.

The asymmetric ban mirrors concerns raised by antitrust authorities, whereby certain DSPs may have alternative ways to gather and use ad-relevant information about users even when third-party cookies are blocked. For instance, DSPs affiliated with prominent publishers may have extensive user information through *first-party* cookies, which are typically enabled even by browsers that block third-party cookies by default. They may be able to leverage this rich first-party information about users for placing ads not only on their own websites but also on third-party websites, thereby obtaining a large information advantage over DSPs without similar capabilities.[24] A prominent example is Google, which possesses large amounts of first-party information on many internet users via its extensive web ecosystem encompassing the Google search engine, Gmail, YouTube, and more. This unique access to first-party information may allow Google to circumvent the effects of the Chrome third-party cookie ban and perhaps even to benefit from such a ban.[25]

To implement the counterfactual simulations, we draw a random sample of 10,000 auctions of impressions from the data. Importantly, this sample includes impressions from all browsers because we want to investigate the market-wide impact on the advertising market. For Chrome impressions (about 58% in the drawn pool), we manipulate their impression characteristics to emulate scenarios of the Cookiepocalypse.

[24] This alternative information collection can be implemented with "digital fingerprinting" methods that track users via IP addresses or device IDs, thus sidestepping cookies altogether. Peukert, Bechtold, Batikas, and Kretschmer (2022) observe that the drop in third-party cookie requests after the enactment of GDPR in the European Union was accompanied by a rise in first-party cookie requests.

[25] The anti-competitive implications of Google's plan on the ad supply chain have been closely scrutinized by government agencies. See Jeon (2020) for a more detailed discussion on the market power of Google in the online advertising markets.

For each Chrome auction and each specification, we draw valuations based on the true user characteristics for each bidder and, depending on the scenario and the bidder, mask any third-party cookie information for each user to simulate the effects of the ban. (That is, the cookie availability variable is set to zero. Other user characteristics associated with third-party cookies are set to either unknown or zero.) Given the counterfactual valuation distributions, we compute the bidding strategies by solving the system of ordinary differential equations (3) that characterizes the equilibrium.

## 5.1 Cookiepocalypse, blocking third-party cookies on Chrome

We first investigate the effect of Cookiepocalypse on submitted bids, the number of bidders, the winning bid (which translates into the publisher's revenue), and bidders' surplus. The results of this counterfactual simulation are presented in Table 7a. We find that the average bid falls from \$0.92 in the benchmark to \$0.56, representing a 39% decrease, and the number of bidders decreases from 7.4 to 4.8. Altogether, this results in about a halving (-54%) of the average publisher revenue from \$2.4 down to \$1.1. This estimate is consistent with several studies investigating the potential effect of removing third-party cookies including industrial studies.[26] On the buyer side, advertisers acquiring impressions through DSPs suffer a substantial 40% reduction in their surplus (the difference between valuation and bid), from an average of \$3.7 in the benchmark to \$2.2 in the first counterfactual.

We next investigate the distributional effect among bidders in terms of their winning frequency and surplus to highlight the unequal impact of the Cookiepocalypse. In Table 7b, we report the outcome variables in the asymmetric ban counterfactual scenario separately for the Big Tech DSP and the other five bidder groups. (Recall that the Big Tech DSP is drawn from the large general-purpose DSP group in the benchmark.) In terms of winning frequency, the Big Tech DSP wins twice as often (15.4%) in this scenario compared to the benchmark (8.3%), thanks to its informational advantage of having sole access to the behavioral information of Chrome users. Its total surplus also increases accordingly from \$31,800 in the status quo to \$48,900

[26] Several papers study the effect of restricting third-party cookies in online advertising and find a loss ranging from 4 percent to 66 percent (Beales and Eisenach, 2014; Marotta, Abhishek, and Acquisti, 2019; Johnson, Shriver, and Du, 2020). The industry estimate is closer to the upper end, where a study by Google finds that disabling third-party cookies results in an average loss of 52% (Ravichandran and Korula, 2019).

Table 7: Counterfactual simulation of Cookiepocalypse

(a) Simulated outcome

| | Status quo | Symmetric ban | Asymmetric ban |
|---|---|---|---|
| Bid | 0.917 | 0.558 | 0.588 |
| | (1.487) | (0.761) | (0.831) |
| No. bidders | 7.383 | 4.771 | 4.918 |
| | (3.965) | (2.879) | (2.865) |
| Publisher revenue | 2.433 | 1.101 | 1.208 |
| | (2.765) | (1.250) | (1.399) |
| Bidder surplus | 3.703 | 2.234 | 2.465 |
| | (5.604) | (4.367) | (4.629) |

(b) Welfare distribution

| | Status quo | Symmetric ban | Asymmetric ban |
|---|---|---|---|
| Winning frequency | | | |
| Big Tech DSP | - | - | 0.152 |
| Large general-purpose | 0.083 | 0.082 | 0.076 |
| Small general-purpose | 0.003 | 0.003 | 0.003 |
| Rebroadcaster | 0.048 | 0.048 | 0.045 |
| Large specialized | 0.028 | 0.026 | 0.024 |
| Small specialized | 0.004 | 0.004 | 0.003 |
| Surplus | | | |
| Big Tech DSP | - | - | 48,900 |
| Large general-purpose | 31,800 | 18,700 | 17,600 |
| Small general-purpose | 928 | 559 | 476 |
| Rebroadcaster | 20,200 | 12,800 | 12700 |
| Large specialized | 5,030 | 2,150 | 1920 |
| Small specialized | 875 | 420 | 369 |
| Full-information surplus | | | |
| Big Tech DSP | | - | 48,900 |
| Large general-purpose | | 31,000 | 29,300 |
| Small general-purpose | | 749 | 645 |
| Rebroadcaster | | 18,500 | 18,000 |
| Large specialized | | 4,890 | 4,260 |
| Small specialized | | 651 | 548 |

*Notes:* Simulated results are based on 10,000 auctions randomly drawn from the data. The Big Tech DSP is drawn from the large general-purpose DSP group. For Chrome impressions, auction characteristics are masked for all bidders in the symmetric ban scenario and are available exclusively to the Big Tech DSP in the asymmetric ban scenario. For each scenario, valuations are updated according to counterfactual characteristics, and outcomes are recomputed using the equilibrium characterization.

under the asymmetric ban, a 54% increase. At the same time, all the other bidders are impacted negatively by the ban, winning less frequently and receiving lower surpluses compared to the status quo and symmetric ban scenarios. Our results demonstrate that the third-party cookie ban leads to divergent experiences for the informational advantaged and disadvantaged bidders, where the former benefit from the ban at the cost of the latter.

To further decompose this redistributive effect, we also calculate the "full-information" surplus, that is, the difference between the valuation under cookie availability and the bid in the counterfactual scenario. The gap between the full-information and limited-information surpluses quantifies the loss in bidder welfare due to the inability to make precise matches when DSPs lose the ability to accurately evaluate and target users following the cookie ban. Comparing this difference in Table 7b, we see that welfare loss stems primarily from the diminished ability of affected DSP to effectively target users post-cookie ban. The primary factor responsible for the welfare redistribution is the inability of disadvantaged bidders to match with the most appropriate advertisements, rather than the Big Tech DSP monopolizing all the valuable impressions in the market.

## 5.2 Privacy Sandbox, alternative tracking technologies

In the second counterfactual, we replace third-party cookies with an alternative privacy-friendly tracking technology that allows bidders to acquire some behavioral information on the users, albeit without the precision and granularity of the cookie-generated information. Google has proposed a few alternative tracking technologies under its *Privacy Sandbox* initiative since 2021, shortly after its announcement of a third-party cookie ban. A prominent proposal is the *Topic API*.[27] With Topics, the browser will infer a handful of recognizable, interest-based "categories" for the user (such as automotive, literature, rock music, etc.) based on recent browsing history to help sites serve relevant ads. However, the specific sites the user has visited are no longer shared across the web like they might have been with third-party cook-

[27]See `https://privacysandbox.com/`. Several techniques have been or are being proposed, developed, and experimented with. Google initially experimented with the Federated Learning of Cohorts (FLoC) in 2021 and "received valuable feedback from regulators, privacy advocates, developers and industry. The new Topics API proposal addresses the same general use case as FLoC, but takes a different approach intended to address the feedback received for FLoC. Chrome intends to experiment with the Topics API and is no longer developing FLoC."

ies. In essence, this new method allows for tracking and targeting but in a more privacy-conscious and less precise manner than traditional third-party cookies.

In our implementation, because the exact alternative technology has not been finalized and we do not observe the user's interest categories, we follow the overarching principle of these proposed technologies that seek the best of the two worlds. On the one hand, users are afforded some degree of privacy; on the other hand, advertisers continue to observe user characteristics, albeit coarser ones. Specifically, we model this compromise between privacy and personalization by replacing Chrome users' behavioral characteristics with the average characteristics for each Yahoo website (e.g. Yahoo Mail, Yahoo Finance, Yahoo News, etc.). For example, the gender information of a Chrome user visiting Yahoo Finance is replaced by the website's proportions of male and female users. The Big Tech DSP, on the other hand, continues to observe Chrome users' exact characteristics.

The rightmost column of Table 8a contains the summary outcomes of the asymmetric ban under the Privacy Sandbox counterfactual. We find that the average bid has fallen from \$0.92 in the benchmark to \$0.82 in the counterfactual, and the number of bidders has decreased from 7.4 to 6.9. Altogether, this results in a 13% drop in the average revenue per auction from \$2.4 to \$2.1, and the bidder (advertiser) surplus drops by 8% from \$3.7 to \$3.4. In a word, the Privacy Sandbox still results in sizable welfare losses for both the publisher and the advertiser–an expected consequence given the coarser information in the market. On the other hand, the impact is a lot more cushioned compared to that of the Cookiepocalypse counterfactual under which the publisher and the advertiser bear a much heavier loss of 54% and 40%, respectively.

Table 8b presents the differentiated impact on DSP groups. Compared to the Cookiepocalypse counterfactual in table 7b, Privacy Sandbox alleviates the anticompetitive redistribution as well as the rising market concentration in favor of the Big Tech DSP. For the Big Tech bidder under the asymmetric ban, both its winning frequency (9.4%) and total surplus (\$36,000) increase compared to the benchmark (8.3% and \$31,800, respectively), representing a more than 10% gain, though the advantage is substantially attenuated compared to that under Cookiepocalypse. The disadvantaged bidders also experience noteworthy improvement compared to Cookiepocalypse. Their metrics under either symmetric or asymmetric ban are much closer to the status quo level: Under the asymmetric ban, for example, large general-

Table 8: Counterfactual simulation of Privacy Sandbox

(a) Simulated outcome

| | Status quo | Symmetric ban | Asymmetric ban |
|---|---|---|---|
| Bid | 0.917 | 0.815 | 0.824 |
| | (1.487) | (1.276) | (1.298) |
| No. bidders | 7.383 | 6.838 | 6.872 |
| | (3.965) | (3.636) | (3.636) |
| Publisher revenue | 2.433 | 2.061 | 2.099 |
| | (2.765) | (2.309) | (2.363) |
| Bidder surplus | 3.703 | 3.378 | 3.442 |
| | (5.604) | (5.380) | (5.475) |

(b) Welfare distribution

| | Status quo | Symmetric ban | Asymmetric ban |
|---|---|---|---|
| Winning frequency | | | |
| Big Tech DSP | - | - | 0.094 |
| Large general-purpose | 0.083 | 0.083 | 0.082 |
| Small general-purpose | 0.003 | 0.003 | 0.003 |
| Rebroadcaster | 0.048 | 0.048 | 0.048 |
| Large specialized | 0.028 | 0.028 | 0.028 |
| Small specialized | 0.004 | 0.004 | 0.003 |
| Surplus | | | |
| Big Tech DSP | - | - | 36,000 |
| Large general-purpose | 31,800 | 28,700 | 28,600 |
| Small general-purpose | 928 | 878 | 826 |
| Rebroadcaster | 20,200 | 18,700 | 18,800 |
| Large specialized | 5,030 | 4,230 | 3,940 |
| Small specialized | 875 | 745 | 715 |
| Full-information surplus | | | |
| Big Tech DSP | | - | 36,000 |
| Large general-purpose | | 32,600 | 32,500 |
| Small general-purpose | | 951 | 898 |
| Rebroadcaster | | 20,500 | 20,500 |
| Large specialized | | 5,400 | 5,030 |
| Small specialized | | 843 | 805 |

*Notes:* Simulated results are based on 10,000 auctions randomly drawn from the data. The Big Tech DSP is drawn from the large general-purpose DSP group. For Chrome impressions, auction characteristics are averaged at the website level for all bidders in the symmetric ban scenario. Exact characteristics are available exclusively to the Big Tech DSP in the asymmetric ban scenario. For each scenario, valuations are updated according to counterfactual characteristics, and outcomes are recomputed using the equilibrium characterization.

purpose DSPs enjoy a surplus of \$28,600, below the status quo level of \$31,800, but a substantial alleviation compared to \$17,600 under Cookiepocalypse. Although still a heavy 10% loss from the advertiser's perspective, this set of results suggests that advertising surplus and user privacy may not necessarily be at great odds. DSPs can rely on privacy-friendly technologies and coarser information to implement targeted ads without severely hurting their bottom lines. The anticompetitive redistribution effect, although much ameliorated compared to Cookiepocalypse, is still present and significant.

# 6 Conclusion

We study the impact of privacy protection on online advertising markets. As privacy concerns have mounted in recent years, internet browsers are increasingly moving away from third-party cookies, a widely-used tool to track online user behavior across the web and implement targeted ads. In this paper, we investigate the impact of a third-party cookie ban by analyzing online banner ad auctions using a detailed bid-level dataset from Yahoo. We find that auction participation, submitted bids, and revenue are higher when third-party cookies are available. This initial set of results demonstrates the pivotal role of third-party cookies in facilitating online advertising.

We next construct an empirical auction model, analytically characterize the equilibrium, and structurally recover valuation distributions from observed bids in the dataset. To evaluate the impact of the planned phasing-out of third-party cookies from Google Chrome, we perform counterfactual analyses based on the recovered structural parameters. Our results indicate that an outright ban–Cookiepocalypse–would reduce publisher revenue by 54% and advertiser surplus by 40%. However, the introduction of alternative, privacy-conscious tracking technologies under Google's Privacy Sandbox initiative, which delivers coarser user information to advertisers, would mitigate these losses.

We also quantify the redistribution of welfare resulting from the third-party cookie ban in which some large, informationally advantaged bidders could leverage their rich information over their competitors in online ad auctions. We find that these advantaged bidders stand to reap a larger surplus from the ban, whereas other bidders have no such recourse. Because of big tech firms' substantial presence in the ad supply chain and their abundant user information, the plan to eliminate third-party cookies raises legitimate antitrust concerns regarding competition and monopoly power in online advertising markets.

# References


Abraham, I., S. Athey, M. Babaioff, and M. D. Grubb (2020): "Peaches, lemons, and cookies: Designing auction markets with dispersed information," *Games and Economic Behavior*, 124, 454–477.

Acquisti, A., C. Taylor, and L. Wagman (2016): "The Economics of Privacy," *Journal of Economic Literature*, 54(2), 442–492.

Aridor, G., Y.-K. Che, and T. Salz (2020): "The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from GDPR," .

Athey, S., and P. A. Haile (2007): "Nonparametric approaches to auctions," *Handbook of econometrics*, 6, 3847–3965.

Athey, S., J. Levin, and E. Seira (2011): "Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions," *The Quarterly Journal of Economics*, 126(1), 207–257.

Barth, S., and M. D. T. de Jong (2017): "The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review," *Telematics and Informatics*, 34(7), 1038–1058.

Beales, H., and J. A. Eisenach (2014): "An Empirical Analysis of the Value of Information Sharing in the Market for Online Content," .

Brown, I. (2016): "The economics of privacy, data protection and surveillance," in *Handbook on the Economics of the Internet*, pp. 247–261. Edward Elgar Publishing.

Campbell, J., A. Goldfarb, and C. Tucker (2015): "Privacy Regulation and Market Structure," *Journal of Economics & Management Strategy*, 24(1), 47–73.

Celis, L. E., G. Lewis, M. Mobius, and H. Nazerzadeh (2014): "Buy-It-Now or Take-a-Chance: Price Discrimination Through Randomized Auctions," *Management Science*, 60(12), 2927–2948.

Decarolis, F., M. Goldmanis, and A. Penta (2020): "Marketing Agencies and Collusive Bidding in Online Ad Auctions," *Management Science*, 66(10), 4433–4454.

Decarolis, F., and G. Rovigatti (2021): "From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising," *American Economic Review*, 111, 3299–3327.

Ghose, A., and V. Todri-Adamopoulos (2016): "Toward a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior," *MIS Quarterly*, 40(4), 889–910.

Goldberg, S., G. Johnson, and S. Shriver (2019): "Regulating Privacy Online: An Economic Evaluation of the GDPR," .

Goldfarb, A. (2014): "What is Different About Online Advertising?," *Review of Industrial Organization*, 44(2), 115–129.

Goldfarb, A., and V. F. Que (2023): "The Economics of Digital Privacy," .

Goldfarb, A., and C. Tucker (2019): "Digital Economics," *Journal of Economic Literature*, 57(1), 3–43.

Haile, P. A., and Y. Kitamura (2019): "Unobserved heterogeneity in auctions," *Econometrics Journal*, 22(1), C1–C19.

Hong, H., and H. Paarsch (2006): *An introduction to the structural econometrics of auction data*. MIT Press.

Hu, Y., D. McAdams, and M. Shum (2013): "Identification of first-price auctions with non-separable unobserved heterogeneity," *Journal of Econometrics*, 174(2), 186–193.

Hubbard, T. P., R. Kirkegaard, and H. J. Paarsch (2013): "Using Economic Theory to Guide Numerical Analysis: Solving for Equilibria in Models of Asymmetric First-Price Auctions," *Computational Economics*, 42(2), 241–266.

Hubbard, T. P., and H. J. Paarsch (2009): "Investigating bid preferences at low-price, sealed-bid auctions with endogenous participation," *International Journal of Industrial Organization*, 27(1), 1–14.

——— (2014): "On the Numerical Solution of Equilibria in Auction Models with Asymmetries within the Private-Values Paradigm," in *Handbook of Computational Economics*, ed. by K. Schmedders, and K. L. Judd, vol. 3 of *Handbook of Computational Economics Vol. 3*, pp. 37–115. Elsevier.

Jeon, D.-S. (2020): "Market power and transparency in open display advertising – a case study," Discussion paper, EU Observatory on the Online Platform Economy.

Johnson, G. (2022): "Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond," .

Johnson, G., S. Shriver, and S. Goldberg (2022): "Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR," .

Johnson, G. A., S. K. Shriver, and S. Du (2020): "Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry?," *Marketing Science*, 39(1), 33–51.

Kesler, R., M. Kummer, and P. Schulte (2019): "Competition and Privacy in Online Markets: Evidence from the Mobile App Industry," .

Kong, Y. (2020): "Not knowing the competition: evidence and implications for auction design," *RAND Journal of Economics*, 51(3), 840–867.

Krasnokutskaya, E. (2011): "Identification and Estimation of Auction Models with Unobserved Heterogeneity," *The Review of Economic Studies*, 78(1), 293–327.

Krasnokutskaya, E., and K. Seim (2011): "Bid Preference Programs and Participation in Highway Procurement Auctions," *American Economic Review*, 101(6), 2653–2686.

Krishna, V. (2009): *Auction Theory*. Academic Press.

Lebrun, B. (1999): "First Price Auctions in the Asymmetric N Bidder Case," *International Economic Review*, 40(1), 125–142.

Levin, J., and P. Milgrom (2010): "Online Advertising: Heterogeneity and Conflation in Market Design," *American Economic Review*, 100(2), 603–607.

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

Marotta, V., V. Abhishek, and A. Acquisti (2019): "Online Tracking and Publishers' Revenues: An Empirical Analysis," .

Perrigne, I., and Q. Vuong (2019): "Econometrics of Auctions and Nonlinear Pricing," *Annual Review of Economics*, 11(1), 27–54.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2022): "Regulatory Spillovers and Data Governance: Evidence from the GDPR," *Marketing Science*, 41(4), 318–340.

Rafieian, O., and H. Yoganarasimhan (2021): "Targeting and Privacy in Mobile Advertising," *Marketing Science*, 40(2), 193–218.

Ravichandran, D., and N. Korula (2019): "Effect of disabling third-party cookies on publisher revenue," Discussion paper, Google.

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

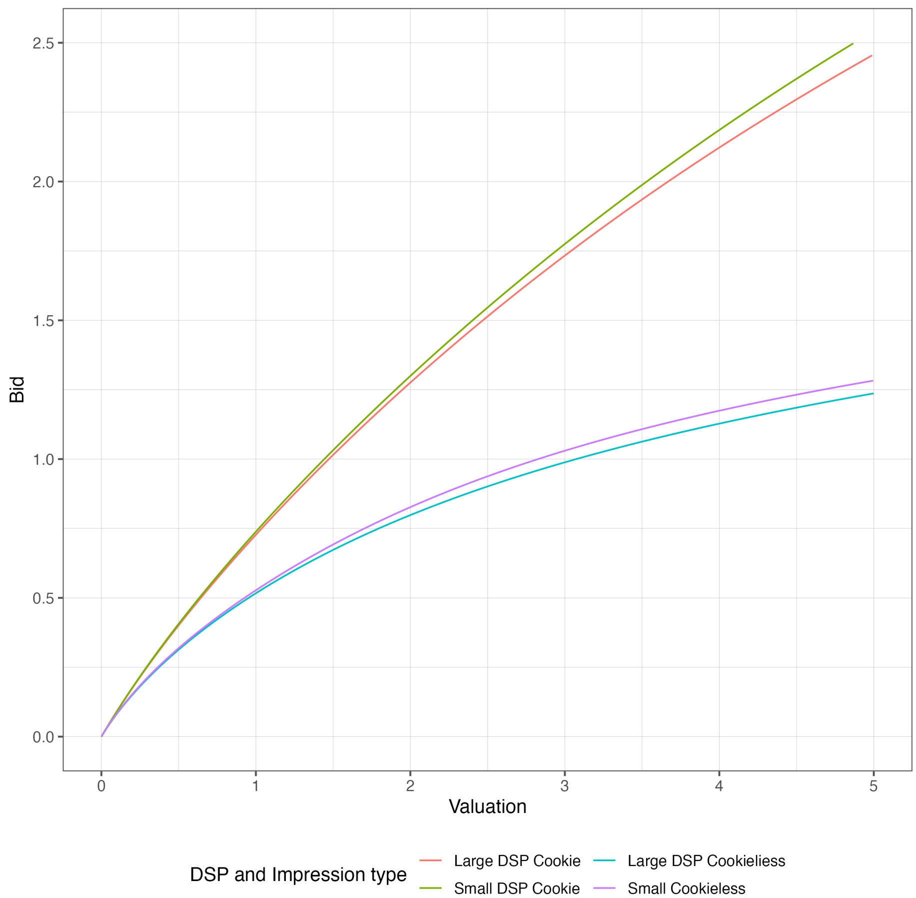
# A Additional Tables and Figures

Table 9: Regression results of logit model of entry decision

| Dependent Variable: | Entry |
|---|---|
| | (1) |
| Cookie | 1.191*** |
| | (0.053) |
| Opt-out | 0.084* |
| | (0.046) |
| Computer | -0.207*** |
| | (0.028) |
| Gender female | 0.164*** |
| | (0.010) |
| Gender male | 0.098*** |
| | (0.015) |
| Age 24 and below | -0.086*** |
| | (0.021) |
| Age 25 to 44 | -0.099*** |
| | (0.020) |
| Age 45 to 64 | -0.113*** |
| | (0.021) |
| Age 65 and above | -0.140*** |
| | (0.023) |
| Interest segments | 0.057*** |
| | (0.008) |
| Months monetized | 0.002*** |
| | (0.000) |
| Total revenue (normalized) | -0.030*** |
| | (0.002) |
| Days in database | -0.037*** |
| | (0.004) |
| *Fixed effects* | |
| Time (hour) | Yes |
| City | Yes |
| Website | Yes |
| Browser | Yes |
| Observations | 2,652,282 |

*Notes:* Estimation results of auction participation using logit model with 10% of the data. The base levels for age and gender are both Unknown. Standard errors are clustered by the hour of the day, the city, and the website and are heteroskedasticity-robust. ***, **, and * indicate statistical significance at the 1, 5, and 10% levels, respectively.

Figure 4: Bidding functions of large and small general-purpose DSPs




*Notes:* Bid functions of large and small general DSPs for cookie and cookieless impressions using estimated parameters at average auction characteristics.