Information Provision and Market Thickness in Auctions: Evidence from the Online Ad Industry *

Tulio Sousa†

January 22, 2026

Abstract

This paper studies how information provision shapes surplus and competition in market design by analyzing a tradeoff between matching and market thickness within the context of online advertising. Granular information allows advertisers to better identify high-value users, raising bids through better match, but enables them to enter fewer auctions, reducing competition. Using a dataset of 1.4 billion auctions, I estimate a model of advertiser entry and bidding with endogenous market thickness. Counterfactual simulations show that eliminating consumer-level information reduces ad spending by 65%, driven by an 89% decline from lower match, partially offset by a 24% increase from stronger competition.

Keywords: Market Design, Information Provision, Auctions, Data Privacy
JEL Codes: D47, D44, L15, L51, L86, M37

*I am extremely grateful to Allan Collard-Wexler, Carl Mela, James Roberts and Daniel Xu for their guidance and support. I thank Pellumb Reshidi, Emily Cuddy, Michael Dinerstein, Srinivas Tunuguntla, John Shin, Seohee Kim, Siddharth Prusty, Levin Zhu, Tirza Angerhofer, several contacts at the advertising industry, and participants of the Duke IO Lunch, Duke Public/IO Seminar, UNC IO Workshop, Triangle Micro Conference 2023, Triangle Applied Micro Conference 2025 and International Industrial Organization Conference 2025 for helpful discussions and conversations.

†Kelley School of Business, Indiana University: tudias@iu.edu

1 Introduction

A fundamental question in market design is how much information about the good being traded should be disclosed to market participants. From the designer's perspective, information provision entails a trade-off: more precise information improves the match between buyers and sellers, raising buyers' willingness to pay and, therefore, the revenue generated. At the same time, finer information can fragment the market by concentrating participation in a narrower set of transactions, reducing entry and market thickness. This reduction in participation softens competition, limiting the seller's ability to extract revenue despite higher match values. This tension between better matching and reduced competition arises across a wide range of markets, from financial exchanges and procurement auctions to online advertising and commodity trading. For instance, in commodity markets such as wheat, cocoa, and diamonds, sellers often aggregate goods of similar quality into standardized grades to attract more bidders and thicken the market. While theory highlights the central role of market thickness for efficiency, there is little empirical evidence on how this information trade-off operates in practice.

This paper studies this trade-off in one of the largest auction markets in the world: the market for online display advertising, where U.S. advertisers spent more than $150 billion with online display ads in 2023 alone. In this industry, advertisers compete daily in billions of auctions to purchase ad impressions targeted to individual consumers. The ability to target consumers is a major driver of advertising spending, which funds most of the free online products we use every day, such as browsers, social media networks and video platforms. Recently, large platforms and regulators have been restricting the ability to identify and reach specific users because of growing data privacy concerns. These policies reduce the precision with which advertisers can target consumers, effectively coarsening the information structure of the market. Such policies have implications that go beyond just consumer privacy protection. Because of this trade-off between revenue extraction and participation, restricting the amount of information available impacts the design of the online

ad industry itself and its entire economic viability.

To study this trade-off, this paper uses a dataset of 1.4 billion ad auctions from the Brazilian market to estimate a structural model of advertiser entry and bidding with endogenous market thickness. Brazil follows a data privacy regime that is closely aligned with those adopted in the U.S. and Europe, and the data offers comprehensive platform-level access to observe bids of hundreds of advertisers. In the model, advertisers first decide whether to enter the segment of personalized, targeted auctions or the segment of broad, non-targeted ones. Entering any segment requires paying a fixed participation cost, which can be interpreted as the managerial or attention cost of monitoring multiple advertising campaigns. Conditional on entry, advertisers participate in a second-price auction to show an ad to a consumer.

Within the model, advertisers optimally adjust their targeting strategies in response to counterfactual changes in the information structure of the market, such as those implied by policies that restrict access to consumer-level identifiers. This setting gives rise to two opposing forces that shape equilibrium winning bids. On one hand, losing the ability to identify and reach the most valuable audiences lowers advertisers' valuations and bids. This "matching effect" arises because, if advertisers can no longer recognize their best prospective consumers, they will not bid as aggressively as before. On the other hand, advertisers respond to the coarsening of information in the market by expanding their targeting strategies toward broader audiences, which increases the number of auctions in which they participate. This second force, the "competition effect", thickens the market and raises equilibrium bids. For example, an advertiser unable to target the small set of users who visited its website may instead target a broader pool of consumers with similar characteristics, thereby entering more auctions and intensifying competition.

The framework allows the study of a range of empirically relevant counterfactuals in the information structure of the market. These include policies that eliminate access to fine consumer identifiers altogether, interventions that increase the cost of exploiting such

information, and changes that reduce the precision of targeting signals without removing them entirely. Each of these interventions corresponds to policies or technological shifts observed in practice, and they operate through distinct equilibrium channels by reshaping advertisers' entry decisions and the thickness of competition. The counterfactual analysis leverages this structure to disentangle how different forms of information restriction affect matching, participation, and revenue.

The central exercise considers a complete ban of third-party cookies, which is one of the main tools that advertisers use to identify individual consumers online. This policy would disrupt many targeting strategies used by advertisers, but its largest impact would fall on retargeting, the practice of showing ads to consumers who previously visited the advertiser's website. I focus on retargeting for two reasons. First, it is among the most valuable targeting strategies: in the data, winning bids from retargeting campaigns are two orders of magnitude higher than those from non-retargeting campaigns. Although retargeting accounts for only 2.3% of the auctions, it represents 39.6% of total ad spending. Second, retargeting relies heavily on third-party cookies and cannot be easily substituted when these identifiers are removed. In this counterfactual, I find that auction revenue would decrease by roughly 65%, with an 89% decline due to weaker matching, partially offset by a 24% increase arising from stronger competition.

This paper makes three contributions to the literature. First, it provides empirical evidence on the trade-off between information provision and market thickness in one of the largest auction markets in the world, the market for online display advertising. This trade-off has been central to the theoretical literature on market and auction design, which highlights the tension between allocative efficiency and participation when information becomes more granular (Vulkan et al. 2013; Levin and Milgrom 2010; Rayo and Segal 2010; Bergemann and Pesendorfer 2007). Despite the prominence of this question, empirical evidence remains scarce. A recent exception is Budish et al. (2024), who examine how aggregating bids across a discrete time bandwidth in financial markets affects participation and market thickness.

In the context of online advertising, however, the relationship between information precision and market thickness has been studied almost exclusively in theory (Bergemann et al. 2021; Bergemann and Bonatti 2015). This paper is the first to quantify that trade-off empirically within this industry.

Second, this paper contributes to the recent literature on the economics of data privacy by quantifying how privacy regulations can unintentionally reshape competition in digital markets. Recent studies show that restricting access to user-level data affects firms' incentives, the allocation of advertising budgets, and ultimately the structure of online markets (Dubé et al. 2025, Deisenroth et al. 2024, Peukert et al. 2022; Aridor et al. 2025). These papers emphasize that policies designed to enhance privacy often generate spillovers, either by raising entry barriers for smaller firms, shifting market power toward integrated platforms, or reducing overall market efficiency. I extend this literature by developing the first structural framework that evaluates a data-privacy policy while accounting for its equilibrium effects on advertiser participation and market thickness. Methodologically, my approach is closest to Johnson (2013) and Alcobendas et al. (2023), who estimate models of advertiser bidding and targeting under different information structures. However, those studies treat the number of bidders as exogenous, whereas my framework endogenizes participation decisions, allowing the counterfactual analysis to capture how privacy policies alter both bidding behavior and market thickness.

Third, the paper contributes methodologically by identifying the parameters governing advertisers' entry decisions in retargeting. The empirical literature on auctions with an entry stage typically requires observing either direct measures of entry costs or the number of actual and potential bidders to recover the distribution of participation costs (Athey et al. 2011; Roberts and Sweeting 2010). The dataset used in this paper contains neither. I overcome this limitation by exploiting variation in the winning probabilities of advertisers' campaigns, obtained from an auction planner tool used by advertisers to predict advertiser campaign performance prior to launch. Through simulation exercises, I show that there is

a direct mapping between the parameters of the entry cost distribution and the unobserved number of active bidders in each auction. This mapping implies a relationship between the latent distribution of parameters and the observed winning probabilities, which allows the entry cost distribution to be identified without observing entry directly.

Related Literature This paper is mostly related to the broader market design literature, which studies how the rules and information environment of a market determine its ability to generate efficient and competitive outcomes. As emphasized by Vulkan et al. (2013), well-functioning markets must (i) provide *thickness*, attracting enough participants to interact; (ii) manage *congestion*, allowing participants to consider a sufficient set of options; and (iii) ensure *safety*, making participation straightforward and strategically secure. The mechanism I study in this paper speaks most directly to the first pillar. The trade-off between matching and competition in online advertising emerges because recently proposed data privacy policies, most notably the removal of third-party cookies, alter the information available to advertisers, thereby affecting equilibrium market thickness.

The logic behind the trade-off between better matching and reduced competition parallels the classic discussion of conflation introduced by Levin and Milgrom (2010). This literature shows that pooling goods with similar characteristics can create thicker markets and higher prices by mitigating the thinness that arises when buyers and sellers trade narrowly defined products. Commodity markets such as wheat, cocoa, and diamonds provide canonical examples: products are deliberately graded into coarser categories to balance information precision against market thickness. Budish et al. (2024) offer a recent empirical analogue in financial markets, where aggregating bids over discrete time intervals reduces the advantage of low-latency traders and increases effective competition. Both cases illustrate how institutional design can trade off informational precision for broader participation, exactly the tension present in this paper in the context of online ad auctions. By estimating how information granularity affects advertisers' entry and bidding decisions, this paper pro-

vides empirical evidence on one of the central design principles emphasized in this literature: that the optimal level of information disclosure depends on how participation responds to it.

Within the online advertising context, this work connects to the literature analyzing how information provision shapes auction outcomes and platform revenue. Models of thin markets in ad auctions, such as Chen and Stallaert (2014) and Hummel and Mcafee (2016), formalize the idea that better match can harm participation: as advertisers target more narrowly defined audiences, fewer bidders compete for each impression, potentially lowering revenue. These theoretical contributions highlight the existence of the trade-off between information precision and market thickness, but they remain qualitative and do not quantify how large these effects might be in equilibrium. Recent empirical studies provide evidence consistent with this prediction. Rafieian and Yoganarasimhan (2021) and Lu and Yang (2020) find that providing advertisers with more granular targeting options does not always increase platform revenue, while Ahmadi et al. (2024) show that non-targeted campaigns can sometimes yield higher advertiser surplus than narrowly targeted ones. Ada et al. (2022) demonstrate that when bidders receive more contextual information about impressions, their bids rise more in thicker markets. Although these papers document non-monotonic relationships between information precision and auction outcomes, they do not model how advertisers' participation responds endogenously to changes in information. By explicitly modeling advertisers' joint entry and bidding decisions, the framework presented in this paper moves beyond partial equilibrium analyses of bidding behavior and provides structural evidence on the magnitude of the trade-off between matching and competition emphasized in this strand of the literature.

This paper also relates to the general literature on targeted ads and the value of information leveraged by consumer tracking. Several papers either document or measure how valuable targeting is to advertisers (Wernerfelt et al. 2022, Goldfarb and Tucker 2011, Marotta et al. 2019, Lambrecht and Tucker 2013, Deisenroth et al. 2024). In particular, Wernerfelt et al. (2022) perform a large-scale experiment with thousands of advertisers on

Meta. They document that advertisers' cost of acquiring a new customer would increase by 37% if they were not able to use off-site data in their targeting settings.[1] In the dataset used in this paper, retargeting provides a substantially higher premium on bids. If one compares campaigns that rely solely on retargeting to campaigns that do not target consumer characteristics, the bids in the dataset are as much as two orders of magnitude higher. Also related to my paper but in the context of Apple's ATT policy, Deisenroth et al. (2024) use difference-in-differences techniques to estimate the causal effect of ATT on several outcome variables, including ad spend and firm entry and exit. The authors find that firms tend to exit the market and overall ad spend decreases, but advertising levels of surviving firms increase. This is in line with the conclusion of this paper that the entry rate of advertisers in non-retargeting campaigns increases following the deprecation of third-party cookies.

I also relate to the recent empirical literature on the impact of data privacy regulations on advertisers and platforms (Johnson 2013, Alcobendas et al. 2023, Johnson et al. 2020, Goldberg et al. 2024, Aridor et al. 2025, Sun et al. 2024, Korganbekova and Zuber 2023, Peukert et al. 2022). I find that the advertising spend in ad auctions would decrease by 65% following the removal of third-party cookies. This number is similar to those found in the literature. For example, in 2019 Google ran an internal experiment by disabling cookies for a randomized set of publishers and found that the ad auction revenue received by those publishers decreased by 52% (Ravichandran and Korula 2019). Similarly, Johnson et al. (2020) find that auctions for users who opt out of cookies close at a price 52% lower than for opt-in users; Alcobendas et al. (2023) find a reduction in publisher revenue of 54% if cookies are removed; and Johnson (2013) finds that advertiser surplus would decline by 45% under a regulation that bans user tracking. I extend this analysis by estimating advertisers' full entry strategies for retargeting and non-retargeting settings, which allows a decomposition of the equilibrium effect into two opposing forces: I find that the 65% decrease in auction revenue is a result of an 89% decrease in revenue due to lower match

[1] In their setting, off-site data is defined as data originating off of Meta's sites and applications.

value between advertisers and consumers, offset by a 24% increase in revenue due to fiercer competition. Conversely, Peukert et al. (2022) find that Google's market share increased after the implementation of the General Data Protection Regulation (GDPR).

Within the branch of literature that measures the impact of data privacy policies, the most relevant papers to my work are Johnson (2013) and Alcobendas et al. (2023). Both studies examine the impact of privacy regulations on the online advertising industry. Johnson (2013) estimates bidding strategies for advertisers that use consumer tracking technologies, with a focus on behavioral targeting. The author simulates counterfactuals where different fractions of the user population can no longer be tracked. Similar to my dataset, his includes bids from hundreds of different bidders placing ads on thousands of websites. However, the author does not observe the precise audience sizes and targeting settings of each advertiser. Consequently, his structural model includes audience sizes as parameters to be estimated, and the advertisers' valuation of impressions includes an unobserved heterogeneity term to account for the targeted user characteristics not observed by the econometrician. In contrast, I observe both audience sizes and advertisers' targeting specifications. This, paired with the data on winning probabilities, allows me to estimate an entry stage to the auction game, where advertisers first decide whether to enter the retargeting segment, and then submit bids. I estimate advertisers' entry strategy in equilibrium, which varies in the counterfactual scenario where targeting is no longer available.

In turn, Alcobendas et al. (2023) study the specific impact of Google's removal of third-party cookies. Similar to Johnson (2013), the authors also estimate bidding strategies of advertisers. I differ from Alcobendas et al. (2023) in two important ways. First, I focus on retargeting strategies, while those authors focus on demographic targeting such as age and gender. Second, advertiser entry in Alcobendas et al. (2023) is interpreted as submitting a bid above the reserve price. That is, the authors take advertisers' decision of what targeting strategy to employ as given in the counterfactual scenario with no cookies. In contrast, I endogenize advertisers' decision of which targeting strategy to use following the removal of

cookies and compute a new market participation equilibrium. This is important because advertisers are likely to change their entry rate in targeting strategies that remain viable in the absence of cookies. Such change happens because advertisers running broader, non-retargeting campaigns will no longer face high-valuation competitors engaging retargeting. Hence, their surplus in those non-retargeting auctions tends to increase, giving them greater ability to bear higher entry costs and more frequently enter non-retargeting campaigns.

The remainder of this paper is organized as follows. Section 2 provides some institutional background on the online ad auctions industry. Section 3 describes the datasets used in the paper and motivates the focus on retargeting strategies. Section 4 details the structural model used in the paper, and Section 5 details how it was estimated. Section 6 discusses what variation in the data identifies the parameters in the model. Section 7 presents the estimation results, while Section 8 shows the counterfactual scenarios under different information structure. Finally, Section 9 concludes.

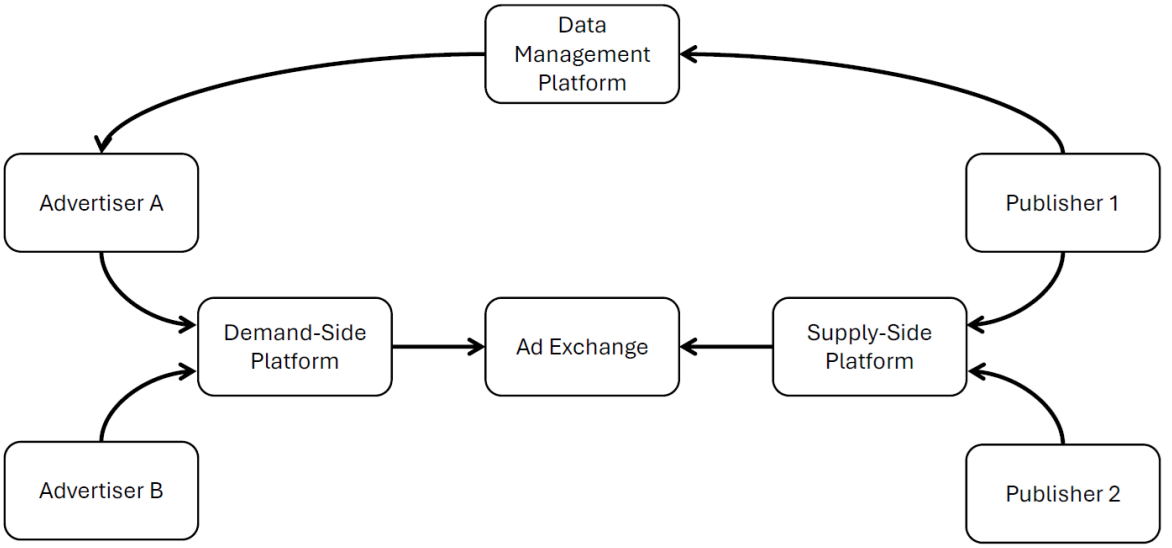
2 Institutional Background

This section provides a brief explanation of how the digital display ad industry works. A more detailed description of the industry can be found in Choi et al. (2020), while Online Appendix OA.1 provides an overview of the regulatory landscape involving the deprecation of third-party cookies.

2.1 The Online Advertisement Ecosystem

The digital display advertising ecosystem is a complex network of various parties involved in the automated buying and selling of ad space on websites and apps. Figure 1 shows a simplified diagram of this ecosystem. On the demand side are advertisers, companies looking to display their ads to targeted audiences. Advertisers set up campaigns, which consist of specific targeting configurations (e.g., location, interests, demographics and retargeting) and

Figure 1: Digital Display Advertising Ecosystem.



This figure represents a simplified version of the digital display advertising ecosystem. Advertisers contact Demand-Side Platforms with their targeting settings and bidding strategies. Data Management Platforms collect user data from publishers and sell it to advertisers. Publishers make their inventory available through Supply-Side Platforms, which list it at the Ad Exchange. The Ad Exchange runs the auctions.

a maximum bid amount they are willing to pay for reaching those targeted audiences. For instance, an advertiser may create a campaign targeting consumers in a particular region of the country with certain inferred interests, setting a maximum bid of, say, $1.50 per thousand auctions for that audience segment.

The other relevant players on the demand side of the market are the demand-side platforms. They work with advertisers, automating the bidding process across numerous auction opportunities based on the targeting settings and bidding strategies defined in the advertisers' campaigns. Typically, demand-side platforms offer advertisers data on consumer characteristics, but advertisers can also obtain these data from data management platforms, whose purpose in the ecosystem is to gather and sell information about consumers online.

On the supply side are publishers, the website and app owners seeking to monetize their ad inventory space by displaying ads. Publishers make their inventory available through supply-side platforms, which act as consolidated catalogs of potential ad placements. The ad exchanges connect the demand side (advertisers and demand-side platforms) with the supply side (publishers and supply-side platforms) and run the auctions to allocate ad impressions

to the highest bidders.

The process leading to an ad impression begins when a user visits a publisher's website. An auction opportunity arises when the page is loaded, and the supply-side platform sends the auction details, including available user data, to the ad exchange. The exchange queries the connected demand-side platforms, which respond with bids from advertisers targeting those specific user characteristics based on their campaign settings. The exchange then allocates the ad impression to the highest bidder. Traditionally, the standard auction format used in the allocation process has been the second-price auction. However, the industry is gradually shifting toward first-price auctions in an effort to provide greater transparency to advertisers and publishers.[2]

When an ad auction opportunity arises, a critical piece of information sent by publishers to the ad exchange is the user's cookie ID. Cookies are small text files websites store on users' devices to track the latter's online activities and preferences. There are two main types: first- and third-party cookies. First-party cookies are set directly by the visited website, allowing that site to recognize users across visits and enhance their experience through features such as saved settings, shopping carts, and automatic logins. Third-party cookies, however, are set by third-party entities with code integrated into the website. When a user visits a site with such code, the third-party server places a cookie associated with the third-party's domain on the user's device in response to a request from the user's browser. This cookie can be accessed and updated when the user visits other websites with the same third-party code, allowing these companies to track users' browsing activities across multiple sites and build comprehensive user profiles. Advertisers leverage these third-party cookies to create targeted audiences based on users' online behavior and interests.

[2] `https://clearcode.cc/blog/first-price-second-price-auction/` (Archive.org, accessed on 06/21/2024)

3 Data

A necessary condition for the tension between matching and competition effects to exist is that markets are not too thick. In a thick market, changes in competition following a shift in the information structure faced by advertisers would have little impact on final bids, and the total effect would be approximately equal to the matching effect. To demonstrate that online ad auctions involving retargeting are not exceedingly thick, this section provides descriptive evidence showing that advertisers' winning probabilities are significantly influenced by whether they are running retargeting campaigns, even after controlling for the size of the audience reached. This finding suggests that advertisers engaging in retargeting face less competition than those running non-retargeting campaigns, as they are targeting a more specific and limited audience. Consequently, the reduced competition in retargeting campaigns can be attributed to their inherent characteristic, which focus on a narrower set of consumers who have previously interacted with the advertiser's website. This evidence supports the existence of a meaningful interplay between matching and competition effects in the context of online ad auctions.

3.1 Data Sources

This paper uses two complementary datasets covering online display advertising auctions in Brazil. The first dataset is an auction-level dataset containing approximately 1.4 billion ad auctions that happened in the Brazilian market between November 2021 and March 2022. The data record only auctions in which an advertiser using the platform wins the impression. For these auctions, I observe the winning bid, the targeting settings used by the advertiser, the size of the targeted audience, and a rich set of impression-level controls. This dataset forms the basis for the descriptive evidence on bidding behavior and winning probabilities and subsequent model estimation.

The second dataset comes from an auction planner tool provided by the demand-

side platform. This tool reports predicted winning probabilities based on historical platform data as a function of advertisers' bids and targeting settings. The auction planner data is necessary because the auction-level data does not contain information on the number of bidders or losing bids, preventing direct measurement of competitive intensity. Section 6 uses variation in those winning probabilities to identify the distribution of entry costs faced by advertisers.

Additional details on data construction, variable definitions, and institutional features of the platform are provided in Online Appendix OA.2.

3.2 Descriptive Analysis

The auction-level data reveals a market that is both large in scale and highly heterogeneous. Even within a 1% random sample from a given week, the dataset contains almost 200 advertisers and thousands of campaigns and publishers, with substantial dispersion in activity across entities. As shown in Table 1, while many advertisers appear only infrequently, a small number participate in a very large number of auctions, reflecting the coexistence of small and large advertisers in the dataset.

Winning bids and payments are small on average but display a pronounced right

[3]To exemplify the distinction between publisher and URL, one can think of ESPN.com being a publisher, while ESPN.com/nfl and ESPN.com/nba being two different URLs within the same publisher.

Table 1: Frequency Distribution of Various IDs in the Dataset for a 1% Random Sample from Nov 20th to Nov 26th 2021

	Number of Unique IDs	Avg Freq	Std Dev Freq	Min	25%	50%	75%	Max
Advertisers	196	3,891	8,409	22	560	1,226	3,084	52,942
Campaigns	1,958	390	2,323	1	2	5	41	41,577
Publishers	2,571	182	3,263	1	1	2	10	126,152
URLs[3]	4,681	163	2,771	1	1	2	9	123,108

This table shows the distribution with which each unique player or object appears in the data. For example, the 1% random sample of the dataset contains 196 unique advertisers and, on average, each advertiser wins 3,891 auctions. The least frequent advertiser won 22 auctions, and the most frequent advertiser won 52,942 auctions.

Table 2: Descriptive Statistics of Numerical Variables in the Dataset for a 1% Random Sample from Nov 20th to Nov 26th 2021

	Mean	Std	Min	25%	50%	75%	Max
Buyer Bid (N$)	0.09	0.96	0.0003	0.01	0.02	0.03	254.14
Buyer Payment(N$)	0.06	0.50	0	0.01	0.02	0.03	133.68
Impression	1.00	0.03	0	1.00	1.00	1.00	1.00
View	0.76	0.42	0	1	1	1	1
Click ($\times$100)	0.13	4.00	0	0.00	0.00	0.00	100

This table shows summary statistics for all relevant numeric variables in the dataset. The total number of auctions in the 1% random subsample for one week of the data is 762,728. Monetary variables are multiplied by a constant and stated in terms of a numeraire currency per thousand auctions (N$).

Table 3: Representativeness of Each Targeting Setting in the Dataset for a 1% Random Sample from Nov 20th to Nov 26th 2021

Targeting Setting	% Campaigns	% Auctions	% Value Spent
Retargeting	81.8	2.3	39.6
Geography	6.7	22.9	10.2
Device Type	68.1	99.2	73.1
Inventory Type	99.2	96.9	73.7
Operational System	1.4	1.8	0.6

This table shows how representative each targeting setting is in the dataset. Each row represents a dummy for whether campaigns contains the respective targeting setting. For example, 81.8% of all campaigns involve retargeting. They represent 2.3% of the 762,728 auctions and account for 39.6% of the total value spent in the auctions.

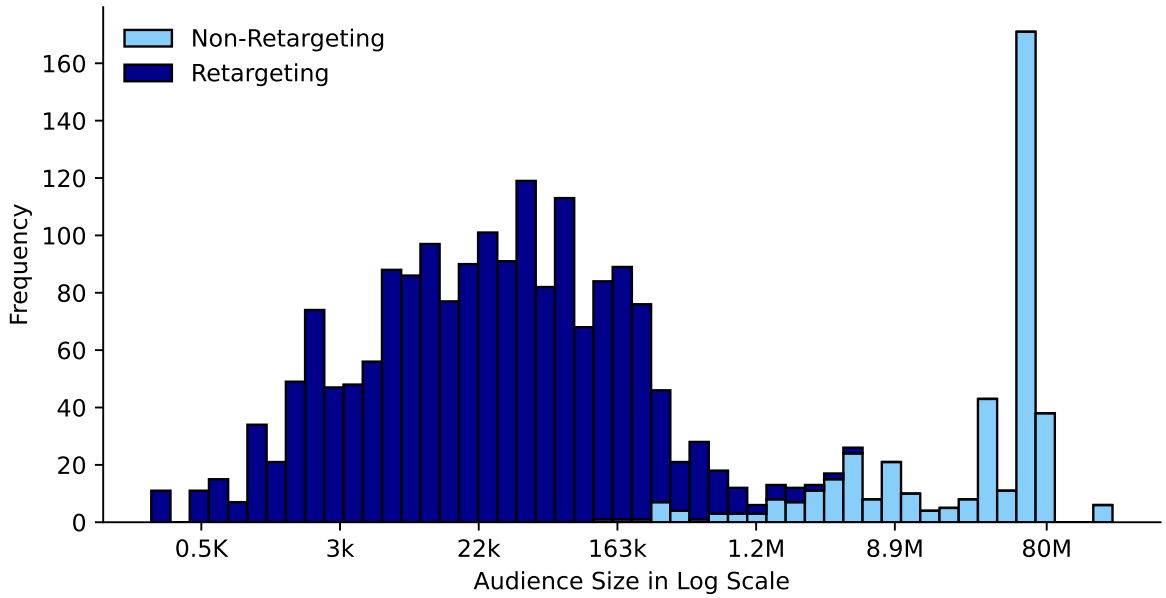
tail, as shown in Table 2. This indicates a large variation in advertisers' willingness to pay across impressions. Nearly all auctions result in an impression being shown, whereas clicks are relatively rare, consistent with standard features of display advertising markets.

A central dimension along which advertisers differ is their use of targeting. Table 3 shows that retargeting is used by a large share of campaigns (81.8%) but accounts for only a small fraction of observed auctions (2.3%). Despite this limited scope, retargeting represents a disproportionate share of total advertising spending (39.6%), indicating that advertisers place particularly high value on reaching consumers who have previously interacted with their products. Consistent with this, the average bid in retargeting campaigns is two orders of magnitude higher than in campaigns that do not use any targeting. These patterns suggest

that retargeting combines high valuations with very narrow reach, operating over much smaller consumer sets than broad, non-targeted campaigns. This combination of high value and limited scale motivates the paper's focus on retargeting and suggests that retargeting campaigns interact with a smaller set of competitors in any given auction environment.

These differences in scope are directly reflected in the distribution of campaign audience sizes. Figure 2 shows that audience sizes span several orders of magnitude, with retargeting campaigns concentrated among the smallest audiences and non-retargeting campaigns targeting a much larger number of consumers. Because audience size is related to the set of advertisers competing for a given impression, this dispersion implies systematic differences in market thickness across the two targeting strategies.

Figure 2: Distribution of Audience Size in Log Scale



This figure shows the distribution of the audience each advertiser reaches with their campaigns. Audience size is displayed on the x-axis in log scale, and the frequency of each bin is displayed on the y-xis. Bins in dark blue represent retargeting campaigns, while bins in light blue represent non-retargeting campaigns.

Two additional pieces of evidence connect these differences in scope to bidding behavior and competitive intensity. First, bid regressions using the auction-level dataset, on Table 4, show that retargeting is the targeting strategy most strongly associated with higher bids, even after progressively introducing rich fixed effects. The regression coefficients

from Table 4 reveal that, in the most comprehensive specification, the predicted bid for a retargeting campaign on mobile devices that reaches 100K consumers is N$5.92,[4] whereas that for a non-retargeting campaign on mobile devices that reaches 80M consumers is N$0.26.

Table 4: Regression on the Winning Bids in the Observed Dataset

Dependent Variable Log(Bid)	(1)	(2)	(3)
Intercept	-4.159***		
	(0.867)		
Retargeting	3.570***	4.335***	2.380***
	(0.265)	(0.421)	(0.332)
Geography	-0.062	-0.053	0.293**
	(0.174)	(0.161)	(0.096)
Device Type	0.577**	0.704***	0.646***
	(0.183)	(0.205)	(0.159)
Inventory Type	0.520	-0.230	-1.491***
	(0.294)	(0.412)	(0.398)
Operational System	-0.001	0.074	0.017
	(0.165)	(0.147)	(0.247)
Log(Audience Size)	-0.043	-0.050	-0.108**
	(0.046)	(0.031)	(0.036)
Fixed effects:			
Device Type		✓	✓
Inventory Type		✓	✓
Operational System		✓	✓
Advertiser			✓
Publisher			✓
R2	0.481	0.523	0.839
Cluster Level: Campaign			
Number of Auctions: 684,318			

This table shows results of three different regressions on the logarithm of the winning bids observed in the 1% random subsample of the dataset. The first five control variables represent dummies for whether the auction is part of a campaign that uses a specific targeting setting. The variable Log(Audience Size) controls for the number of potential consumers reached by the campaign. I gradually include auction-level fixed effects for each type of device, inventory and operational system, as well as fixed effects for advertisers and publishers. Standard errors are clustered at the campaign level. Significance levels: $* \ p < 0.05, ** \ p < 0.01, *** \ p < 0.001$.

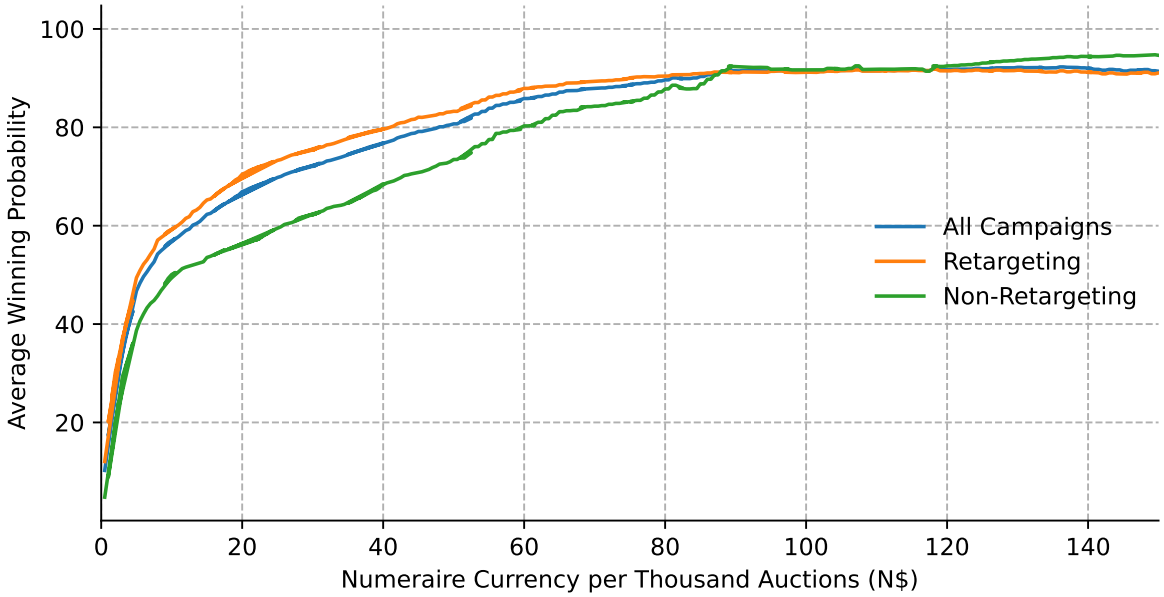
These findings suggest that the heterogeneity in bids observed in the dataset cannot be solely attributed to a large dispersion of valuations across different bidders. When

[4] I multiply all monetary values in the dataset by a constant for confidentiality reasons, and state them in terms of a fictitious numeraire currency (N$).

an advertiser chooses to employ retargeting, it is effectively attempting to buy a better match with consumers, which is embodied in the signal that the consumer visited the advertiser's website and is aware of its products. This provides the microfoundation used to specify bidders' valuations in Section 4.

Second, competitive conditions inferred from the auction planner vary sharply across targeting environments. Figure 3 plots average winning probabilities as a function of bids across campaign types.

Figure 3: Average Winning Probabilities Across Different Campaigns



This figure shows the average winning probability across different campaigns as a function of advertisers' bid. This data comes from the auction planner, a tool provided by the Demand-Side Platform to help advertisers plan their campaigns. On the x-axis I display the bids in a numeraire currency per thousand auctions (N$), and on the y-axis I display the corresponding average winning probability. The blue line is the average across all campaigns, the orange line is the average only for retargeting campaigns, and the green line is the average only for non-retargeting campaigns.

For most of the left portion of the average CDFs, where bids are close to those observed in the data, winning probabilities for retargeting campaigns are above those for non-retargeting campaigns. This means that a given bid ensures, on average, a higher winning probability for retargeting than for non-retargeting campaigns. In my model, this is rationalized by fiercer competition faced in non-retargeting campaigns. Regression evidence from the planner dataset, on Table 5, reinforces this conclusion: retargeting is associated

Table 5: Regression on the Winning Probability in the Auction Planner Dataset

Dependent Variable Winning Probability	(1)	(2)	(3)	(4)
Intercept	0.102***	1.454***	0.796***	
	(0.026)	(0.048)	(0.121)	
Retargeting	0.582***		0.318***	0.236***
	(0.019)		(0.046)	(0.039)
Geography	0.028	-0.122**	-0.028	-0.132***
	(0.020)	(0.037)	(0.029)	(0.033)
Device Type	-0.091**	-0.113***	-0.089**	-0.050*
	(0.028)	(0.029)	(0.028)	(0.020)
Inventory Type	0.003	0.029	-0.009	-0.173**
	(0.012)	(0.022)	(0.012)	(0.064)
Operational System	0.004	-0.188***	-0.078***	-0.046
	(0.010)	(0.028)	(0.023)	(0.031)
Log(Audience Size)		-0.074***	-0.039***	-0.054***
		(0.003)	(0.007)	(0.005)
Fixed Effects:				
Advertiser				✓
R2	0.587	0.576	0.606	0.758
Number of Campaigns	1,807	1,739	1,739	1,739
Average Winning Probability (Retargeting): 62.8%				
Average Winning Probability (Non-Retargeting): 2.1%				
Cluster Level: Advertiser				

This table shows results of four different regressions on the campaign's winning probability, using campaign-level data collected from the auction planner. The first five control variables represent dummies for whether the campaign uses a specific targeting setting. The variable Log(Audience Size) controls for the number of potential consumers reached by the campaign. In the last column the regression includes fixed effects for advertisers. Standard errors are clustered at the advertiser level. Significance levels: $* \ p < 0.05, * * \ p < 0.01, * * * \ p < 0.001$.

with significantly higher winning probabilities even after controlling for audience size and advertiser fixed effects, and smaller audiences are associated with higher winning probabilities. Taken together, this evidence indicates that retargeting operates in relatively thin markets, where changes in participation can materially affect equilibrium competition.

These descriptive patterns establish that retargeting combines high valuations with limited participation, creating an environment in which changes in information precision can affect both matching and competition. The next section formalizes these mechanisms in a model that endogenizes advertisers' participation decisions across targeting segments.

4 Model

This section presents the structural model employed in the paper. It builds on the literature on auctions with an entry stage (Samuelson 1985, Levin and Smith 1994, Li and Zheng 2009, Athey et al. 2011, Roberts and Sweeting 2016). The model captures the two-stage decision process faced by advertisers in the context of personalized retargeted ad auctions. In the first stage, advertisers must decide whether to enter the segment of auctions for personalized ads, incurring a fixed managerial (or attention) cost to do so. This cost represents the resources required to monitor the performance of various advertising campaigns. In the second stage, those who choose to enter participate by submitting bids in second-price auctions with no reserve price.

The focus of the model is on advertisers' decision to either engage in retargeting or forgo targeting altogether. Although this simplification excludes other targeting strategies, such as geographic targeting, it preserves the core economics of the entry decision while reducing computational complexity. By concentrating on the fundamental choice between retargeting and no targeting, the model captures the essential tradeoffs faced by advertisers in the context of targeted online ads.

4.1 Advertiser Valuations

There are N potential bidders indexed by i and J potential consumers indexed by j. Advertisers compete in second-price auctions.[5] Assume that the valuation of bidder i for showing an ad to potential consumer j is:

[5]This is an assumption that is widely used in the quantitative marketing literature (Balseiro and Gur 2019, Balseiro et al. 2015, Choi and Mela 2019, Choi and Sayedi 2019), although the industry has been moving to a first-price auction format (Despotakis et al. 2021, Zeithammer 2019, Alcobendas and Zeithammer 2021). The dataset used in this paper contains a mix of first and second price auctions. Importantly, it is possible to distinguish between these formats on an auction-level basis, which allows me to only use second-price auctions in any empirical exercise.

$$v_{ij} = \beta_r \mathbb{1}_{(j \in ret_i)} + \epsilon_{ij}, \quad \epsilon_{ij} \sim LogNormal(\mu_\epsilon, \sigma^2_{\epsilon,s}) \quad \text{iid} \quad \forall i, j, s \tag{1}$$

That is, valuations are composed of an idiosyncratic shock ϵ_{ij} and a retargeting premium β_r if potential consumer j is in the retargeting audience of advertiser i.

The auction happens in an independent private values (IPV) environment: the idiosyncratic valuation shocks ϵ_{ij} are independently distributed and are private information at the time of bidding, while their distribution is common knowledge. Conditional on the targeting setting and audience membership, advertisers do not receive additional signals about competitors' valuations. The IPV assumption is tested and validated with a field experiment in a companion paper (Sousa 2026).

The specification in Equation 1 is motivated by the results of Section 3. Those reduced-form results demonstrate that auctions won by retargeting campaigns close at substantially higher prices than those won by non-retargeting campaigns. This price differential suggests that the observed gap cannot be solely attributed to idiosyncratic valuation heterogeneity across bidders. Retargeting strategies provide advertisers with a highly informative signal, indicating that the targeted audience comprises the most valued consumers for the advertiser. By leveraging retargeting, advertisers can identify and focus their efforts on the subset of consumers who have previously expressed interest in their products or services, thereby justifying more competitive bidding behavior.

Idiosyncratic shocks for retargeting and non-retargeting campaigns are drawn from lognormal distributions with a common mean but potentially different variances. I do not allow for different means because retargeting valuations are already shifted by the retargeting premium β_r.

4.2 Auction Entry

Before bidding, advertisers must decide whether to enter the auction. Advertisers are allowed to do retargeting (indexed by r) and non-retargeting (indexed by nr).[6] Therefore, each advertiser can have 0, 1 or 2 campaigns running simultaneously. The decision to enter auctions under different targeting settings is made independently of each other, that is, advertisers' decisions are a binary outcome to enter each type of targeting setting. Participation in an auction with targeting setting $s \in \{r, nr\}$ requires paying a per-auction fixed cost κ_{ij}^s. This term represents an upfront managerial, or attention, resource required to monitor the performance of the auction. I assume that

$$\kappa_{ij}^s \sim LogNormal(\mu_\kappa, \sigma_\kappa^2) \quad \text{iid} \quad \forall i, j, s \tag{2}$$

This means that advertisers independently draw a value for the participation cost in each auction. Participation costs are private information, known before the entry decision, but their distribution is common knowledge.

4.3 Retargeting Audiences

Each advertiser i is endowed with an audience size $J_i \leq J$. This represents the number of consumers who have visited advertiser i's website in the past. When advertisers have access to individual-level identifiers, such as third-party cookies, they can create a retargeting campaign to reach the exact group of J_i potential consumers who previously visited their website. A potential consumer j can belong to audiences of multiple advertisers at the same time.[7] I assume that J_i and J are common knowledge to all bidders. Both J_i and J are

[6] In the dataset, advertisers run on average 10 simultaneous retargeting campaigns. I abstract from this empirical regularity, as allowing for a higher number of simultaneous campaigns does not change the core economical tradeoffs captured in the model

[7] I assume that this happens independently: I do not model any correlation between audiences of different advertisers. Therefore, when an advertiser i enters a non-retargeting campaign nr for a consumer j, the probability that consumer j belongs to advertiser i's audience J_i is $\frac{J_i}{J}$, whereas for a retargeting campaign this probability is 1.

observed in the dataset.

4.4 Optimal Bidding

The optimal strategy in the bidding stage of the is to bid one's expected valuation conditional on winning (Krishna 2009):

$$b_{ij}(v_{ij}) = \mathbb{E}[v_{ij}|i \text{ wins}] = \beta_r \mathbb{E}[\mathbb{1}_{(j \in ret_i)}|i \text{ wins}] + \epsilon_{ij} \tag{3}$$

In particular, I define the nonrandom component of the bid as:

$$\bar{b}_{is} \equiv \beta_r \mathbb{E}[\mathbb{1}_{(j \in ret_i)}|i \text{ wins}] \tag{4}$$

In this way, I can rewrite the optimal bid in Equation 3 as

$$b_{ij}(v_{ij}) = \bar{b}_{is} + \epsilon_{ij} \tag{5}$$

This notation will be important in Section 5, when I discuss estimation of the model. The nonrandom component of bids $\bar{b}_{is}$ is an equilibrium object, as it depends on the expectation that advertisers have about the chance of winning an auction for a consumer that belongs to their retargeting audience. This expectation depends on how other advertisers enter auctions, which is defined as an equilibrium in the first stage of the game.

4.5 Optimal Entry

In the entry stage, advertisers' optimal strategy is characterized by a cutoff threshold k^*_{is} for the entry costs. If the entry cost of advertiser i for participating an auction for potential consumer j with targeting setting s exceeds this threshold ($\kappa^s_{ij} > k^*_{is}$), than the advertiser does not enter the auction. This threshold is given by the value of the entry cost that makes advertiser i indifferent, in expectation, about entering an auction with targeting setting s.

Two points about this cutoff strategy are noteworthy. First, k^*_{is} is advertiser and setting specific.[8] This means that, for a given advertiser with audience size J_i, there will be two thresholds determining whether i enters each targeting setting: one for retargeting ($k^*_{i,r}$) and one for non-retargeting ($k^*_{i,nr}$). Second, the entry decision is made without knowledge of the idiosyncratic shocks ϵ_{ij}. This implies that there is no selection on idiosyncratic shocks for bidders that decide to participate in auctions. This timing closely matches how the online ad industry works. Advertisers must decide whether to create ad campaigns much before participating in auctions. At the time of this decision, they do not know their exact valuation for each of the millions of auctions in which they will participate.

In summary, the timing of this two-stage game is as follows. First, advertisers draw an entry cost κ^s_{ij} for both retargeting and non-retargeting campaigns. They compare this value to the equilibrium thresholds k^*_{is} and decide whether to enter the auctions. Next, entrants draw their value for the idiosyncratic shock ϵ_{ij} and submit bids. Finally, the ad is allocated to the highest bidder, and payments are realized.[9][10]

4.6 Equilibrium Definition

Let $\pi_i(b(v, \mathbf{k^*_s}), \mathbf{k^*_s})$ denote advertiser i's ex ante expected payoff of entering an auction using targeting setting s, given some bidding function $b(v, \mathbf{k^*_s})$ and a vector of equilibrium entry cost thresholds $\mathbf{k^*_s}$. An equilibrium in this two-stage game is composed of an entry rule $\sigma(\kappa, k^*_{is})$ and a bidding function $b(v, \mathbf{k^*_s})$ such that:

[8] This is a slight abuse of notation, as the entry strategy is not advertiser specific but rather audience-size specific. This means that if two advertisers have exactly the same audience size $J_a = J_b$, then their entry thresholds will be identical.

[9] The timing of this model closely resembles that in Levin and Smith (1994), where the entry decision is made without full knowledge of valuations. However, players in Levin and Smith (1994) use a mixed strategy, whereas in my model the strategy is deterministic conditional on the entry cost draw. This timing closely resembles model 2 in Li and Zheng (2009).

[10] One could also create another model where advertisers do not take two random draws, for ϵ and κ. Instead, entry would be contingent only on advertisers' idiosyncratic shocks, ϵ_{ij}, being higher than a threshold. I do not pursue this approach because this assumption requires advertisers to know beforehand the exact valuation of each impression. This is not a reasonable assumption in light of the industry context, as individual valuations are determined on an impression basis, at the time of the auction, long after the entry decision has been taken.

1. The entry rule $\sigma(\kappa, k_{is}^*) : (\kappa, k_{is}^*) \to \{0, 1\}$ is given by an equilibrium threshold k_{is}^* such that $\sigma(\kappa, k_{is}^*) = 1$ if, and only if, $\kappa \leq k_{is}^*,\ s \in \{r, nr\}$;

2. The equilibrium entry cost threshold k_{is}^* is given by the indifference condition

$$\pi_i(b(v, \mathbf{k}_\mathbf{s}^*), \mathbf{k}_\mathbf{s}^*) = k_{is}^* \tag{6}$$

   for a given bidding function $b(v, \mathbf{k}_\mathbf{s}^*)$

3. The equilibrium bidding function $b(v, \mathbf{k}_\mathbf{s}^*) : v, \mathbf{k}_\mathbf{s}^* \to \mathbb{R}^+$ is given by Equation 3

4. The entry rule $\sigma(\kappa, k_{is}^*)$ and the equilibrium entry threshold k_{is}^* are consistent with the ex ante expected payoffs generated by the bidding function $b(v, \mathbf{k}_\mathbf{s}^*)$

5. The bidding function $b(v, \mathbf{k}_\mathbf{s}^*)$ is consistent with the selection induced by the entry rule $\sigma(\kappa, k_{is}^*)$

5 Estimation

This section details the estimation procedure used to recover the parameters of interest from the structural model. The set of parameters estimated is $\Theta = \{\beta_r, \mu_\epsilon, \sigma_{\epsilon,r}^2, \sigma_{\epsilon,nr}^2, \mu_k, \sigma_k^2\}$, namely, the valuation premium for retargeting as well as the distribution of idiosyncratic shocks and entry costs. These parameters allow for the computation of the objects determining the equilibrium of the model: the optimal bidding policies from Equation 5 and the entry cost thresholds given by Equation 6.

Estimation is based on a subset of mobile display ad auctions that isolates the choice between retargeting and no targeting, and is restricted to mature campaigns participating in second-price auctions. This sample selection ensures a clean comparison across targeting settings and aligns the data with the assumptions of the model. Additional details on the construction of the estimation sample are provided in Online Appendix OA.3.

The model is estimated using Simulated Method of Moments with an Adaptive Nelder-Mead minimization algorithm in Python. Formally, let M be a $k \times 1$ vector of moments from the true dataset, and $\widetilde{M}(\Theta)$ be the analogous vector of moments from the simulated dataset, which depends on the vector of model parameters. Let Ω be a $k \times k$ positive-definite weight matrix. Then, the SMM estimator is defined as:

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \left[M - \widetilde{M}(\Theta)\right]^\intercal \Omega \left[M - \widetilde{M}(\Theta)\right] \tag{7}$$

This paper uses an identity matrix for Ω. Nonetheless, the estimates are consistent regardless of the choice of weights.

5.1 Moments Used in Estimation

The Simulated Method of Moments estimator uses two sets of moments. The first set consists of the average and standard deviation of the first- and second-highest bids for retargeting and non-retargeting campaigns. This set comprises eight moments. Table 6 shows the value of each moment in the selected data for estimation. Similar to the descriptive statistics of the full data, the average bids for retargeting campaigns are two orders of magnitude higher than those for non-retargeting campaigns. In particular, the average of the highest bid for retargeting is 100 times higher than that for non-retargeting, demonstrating how much advertisers value the informative signal embodied in retargeting audiences. Retargeting campaigns also show a substantially larger variance in bids, with a standard deviation of approximately N$1.98. In contrast, there is considerably less bid heterogeneity for non-retargeting campaigns, where the standard deviation of bids is approximately N$0.01.

The second set of moments used is the winning probabilities conditional on bidding, taken from the auction planner. Those probabilities are computed at the average bid of each campaign in the dataset. Each simulated bidder in the algorithm is assigned an OLS projection of their empirical winning probability based on whether the bidder is

Table 6: First Set of Moments from the Data Used in Estimation

Moment	Value (N$)
Average highest non-retargeting bid	0.012
Average second-highest non-retargeting bid	0.011
Average highest retargeting bid	1.133
Average second-highest retargeting bid	0.699
Std. of highest non-retargeting bid	0.013
Std. of second-highest non-retargeting bid	0.009
Std. of highest retargeting bid	1.989
Std. of second-highest retargeting bid	1.137

This table shows the first 8 moments used in estimation. Those moments are computed using 12,672,902 auctions from 265 campaigns in the data, and then matched by the Simulated Method of Moments estimator.

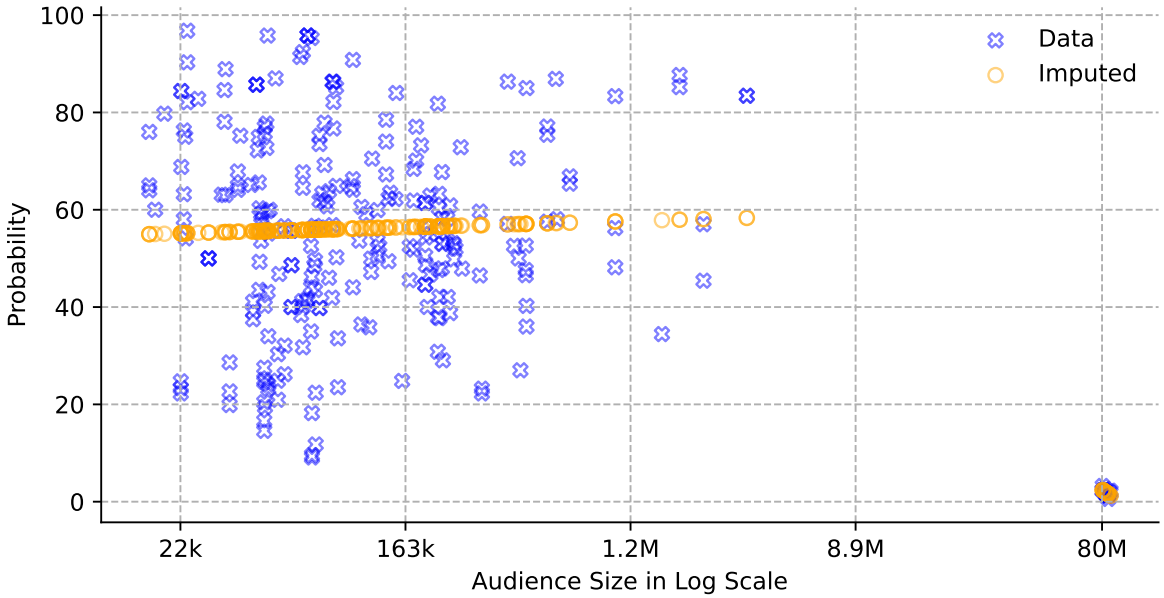
adopting retargeting and on the bidder's audience size. Figure 4 shows how the imputation is done. The cluster of campaigns with an audience size of 80M are those non-retargeting campaigns that reach all mobile device users. The other cluster, with an audience size of approximately 100K, are the retargeting campaigns. Each simulated bidder is imputed the predicted winning probability as a function of its audience size and whether it is adopting retargeting. Then, each simulated bidder is assigned the predicted winning probability from this projection.

5.2 Simulation Algorithm to Solve the Model

The equilibrium objects are the nonrandom component of bids $\bar{b}_{is}$ and the entry cost thresholds k^*_{is}. A nested fixed point algorithm is required to recover them.

The algorithm is initialized with a given set of parameters $\Theta = \{\beta_r, \mu_\epsilon, \sigma^2_{\epsilon,r}, \sigma^2_{\epsilon,nr}, \mu_k, \sigma^2_k\}$ and a number of potential bidders N. The current version of this paper uses $N = 10$ potential bidders, and Appendix C shows robustness checks relaxing this number. Then, for an initial guess of $\mathbf{k}^*_{\mathbf{s,0}}$, I simulate a large number of auctions. For each auctioned consumer j, bidders $i \in \{1, \cdots, N\}$ draw an entry cost κ^s_{ij} and compare this value to their respective (guessed) entry threshold $k^*_{is,0}$ to make the entry decision. Then, each entrant draws its idiosyncratic shock ϵ_{ij} and realizes whether consumer j belongs to its retargeting audience

Figure 4: Imputed Winning Probabilities as a Function of Audience Size



This figure shows the campaigns' winning probability as a function of audience size. On the x-axis, it displays the size of the audience reached by each campaign in log scale. On the y-axis, it shows the winning probability associated with each campaign. Each blue cross marks a campaign's winning probability and audience size, collected from the auction planner. Each orange circle denotes a campaign's predicted winning probability using a simple OLS regression.

J_i. This happens with probability $\frac{J_i}{J}$ for each bidder. Next, each player bids and payoffs are realized.

The procedure above is repeated a sufficiently large number of times to compute the equilibrium nonrandom component of bids $\bar{b}^0_{is} \equiv \beta_r \mathbb{E}^0[\mathbb{1}_{(j \in ret_i)} | i \text{ wins}]$ (Equation 4). I repeat this computation until convergence of the nonrandom component of bids is achieved. This configures the inner loop of my algorithm. Note that this quantity depends on the guessed vector of entry thresholds $\mathbf{k}^*_{\mathbf{s,0}}$. With $\bar{b}^0_{is}$, I then update the equilibrium entry cost thresholds by computing each bidder's ex ante expected payoff $\pi^0_i(b^0_{ij}, \mathbf{k}^{*,\mathbf{old}}_{\mathbf{s,0}}) = k^{*,new}_{is,0}$ (Equation 6). Finally, I repeat the updating of entry thresholds until it converges. This configures the outer loop of the algorithm, which recovers the equilibrium entry cost thresholds. A pseudo-code that implements the computation of equilibrium objects, detailing the procedure described above, is in Appendix B. More details about the simulation process are in Online Appendix OA.4.

6 Identifying Variation

This section discusses what variation in the data identifies the key structural parameters of the model: retargeting valuation premium (β_r), mean and variance of idiosyncratic shocks ($\mu_\epsilon, \sigma^2_{\epsilon,r}$ and $\sigma^2_{\epsilon,nr}$), and mean and variance of entry costs (μ_k and σ^2_k). These parameters are identified from the variation in bids and winning probabilities across different types of advertisers and impressions, combined with the equilibrium restrictions implied by the model.

6.1 Valuation Parameters

First, the retargeting premium β_r is identified from the difference in average bids between retargeting and non-retargeting campaigns. The model assumes that valuations include an additive term β_r for retargeted impressions. Therefore, the average difference in bids between these two types of targeting settings identifies the premium that advertisers place on retargeting.

Second, the mean and variances of idiosyncratic shocks, $\mu_\epsilon, \sigma^2_{\epsilon,r}$ and $\sigma^2_{\epsilon,nr}$, are identified from the mean and variance of bids within retargeting and non-retargeting campaigns, after controlling for the retargeting premium β_r. Intuitively, μ_ϵ determines the average level of bids, while $\sigma^2_{\epsilon,r}$ and $\sigma^2_{\epsilon,nr}$ govern the dispersion of bids around this mean. By matching the observed bid moments within targeting setting, the model can separate the contribution of the idiosyncratic shocks from the retargeting premium.

6.2 Entry Cost Parameters

The identification of entry cost parameters, μ_k and σ^2_k, requires more careful consideration. In auctions with selected entry, identification of entry costs typically relies on observing either a measure of entry costs or the number of actual and potential bidders in each auction (Roberts and Sweeting 2010, Athey et al. 2011, Roberts and Sweeting 2016). However, nei-

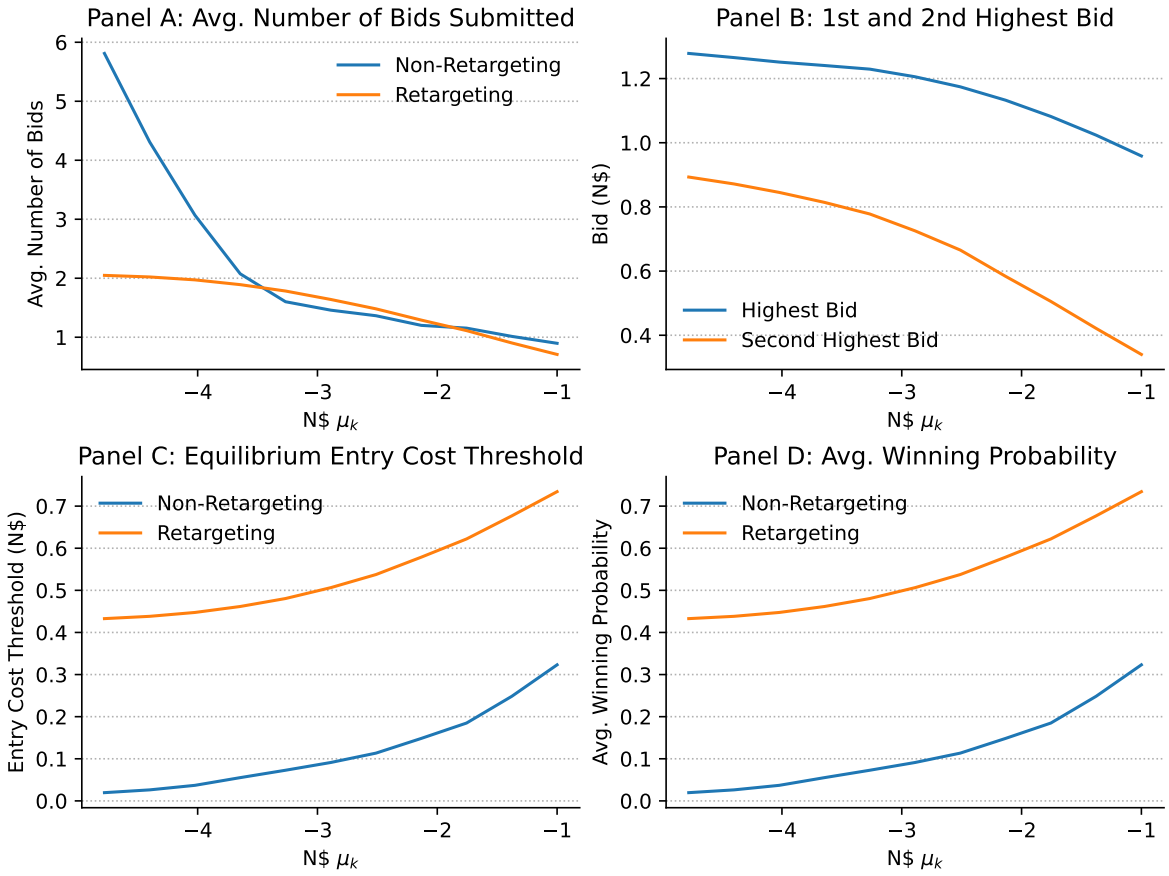
ther is observed in the dataset used in this paper. Instead, the winning probabilities of each campaign are available. Consequently, the identification of entry costs in this case involves linking variation in the unobserved numbers of bidders, driven by differing parameters for the entry cost distribution, to variations in observed winning probabilities.

The relationship between the entry cost parameters and key outcome variables of the model is illustrated using simulated datasets with the estimated parameters from Table 7. While the number of actual bidders is not observed in the data, they are observed in the simulated data. This allows tracing the effect of varying the mean and standard deviations parameters from the (unobserved) number of actual bidders to the (observed) winning probabilities.

Figures 5 and 6 show simulation results for four quantities in the model: number of actual bidders (Panel A); average highest and second-highest bids (Panel B); equilibrium entry cost thresholds (Panel C); and winning probability conditional on entry (Panel D). A higher mean parameter of the entry cost distribution has an unambiguously negative effect on the number of bids submitted. Advertisers running both retargeting and non-retargeting campaigns tend to participate less frequently in auctions, as their entry costs are generally increasing. With fewer participants, the average highest and second-highest bids in the auctions tend to decrease. This leads to two consequences for advertisers: first, the surplus in the auctions they win tend to increase, which in turn increases their equilibrium entry cost thresholds; second, their winning winning probabilities tend to increase, as they face lower competing bids. Therefore, higher mean parameters for the entry cost distribution exhibit a positive correlation with the observed winning probabilities in the dataset used in this paper.

The effect of increasing the parameters for the standard deviation is shown in Figure 6. Contrary to what is observed when one varies the mean parameter of the entry cost distribution, a higher variance parameter has an ambiguous effect on the number of bids submitted. This happens because, as one increases the variance of the entry cost

Figure 5: Simulations Results of Varying Mean Parameter of Entry Cost Distribution

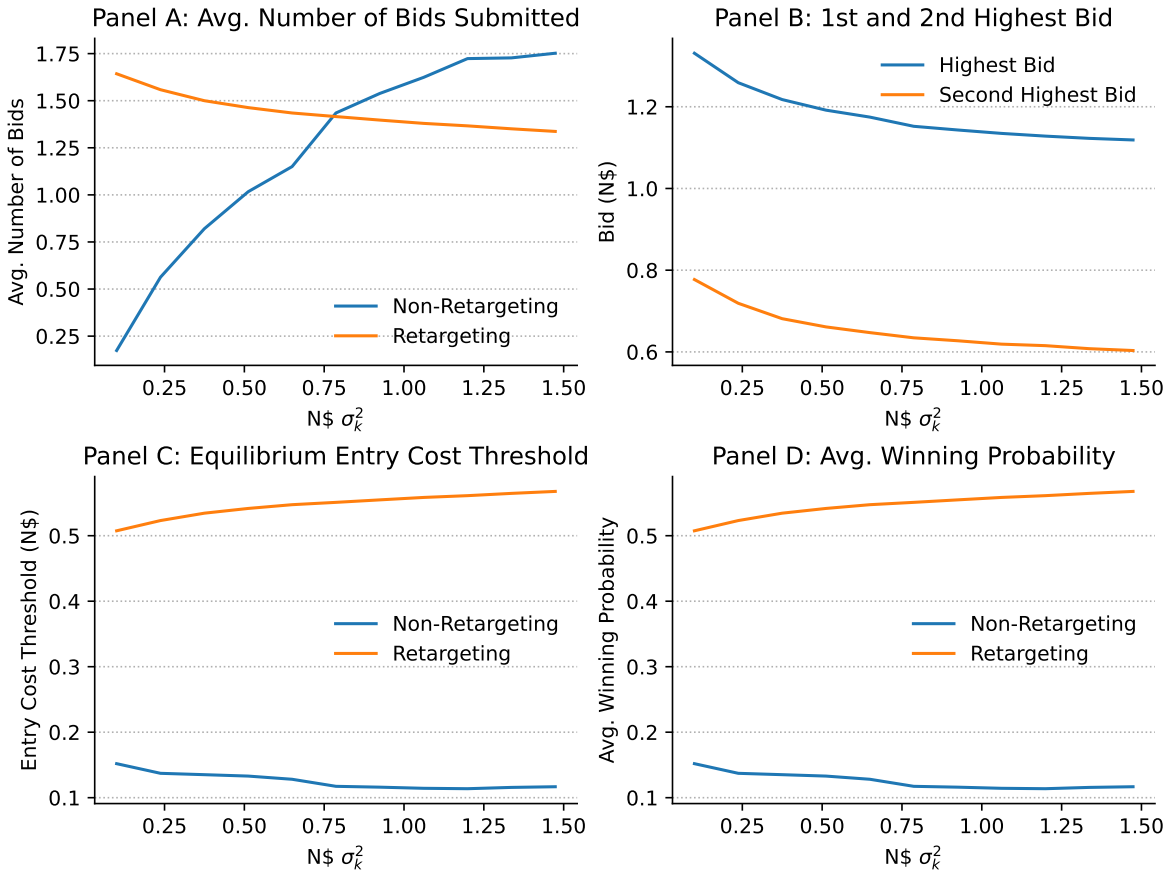


This figure shows simulation results that indicate the effect of varying the mean parameter of the entry cost distribution on four key quantities of the model: average number of submitted bids (Panel A); highest and second highest bid (Panel B); equilibrium entry cost threshold (Panel C); and average winning probability conditional on entry (Panel D). The mean parameter of the entry cost distribution, on the x-axis of each panel, is multiplied by a constant for confidentiality and displayed in a numeraire currency per thousand auctions.

distribution, some advertisers are more likely to receive very low entry cost draws, which would have previously been prohibitive of entry. Those tend to be the advertisers that would only enter under very low entry costs, which are typically advertisers running non-retargeting campaigns. Because of this ambiguous effect on participation, a higher variance parameter of the entry cost distribution also leads to an ambiguous effect on winning probabilities.

These simulation exercises establish a link between the effects of different parameters for the entry cost distribution on the unobserved number of bidders and the observed winning probabilities. A higher mean parameter has a negative effect on the number of bids

Figure 6: Simulations Results of Varying Variance Parameter of Entry Cost Distribution



This figure shows simulation results that indicate the effect of varying the variance parameter of the entry cost distribution on four key quantities of the model: average number of submitted bids (Panel A); highest and second highest bid (Panel B); equilibrium entry cost threshold (Panel C); and average winning probability conditional on entry (Panel D). The variance parameter of the entry cost distribution, on the x-axis of each panel, is multiplied by a constant for confidentiality and displayed in a numeraire currency per thousand auctions.

submitted, leading to increased equilibrium entry cost thresholds and winning probabilities for advertisers. In contrast, a higher variance parameter has an ambiguous effect on participation and winning probabilities due to the increased likelihood of some advertisers receiving very low entry cost draws. Therefore, separate identification of the mean and variance of the entry cost distribution is possible by connecting the variations in unobserved bidder numbers to the variations in observed winning probabilities.

7 Results

This section reports the results of the SMM estimation. First, it presents the estimated parameters that compose the model. Next, it shows the equilibrium entry cost threshold policy used by advertisers in this market.

7.1 Parameter Estimates and Moment Comparison

Table 7 shows the vector of estimated parameters, normalized by a constant for confidentiality reasons.[11] The parameter for the retargeting premium, β_r, is N$1.28. This represents the monetary difference in valuations between non-retargeting and retargeting. The estimated distribution for idiosyncratic shocks is Lognormal(-8.03, 0.02) for non-retargeting bidders and Lognormal(-8.03, 9.38) for retargeting bidders, while the entry costs distribution is Lognormal(-3.88, 2.16). This raises two observations. First, applying these parameters to Equation 3 implies that the average bid for a non-retargeting advertiser is N$0.012 while that for a retargeting bidder is N$1.33. This is in line with the summary statistics presented in Section 3, where the winning bids for retargeting campaigns are around 100 times higher than those for non-retargeting campaigns. This demonstrates that the model is able to capture the large premium advertisers derive from retargeting. Second, the average entry cost drawn by advertisers is N$0.06. This represents approximately 4% of the typical retargeting bid, but as much as five times the typical non-retargeting bid. This is also in line with the empirical regularity that advertisers simultaneously run various retargeting campaigns but typically only run one broad, non-retargeting campaign. Because of the large retargeting premium in their valuations, advertisers generally derive a larger surplus in retargeting campaigns. Therefore, they can afford to pay higher entry costs to run many of those campaigns.

[11] To avoid reporting confidential level information, all level parameters are expressed in normalized units. Specifically, let X be a random variable such that $\log(X) \sim N(\mu, \sigma^2)$. Let c be an arbitrary constant. Then, defining $\widetilde{X} = c \times X$ implies $\log(\widetilde{X}) \sim N(\log(c) + \mu, \sigma^2)$. Thus, the normalization shifts the mean of $\log(X)$ but leaves its variance unchanged. All reported parameters are therefore reported up to a multiplicative constant.

Table 7: Estimated Parameters

Parameter	Interpretation	Estimate (N$)
β_{ret}	Valuation premium for retargeting	1.28
μ_k	Mean parameter for entry cost distribution	-3.88
μ_ϵ	Mean parameter for idiosyncratic shock distribution	-8.03
σ_k^2	Variance parameter for entry cost distribution	2.16
$\sigma_{\epsilon,nr}^2$	Variance parameter for idiosyncratic non-retargeting	0.02
$\sigma_{\epsilon,r}^2$	Variance parameter for idiosyncratic retargeting	9.38

This table shows the parameters estimated in the model, their interpretation and estimated value in a numeraire currency per thousand auctions (N$).

Table 8: Winning Probability Conditional on Bidding

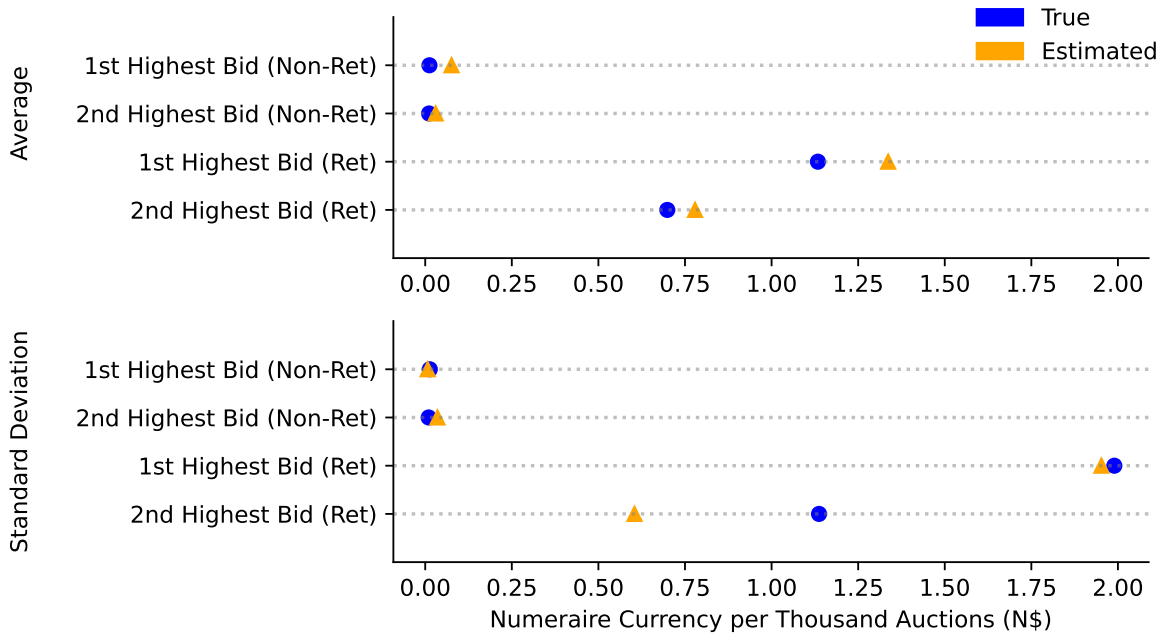
Equilibrium object	Data	Model
Average winning probability for retargeting (%)	55.7	51.9
Average winning probability for non-retargeting (%)	1.9	11.1

This table compares the average winning probability conditional on bidding observed in the data to the one predicted by the model, in each targeting setting.

Figure 7 compares the first set of simulated moments against their counterpart in the true dataset. The model fits those moments well for both the level of bids as well as their variation. For example, the average observed second highest bid for retargeting is N$0.70, and the model simulates an average of N$0.78. The observed average highest bid for non-retargeting is N$0.012 and the model simulates it at N$0.075.

Table 8 compares the average winning probabilities observed in the data to the averages predicted by the model for each targeting setting. The model is able to fit the winning probabilities for retargeting well, but it overestimates the winning probability for non-retargeting bidders. The average winning probability imputed from the data for the simulated potential bidders in retargeting is 55.7%, and the average predicted by the model is 51.9%. However, for non-retargeting campaigns the imputed mean is 1.9% while that simulated by the model is 11.1%.

Figure 7: True and Estimated Moments - Bids



This figure compares the true set of bid moments used in estimation to their simulated counterparts. The value of moments is displayed on the x-axis in a numeraire currency per thousand auctions (N$). The first subplot shows moments pertaining to the average for highest and second highest bids of retargeting and non-retargeting campaigns, while the last subplot shows the moments for standard deviations. Blue dots represent the true moments, and orange triangles represent the simulated moments.

7.2 Equilibrium Entry Cost Threshold

The equilibrium entry cost thresholds are computed as the value of the entry cost that would make bidders indifferent about entering an auction (Equation 6). Table 9 shows the average estimated thresholds for simulated advertisers in each targeting setting.

Advertisers that decide to employ retargeting exhibit higher entry cost thresholds than those not participating in retargeting. Specifically, the average entry threshold for the retargeting setting is N$0.060, whereas for non-retargeting it is N$0.003. This disparity arises because retargeting provides greater surplus for bidders. Advertisers running retargeting campaigns are assured that their ads will be shown to a relevant consumer if they win the auction. The high information value of this targeted audience justifies the higher entry cost thresholds for advertisers.

Table 9: Equilibrium Entry Cost Thresholds

Equilibrium Object	Value (N$)
Average equilibrium entry cost threshold for retargeting	0.060
Average equilibrium entry cost threshold for non-retargeting	0.003

This table shows the average equilibrium entry cost thresholds predicted by the model in each targeting setting. Values are multiplied by a constant and displayed in a numeraire currency per thousand auctions (N$).

8 Counterfactual Exercises

This section presents the results of four counterfactual exercises. These exercises investigate how changes to the information environment governing targeted advertising reshape equilibrium bidding and participation in the online ad industry. A central theme in markets with heterogeneous goods is that the degree of information used to segment the market determines both the quality of matches and the intensity of competition. In the display advertising industry, third-party cookies are a key mechanism through which advertisers segment consumers into granular retargeting audiences. Because such segmentation affects who participates in which auctions, modifying the information structure available to advertisers has direct implications for auction competition. This perspective motivates all four counterfactuals below. First, subsection 8.1 describes the procedure used to separately quantify the matching and the competition effect. Then, the four counterfactuals are presented.

8.1 Disentangling Match from Competition

Separately quantifying the matching and competition effects requires isolating how these mechanisms operate in equilibrium. To separate these equilibrium forces, the analysis proceeds in two steps. First, compute expected auction revenue under the status quo, with no change in the information structure. Next, for each counterfactual, compute an intermediate scenario in which advertisers face the altered information structure, but their equilibrium entry thresholds are held fixed at status-quo levels. The only change relative to the status

quo is thus the deterioration in the match between advertisers and consumers. The difference in revenue (and advertiser surplus) between this intermediate scenario and the status quo defines the matching effect. Then, compute expected auction revenue under the full counterfactual equilibrium, allowing advertisers to optimally adjust their entry strategies to the new environment. The further change in revenue, between the full counterfactual equilibrium and the intermediate scenario, defines the competition effect, capturing how strategic entry adjustments amplify or mitigate the initial loss of information. By construction, the total effect of the counterfactual is the sum of these two components.

Formally, let $Z(I, \boldsymbol{k}^*)$ denote expected auction revenue under a given information structure I and equilibrium entry thresholds $\boldsymbol{k}^*$. Then, the decomposition for expected change in revenue follows:

$$\underbrace{[Z(I^{cf}, \boldsymbol{k}^{*,\boldsymbol{cf}}) - Z(I^{sq}, \boldsymbol{k}^{*,\boldsymbol{sq}})]}_{\text{Total Effect}} = \underbrace{[Z(I^{cf}, \boldsymbol{k}^{*,\boldsymbol{sq}}) - Z(I^{sq}, \boldsymbol{k}^{*,\boldsymbol{sq}})]}_{\text{Matching Effect}} + \underbrace{[Z(I^{cf}, \boldsymbol{k}^{*,\boldsymbol{cf}}) - Z(I^{cf}, \boldsymbol{k}^{*,\boldsymbol{sq}})]}_{\text{Competition Effect}}$$

8.2 Counterfactual A: Banning Third-Party Cookies

The first counterfactual removes the information that enables retargeting by shutting down that segment entirely, following the industry's trend to move to a privacy-friendly environment. For example, Safari and Firefox have already blocked third-party cookies by default on their browsers, and in 2020 Google announced it would remove third-party cookies from Chrome – a plan that was later abandoned. Although such ongoing data privacy initiatives make this scenario empirically relevant, the conceptual motivation for this counterfactual is broader: eliminating retargeting collapses a finely segmented market into a coarse one. Doing so forces advertisers to compete over a pooled set of impressions rather than selectively entering only the auctions in which they find the most valuable consumers.

In the model, banning third-party cookies means that advertisers cannot choose to do retargeting, and are restricted to either run non-retargeting campaigns or to leave the market. I run the simulation algorithm to solve the model as described in Section 5.2 with

the only change being that entry costs for retargeting are prohibitively high. That is, I set $\kappa_{ij}^{r} \approx \infty$ but do not change κ_{ij}^{nr}.

The results for the new entry cost thresholds appear in Table 10. There are two noteworthy facts about this result. First, any positive equilibrium entry cost threshold could justify a rational advertiser not to enter the market, given that all drawn entry costs are infinite. That is, no advertiser can ever satisfy $\kappa_{ij}^{nr} \leq k_r^*$ when $\kappa_{ij}^{r} \approx \infty$. This is a natural result in light of Equation 6: because the cost of retargeting is prohibitively high under this counterfactual scenario, any positive ex ante expected payoff of entering retargeting will be lower than the entry cost.

Table 10: Equilibrium Entry Cost Thresholds - Cookies Banned

	Value (N$)	
Equilibrium Object	Cookies Available	Cookies Banned
Avg. eqm. entry cost threshold for retargeting	0.060	-
Avg. eqm. entry cost threshold for non-retargeting	0.003	0.011

This table compares the average equilibrium entry cost thresholds predicted by the model in each targeting setting, in a counterfactual scenario where cookies are banned to the ones when cookies are available. Values are multiplied by a constant and displayed in a numeraire currency per thousand auctions (N$).

Second, the new equilibrium entry cost thresholds for advertisers in non-retargeting campaigns is increased. The intuition for this result is as follows. Before the removal of cookies, advertisers participating in non-retargeting campaigns would frequently compete against other advertisers that used retargeting. However, those in retargeting campaigns hold a substantially higher valuation for the same impression. This premium is captured by β_r in Equation 1. Therefore, on average, the surplus enjoyed by advertisers in non-retargeting campaigns is typically small, evidenced by the average equilibrium entry cost threshold at N$ 0.003. When third-party cookies are removed, advertisers in non-retargeting campaigns no longer compete against those high-valuation advertisers adopting retargeting. Therefore, their surplus when winning auctions is not as small as before, as they only need to outbid advertisers that have a similar valuation as themselves. This allows advertisers in non-retargeting campaigns more room to pay higher entry costs, thereby increasing the equi-

librium entry cost thresholds for non-retargeting in the counterfactual.

The first row in Table 11 shows the separate effects of matching and competition on ad auction revenue following the removal of third-party cookies. The model predicts that the revenue from ad auctions would decrease by 64.7%. This is in line with the existing literature on data privacy regulations studying the deprecation of cookies. For example, Johnson et al. (2020) find that auctions for users who opt out of cookies close at a price 52% lower than for opt-in users, and Alcobendas et al. (2023) find a reduction in publisher revenue of 54% when cookies are removed. Google also ran an internal experiment by disabling cookies for a randomized set of publishers and found that the ad auction revenue received by those publishers decreased by 52% (Ravichandran and Korula 2019).

Table 11: Separate Effects of Match and Competition After a Ban on Third-Party Cookies

% Change in:	Total Effect	Matching Effect	Competition Effect
Auction Revenue	-64.7	-88.7	24.1
Advertiser Surplus	-93.4	-67.4	-26.1

This table shows the decomposition of the total effect of removing the information provided by third-party cookies in the online advertising industry. The total effect is the sum of the matching effect and the competition effect.

I extend this analysis by separately quantifying the importance of the matching and competition effects for the final change in revenue. In particular, the model predicts that if one only considers the lower match between advertisers and consumers resulting from the removal of cookies, ad auction revenue would decrease by 88.7%. However, this decrease in revenue is partially offset by a 24.1% increase in revenue because of fiercer competition. This raises two conclusions. First, publishers and platforms would be severely harmed by such a large decrease in revenue from ad auctions. Usually, the revenue from auctions is proportionally split between the publisher and the ad exchange. Therefore, the 64.7% decline in expected revenue predicted by the model would be translated into a 64.7% lower revenue for publishers and platforms. Second, the decomposition of this number into a matching and competition effect shows that advertisers would be severely harmed if they no longer

had access to retargeting. Advertisers would experience a large decrease in the match value they derive from this market and pay substantially more for non-retargeting impressions as a result of fiercer competition.

To put these magnitudes in perspective, consider the size of the U.S. market. In 2023, advertisers spent roughly $150 billion on online display advertising, of which about one quarter ($38 billion) corresponds to targeted ads allocated via auctions. A 64.7% reduction in this segment therefore represents a loss of approximately $24.6 billion in annual advertising revenue, a contraction large enough to substantially affect both publishers and digital platforms. To illustrate the economic scale of this decline, one can ask how much consumers would have to value their increased privacy benefits in order to "fund" the reduction in market revenue. Spread across roughly 300 million U.S. consumers, offsetting a $24.6 billion decline in targeted ads revenue would require each consumer to value the associated privacy gains at roughly $82 per year.

The second row in Table 11 shows how the decomposition of the matching and competition effects applies to the change in advertiser welfare. I define advertiser welfare as the difference between their valuation derived from winning the auction and the sum of payments and entry costs. For instance, an advertiser that entered the auction but lost has a negative surplus represented by the sunk entry cost paid for participation. Contrary to what happens to ad auction revenue, in this case, both the matching and competition effects act to decrease advertiser welfare. The matching effect represents the lost valuation derived from auctions where advertisers could target the best possible audience for their products. Without third-party cookies, this is no longer possible. As a consequence, winners obtain a lower average surplus due to a lower match with the consumer to whom they show their ad.

In addition, the surplus derived by advertisers from auctions is further reduced by the fact that now they face a higher number of bidders: without the fierce competition from high-valuation retargeting bidders, advertisers conducting non-retargeting campaigns have more incentives to enter the market. This causes the second-highest bid in the auctions

to increase, which further reduces advertisers surplus. In equilibrium, advertisers lose 93.4% of their surplus from the removal of third-party cookies. This effect consists of a 67.4% reduction due to a lower match with consumers and a further 26.1% surplus reduction due to fiercer competition in auctions. This means that the competition effect represents 27.9% of the total decrease in advertiser surplus from the removal of third-party cookies. That is, almost a third of the reason why advertisers are worse-off following a complete ban of third party cookies is due to a squeeze in the auction surplus caused by tougher competition in the industry.

8.3 Counterfactual B: Taxing Third-Party Cookies

The second counterfactual increases the cost of entering the retargeting segment. From a market design perspective, participation costs are a natural regulatory lever: by altering the cost of accessing finely segmented auctions, regulators can directly reshape which advertisers choose to operate in targeted segments and, therefore, how thick those segments become. Two scenarios motivate this exercise. First, a regulator could impose a "cookie tax" that raises the operating cost of personalized targeting for privacy-related reasons, making retargeting financially less attractive relative to broad campaigns. Another scenario motivating this counterfactual is one in which advertisers face high adjustment costs when adopting alternative targeting technologies, such as Google's proposed Privacy Sandbox,[12] which require rebuilding tracking infrastructure and reconfiguring bidding systems.[13] In both cases, the economic mechanism is the same: increasing the cost of using granular information discourages participation in narrowly defined segments, pushing advertisers toward broader auctions and altering both match quality and competitive intensity in equilibrium.

As opposed to banning cookies altogether, this counterfactual exercise increases the entry cost for retargeting by 100%, that is, it sets $\kappa_{ij}^{r,tax} = 2\kappa_{ij}^{r}$. Table 12 displays the

[12] `https://privacysandbox.com/` (Archive.org, accessed on 06/21/2024)

[13] `https://www.aol.com/news/googles-privacy-sandbox-adoption-costs-100300733.html` (Archive.org, accessed on 09/09/2024)

Table 12: Equilibrium Entry Cost Thresholds - Cookies Taxed

Equilibrium Object	Value (N$) Cookies Available	Cookies Taxed
Avg. eqm. entry cost threshold for retargeting	0.060	0.073
Avg. eqm. entry cost threshold for non-retargeting	0.003	0.006

This table compares the average equilibrium entry cost thresholds predicted by the model in each targeting setting, in a counterfactual scenario where cookies are taxed to the ones when cookies are available as usual. Values are multiplied by a constant and displayed in a numeraire currency per thousand auctions (N$).

equilibrium entry cost thresholds under this counterfactual. The result that the entry cost thresholds for both retargeting and non-retargeting campaigns increases appears counterintuitive at first glance. However, recall that the entry cost paid by retargeting campaigns is doubled. Therefore, participation in retargeting campaigns is reduced in this counterfactual exercise, as the increase in the threshold does not offset the effect of doubling entry costs paid for retargeting. As fewer retargeting campaigns compete, non-retargeting campaigns can experience higher surplus in the auctions, thereby increasing their entry cost thresholds and participation in auctions.

Table 13 shows the decomposition of the total effect of doubling entry costs for retargeting. The first row indicates that a cookie tax, or an alternative and more costly targeting technology, is effective in softening the negative effect of cookie deprecation on auction revenue. Under this new scenario, auction revenue decreases by 15.5%, a lower figure compared to the 64.7% decrease when banning cookies and do not replacing them at all. Matching and competition effects are quantitatively relevant, with matching representing a 25.6% decrease in revenue, offset by a 10.2% increase due to tougher competition. These numbers are in line with recent tests conducted by Kobayashi et al. (2024) to assess the effectiveness of Google's Privacy Sandbox. In a series of experiments using thousands of advertisers in the US market, the authors find that Privacy Sandbox dollar conversions decrease by 14%.

The second row in Table 13 shows that advertisers would have their surplus decreased by only 1.2% in this scenario. The matching and competition effects almost entirely

Table 13: Separate Effects of Match and Competition After a "Cookie Tax"

% Change in:	Total Effect	Matching Effect	Competition Effect
Auction Revenue	-15.5	-25.6	10.2
Advertiser Surplus	-1.2	17.9	-19.2

This table shows the decomposition of the total effect of doubling the entry cost to do retargeting. The total effect is the sum of the matching effect and the competition effect.

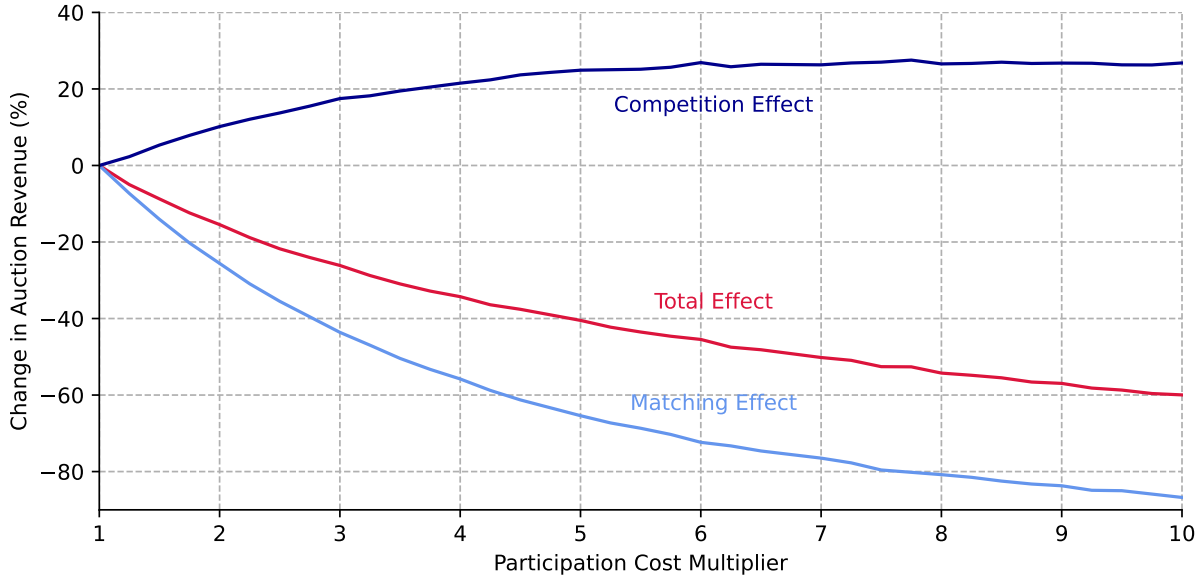
offset each other, with an increase in surplus by 17.9% due to better match and a decrease in surplus by 19.2% due to fiercer competition. This happens because the survivors that still use retargeting are now facing an even thinner market than before. Whenever those bidders enter auctions, they now expect to have a higher surplus as they face fewer bidders also employing retargeting. Non-retargeting bidders are now more incentivized to participate as there are fewer retargeting bidders. This tougher competition tends to squeeze their margins even further. Therefore, the slight decrease in advertiser surplus happens because remaining retargeting bidders that use potential alternative and more costly technologies such as the Privacy Sandbox are benefited at the expense of non-retargeting bidders.

Multiplying entry costs by a factor of two is an approximation to a potential cookie tax or to the adoption costs faced by advertisers that seek to use more privacy-friendly technologies. However, there is no variation in the dataset that allows the identification of a precise cost multiplier that best represents these scenarios. For this reason, Figures 8 and 9 decompose the matching and competition effects on auction revenue and advertiser surplus for different entry cost multipliers.

Recall that banning cookies without any alternative technology, effectively removing retargeting, is estimated to decrease auction revenue by 64.7%. The industry could experience almost half of that effect with an entry cost multiplier as small as 3. This evidences that a substantial portion of the decrease in revenue happens early on, for not so large values of adoption costs or taxes. Therefore, from a market design perspective, any proposal to substitute individual-level identifiers should focus on decreasing adoption costs required to use it. Low adoption costs would decrease the reallocation of campaigns across targeting

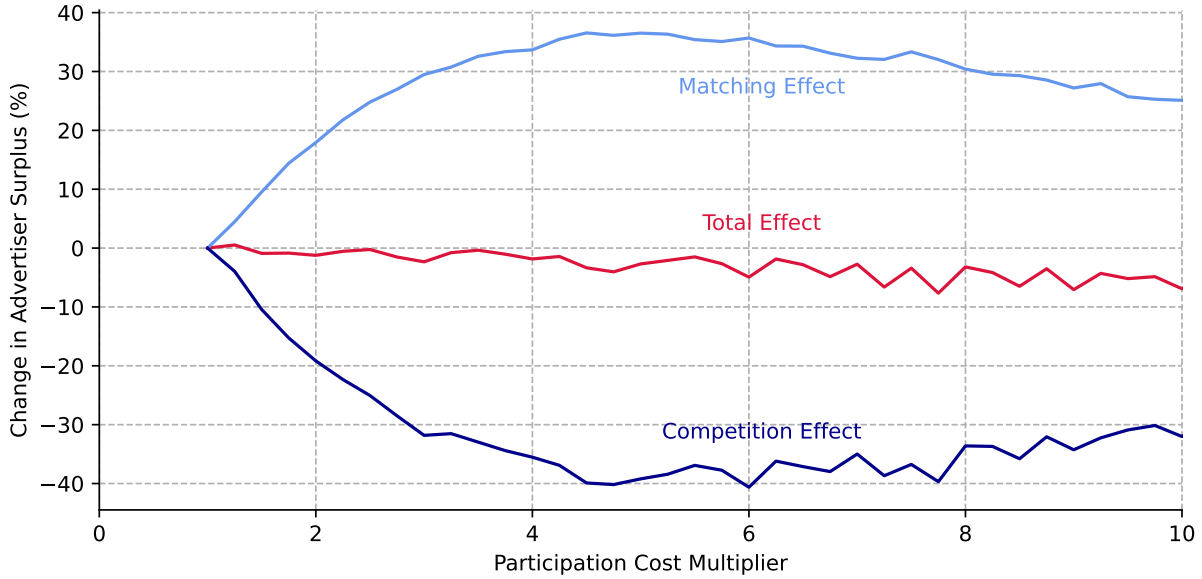
segments, effectively softening the increase in competition in non-retargeting segments.

Figure 8: The Separate Effects of Match and Competition on Auction Revenue for Different Values of Entry Costs Multipliers



This figure shows the decomposition of the equilibrium effect on auction revenue (Y-axis) for different values of an entry cost multiplier (X-axis). The total effect is represented by the red line; the matching effect is represented by the light blue line; and the competition effect is represented by the dark blue line.

Figure 9: The Separate Effects of Match and Competition on Advertiser Surplus for Different Values of Entry Costs Multipliers



This figure shows the decomposition of the equilibrium effect on advertiser surplus (Y-axis) for different values of an entry cost multiplier (X-axis). The total effect is represented by the red line; the matching effect is represented by the light blue line; and the competition effect is represented by the dark blue line.

This softening in competition occurs because the increased cost to use such identifiers reallocates campaigns across targeting segments. The total effect of varying entry costs on advertiser surplus is limited. This happens because the matching and competition effects almost offset each other: higher entry costs make the survivors extract a higher match whenever they show their ads to valuable consumers, but the increased participation of non-retargeting campaigns increases winning bids, squeezing some of that surplus.

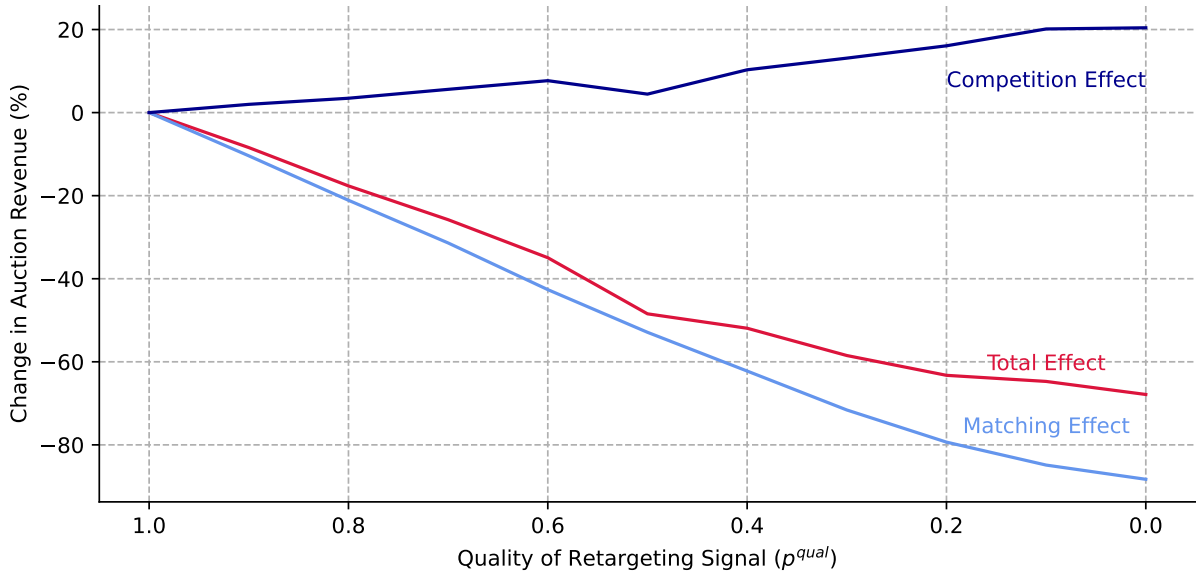
8.4 Counterfactual C: Varying Quality of Retargeting Signal

The third counterfactual is motivated by the increasing number of regulatory requirements for websites to ask consumer consent before placing cookies on their browsers, a movement sparked by Europe's GDPR. This exercise examines how auction outcomes change when the information used to determine which impressions belong to a retargeting audience becomes less reliable. Retargeting depends on a binary signal that identifies whether a consumer previously interacted with an advertiser; reducing the quality of this signal weakens the mechanism that sorts advertisers into their high-match auctions. Concretely, as signal quality deteriorates, some genuine retargeting opportunities fail to be recognized as such, and advertisers lose access to the subset of auctions in which their valuations are highest. The result is a contraction of the retargeting segment and a shift in participation toward broader, more competitive markets.

In this counterfactual, instead of perfectly identifying whether a consumer belongs to an advertiser's retargeting audience, the signal is now correct only with probability p^{qual}. When a consumer j that belongs to advertiser i's audience enters a website and triggers an auction, advertiser i is called to participate in the retargeting auction with probability p^{qual}. When $p^{qual} = 1$, the baseline environment obtains; when $p^{qual} = 0$, the signal conveys no information and retargeting collapses into non-retargeting. Intermediate values correspond to different quality levels of the retargeting signal.

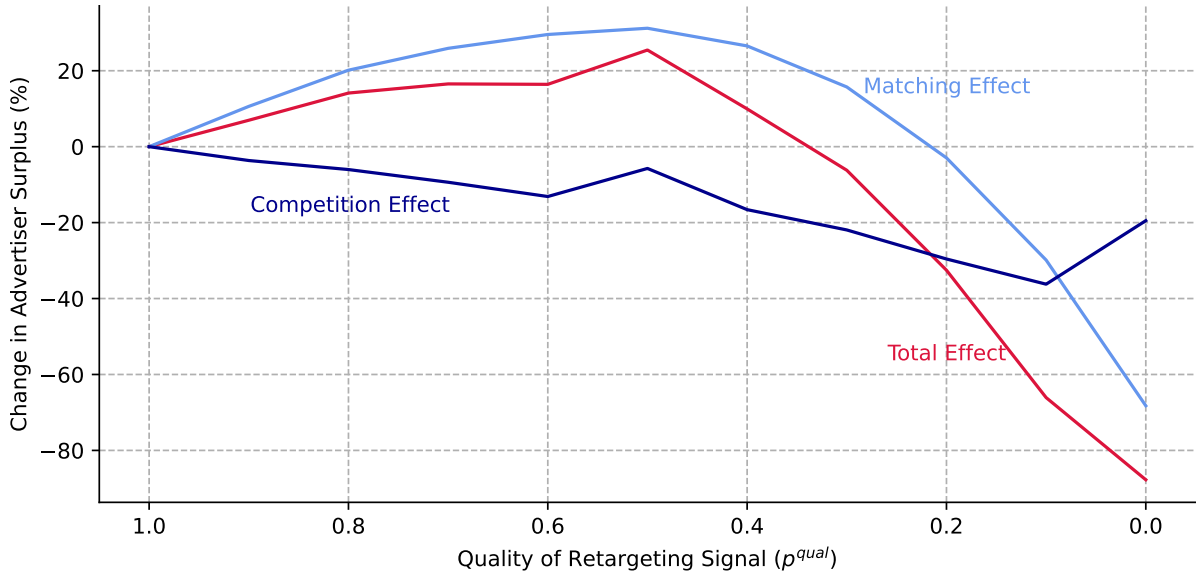
Figures 10 and 11 show the decomposition of the effects of lower signal quality

Figure 10: The Separate Effects of Match and Competition on Auction Revenue for Different Values of Signal Quality



This figure shows the decomposition of the equilibrium effect on auction revenue (Y-axis) for different values of the quality of the retargeting signal (X-axis). The total effect is represented by the red line; the matching effect is represented by the light blue line; and the competition effect is represented by the dark blue line.

Figure 11: The Separate Effects of Match and Competition on Advertiser Surplus for Different Values of Signal Quality



This figure shows the decomposition of the equilibrium effect on advertiser surplus (Y-axis) for different values of the quality of the retargeting signal (X-axis). The total effect is represented by the red line; the matching effect is represented by the light blue line; and the competition effect is represented by the dark blue line.

on auction revenue and advertiser surplus, respectively. As expected, auction revenue falls monotonically: reducing the probability p^{qual} that cookies correctly identify an advertiser's retargeting audience weakens the match between advertisers and consumers, and this loss in match quality dominates the increase in competition that arises as the signal becomes noisier.

A more interesting pattern emerges when decomposing advertiser surplus. In Figure 11, both the total effect and the matching component initially increase as signal quality deteriorates, while the competition component remains negative across all values of p^{qual}. This initial rise in the matching component reflects an intrinsic property of this counterfactual: even though fewer advertisers are correctly brought into the auction, the subset of retargeting campaigns that still find their intended consumers continue to generate high-value matches. For these advertisers, the decline in the number of correctly matched rivals implies that the prices advertisers pay in those successful matches fall. Similarly to Counterfactual B, this tends to increase the surplus of retargeting advertisers that participate in the auctions. However, this time participants do not pay a higher entry cost in the form of a tax or an alternative targeting technology: there is no offsetting negative aspect of participating in a retargeting auction. Thus, conditional on a correct match occurring, the surplus extracted from that match increases because advertisers' valuations and participation costs are unchanged while the second-highest bid becomes lower on average. That is, lower signal quality only serves to make the market thinner than before, and the advertisers that stay active derive a higher surplus compared to the baseline scenario. As signal quality deteriorates further, however, the probability of forming a correct match collapses. When p^{qual} is sufficiently low, the expected number of valuable matches becomes too small to justify paying the entry costs of running retargeting campaigns, causing both the matching component and total advertiser surplus to turn negative.

8.5 Counterfactual D: Varying Precision of Retargeting Signal

The final counterfactual, motivated by a privacy policy that aims at "hiding the consumer in the crowds", introduces noise to the individual-level identifier signal.[14] It varies the precision of the individual-level identifier signal available to advertisers. Unlike signal quality, which determines whether a retargeting campaign is correctly deemed eligible to bid, signal precision governs what happens conditional on being invited to bid. With perfect precision, every impression routed to a retargeting campaign corresponds to a consumer who previously visited the advertiser's website. When precision declines, some of these impressions are in fact generated by consumers outside the advertiser's retargeting audience. As a result, even when the advertiser enters and wins an impression routed to its retargeting campaign, the underlying match value may be low. This counterfactual therefore examines a form of audience conflation: previously distinct retargeting and non-retargeting opportunities become partially pooled. From a market design perspective, this mirrors a regulatory environment in which fine-grained audience segmentation is blurred, reducing the informational advantage of retargeting relative to broad campaigns.

This time, conditional on being routed to a retargeting campaign, the consumer is relevant to advertiser i only with probability p^{prec}. When $p^{prec} = 1$ every impression assigned to a retargeting campaign corresponds to a genuine past visitor, as in the baseline. When $p^{prec} = 0$, the signal provides no useful information, and impressions routed to retargeting campaigns are indistinguishable from broad, non-retargeting opportunities. Intermediate values of p^{prec} capture environments in which audience definitions are partially pooled and the informational advantage of retargeting is diluted. This exercise captures regulatory scenarios in which privacy protection policies reduce the granularity of user-level identifiers without fully eliminating them: for example, by requiring coarser audience buckets, blurring retargeting lists, or introducing noise to avoid re-identification of consumers.

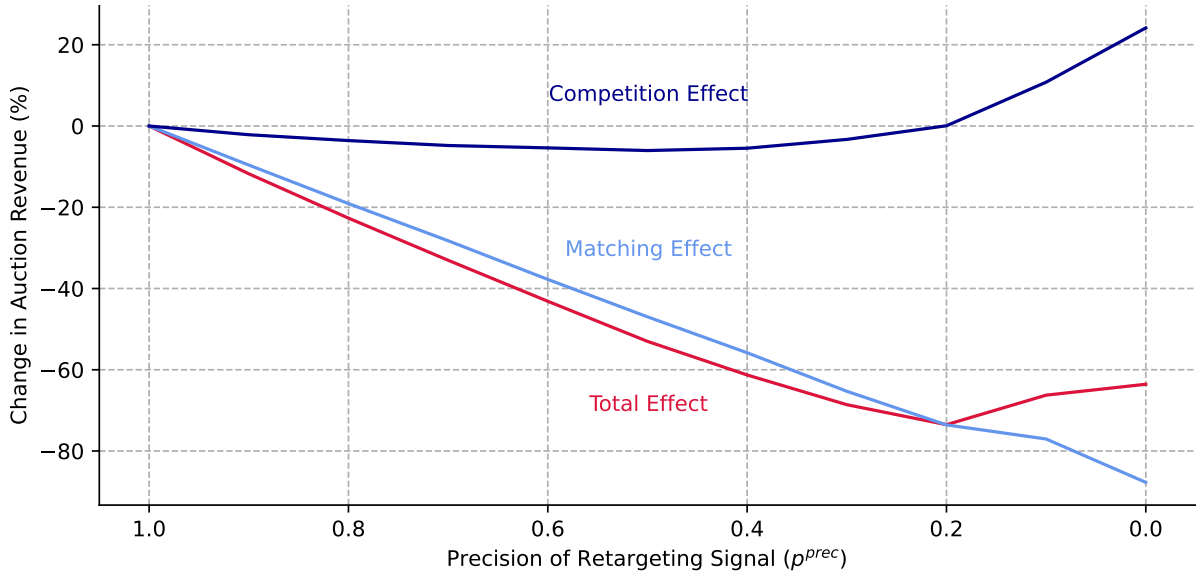
[14] Introducing noise to targeting signals is the main strategy in one of the APIs from Google's Privacy Sandbox, called "Topics". More details can be found at `https://privacysandbox.google.com/private-advertising/topics/why-topics` (Archive.org, accessed on 07/26/2025)

This counterfactual contrasts with the previous one, which reduces the quality of the signal. When signal quality falls, fewer true past visitors are correctly recognized, so the volume of retargeting opportunities shrinks. In contrast, lowering signal precision does not change how often retargeting campaigns are invited to bid. Instead, it changes the composition of those opportunities by mixing true and false positives. Thus, while a lower signal quality reduces the frequency with which advertisers encounter valuable customers, low signal precision preserves that frequency but dilutes the information content of each retargeting impression, conflating retargeting and non-retargeting impressions.

Figures 12 and 13 report how auction revenue and advertiser surplus evolve as the precision of the retargeting signal decreases. For most values of p^{prec}, both revenue and surplus decline in an approximately linear fashion. The total decline in revenue is slightly amplified by the competition effect. The mechanism is straightforward: as signal precision falls, the expected value of a retargeting impression collapses because retargeting advertisers increasingly win auctions involving consumers who are not true past visitors. Since equilibrium bids equal expected valuations conditional on winning, retargeting bidders sharply reduce their bids and participation. Although lower bids induce a modest increase in non-retargeting participation, the decline in both the number and aggressiveness of retargeting bidders is strong enough to pull down auction revenue over most of the range of signal precisions. This pattern reverses once precision falls below roughly 20%. In that region, the remaining retargeting bidders are already bidding extremely low amounts and entering very few auctions. Therefore, their contribution to the second-highest bid becomes negligible. As a result, the dominant force becomes the influx of non-retargeting advertisers, whose increased participation raises competitive pressure and pushes revenue upward. Thus, at very low levels of precision, tougher competition outweighs the loss of match value.

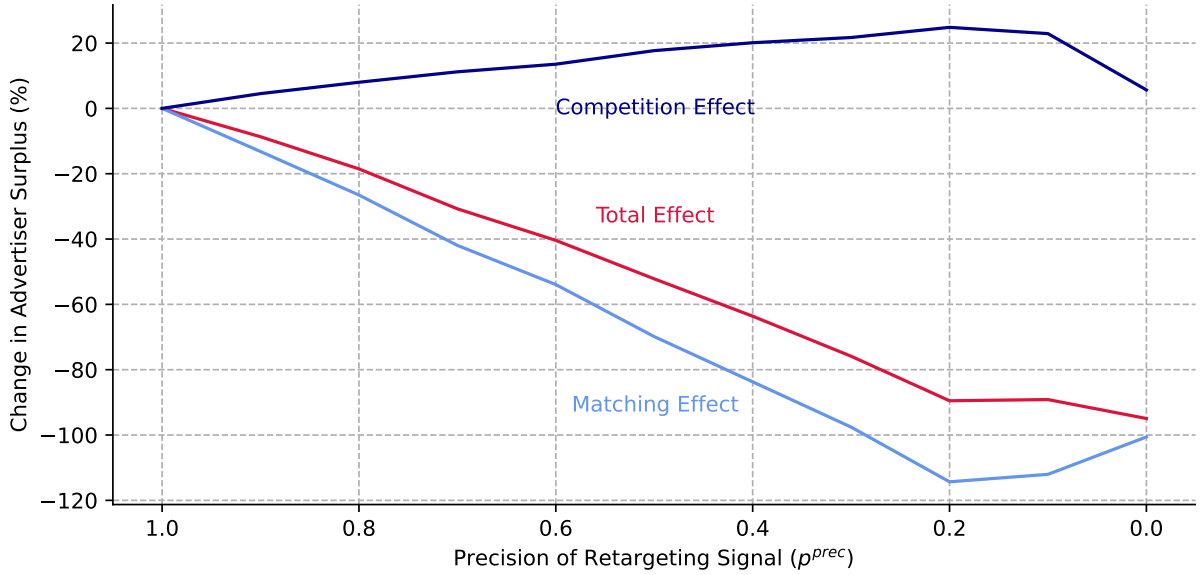
Advertiser surplus exhibits a closely related pattern. For most of the range of p^{prec}, both total surplus and the matching component fall roughly linearly with precision, reflecting the loss of value from incorrectly assigned retargeting impressions. However, the

Figure 12: The Separate Effects of Match and Competition on Auction Revenue for Different Values of Signal Precision



This figure shows the decomposition of the equilibrium effect on auction revenue (Y-axis) for different values of the precision of the retargeting signal (X-axis). The total effect is represented by the red line; the matching effect is represented by the red line; and the competition effect is represented by the dark blue line.

Figure 13: The Separate Effects of Match and Competition on Advertiser Surplus for Different Values of Signal Precision



This figure shows the decomposition of the equilibrium effect on advertiser surplus (Y-axis) for different values of the precision of the retargeting signal (X-axis). The total effect is represented by the red line; the matching effect is represented by the red line; and the competition effect is represented by the dark blue line.

competition effect increases surplus over this region. The same mechanism that depresses revenue (very low retargeting bids) raises the surplus of the retargeting advertisers who still enter: conditional on winning, when they do happen to match with a true past visitor, the gap between their (low) payment and their ex-post valuation is large. At extremely low levels of precision, this effect disappears. The probability of encountering a true retargeting consumer becomes so small that retargeting entry essentially vanishes, and the expansion of non-retargeting participation becomes the dominant force once again, reducing surplus through heightened competition among non-targeting bidders.

9 Conclusion

This paper studies a fundamental market design trade-off in environments where participation is endogenous: providing agents with more informative signals can raise revenue by improving the match between buyers and sellers, while at the same time discouraging participation by fragmenting markets and thinning competition, which decreases revenue. Despite its importance, empirical evidence on this trade-off remains scarce. This paper provides such evidence in the context of one of the largest auction markets in the world – the online display advertising market, where advertisers compete in auctions for individual consumers and where the ability to identify users enables selective entry into high-value auctions while reducing overall market thickness.

To bring these ideas to the data, this paper leverages a novel dataset containing 1.4 billion auctions from the Brazilian market and estimates a model of advertiser bidding with endogenous market thickness. The model replicates salient features of the data: most notably, the large bid premium for retargeting and the substantial heterogeneity in participation thresholds between retargeting and non-retargeting campaigns.

Using the structural estimates, the paper evaluates counterfactual scenarios pertaining to different information structures in the online advertising industry. Specifically,

four scenarios are considered: (i) a complete ban on individual-level identifiers, removing advertisers' ability to do retargeting; (ii) a "cookie tax" that makes retargeting more costly or, alternatively, the use of more costly alternative targeting technologies; (iii) a deterioration of the signal quality provided by individual-level identifiers, reducing the share of times advertisers are correctly called to retargeting auctions; and (iv) a conflation scenario where some non-retargeting consumers are mixed in the retargeting audience, effectively decreasing the precision of the retargeting signal. In the benchmark exercise, scenario (i) finds that auction revenue would decrease by roughly 65%, with an 89% decline due to lower match values and a 24% increase due to fiercer competition as advertisers shift toward broader audiences.

Taken together, the counterfactuals illustrate two fundamentally distinct channels through which policy interventions can affect targeted advertising markets. Interventions that raise participation costs, such as transitioning to alternative targeting technologies or imposing taxes on the use of cookies, primarily operate by selecting which advertisers continue to use retargeting. When costs rise, some advertisers exit, but survivors benefit from softer competition within the remaining retargeting segment. By contrast, interventions that degrade signal precision reshape the informational content of retargeting itself. Lower precision effectively collapses the premium value of retargeted impressions, causing retargeting advertisers to exit not because it becomes expensive, but because it becomes uninformative. This loss of targeting precision crowds out all retargeting advertisers symmetrically, lowering both participation and bids across the board, with no compensating benefit to the advertisers that continue doing retargeting.

These findings highlight the importance of accounting for equilibrium participation responses when evaluating data-privacy policies and technological changes in digital advertising markets. The online ads ecosystem is far broader than the targeted display auctions studied in this paper. Many related environments, from contextual advertising to platform-level optimization tools, exhibit the same interplay between information, value, and participation. I hope this paper encourages further empirical work that brings these

equilibrium forces to the forefront, providing a more complete understanding of how policy tools related to the availability of information in markets reshape incentives, competition, and ultimately market performance.

Appendices

A Experiment Details

This section presents more institutional details surrounding the design and implementation of the experiment.

A.1 Experiment Design

The experiment was implemented in partnership with The Company. Between February 22 and February 25, 2022, I exogenously varied the bids of twelve live campaigns belonging to six advertisers operating in the Brazilian market. The purpose of the experiment was to test whether online ad auctions are better described by an independent private values (IPV) or common values (CV) framework. Establishing this distinction is essential for the structural model, as the IPV assumption rules out forms of selection on unobservables that would otherwise contaminate the recovery of bidders' valuation distributions.

The core idea of the experiment is straightforward. For each selected advertiser, The Company altered the bid submitted to the demand-side platform for two of its active campaigns: one retargeting and one non-retargeting. These changes followed a pre-specified schedule: on Tuesday, February 22, the base bid was increased by 10% relative to the base bid on Monday; on Wednesday, it was increased by 50% relative to Monday; on Thursday, it was decreased by 10% relative to Monday; and on Friday, it was decreased by 50% relative to Monday. In each case, only the base bid level was altered, leaving all targeting settings unchanged. The performance of these treated campaigns is compared to auctions from the same weekday in the prior week, so that the comparison remains within identical calendar conditions. Although the basic intuition is simple, several design choices were required to ensure that the econometric comparison between treated and control periods is valid.

The first step was advertiser selection. I asked The Company to choose six ad-

vertisers, classified internally as two "small," two "medium," and two "large" based on their weekly website traffic and consequently ad spending pattern. The purpose of this request was to ensure that the experiment included advertisers operating at different scales, rather than drawing all observations from a narrow segment of The Company's client base. Advertisers of different sizes may differ along various operational or strategic dimensions, and including a diverse set of advertisers avoids constructing an experimental sample that inadvertently reflects a single type. Since The Company already maintained this classification, I asked them to randomly select two advertisers from each group.

A second institutional detail concerns the object of manipulation. The treatment varies the base bid, defined as the bid transmitted by the advertiser to the demand-side platform. The platform's bidding algorithm then submits auction-level bids that average to (at most) the base bid over the course of the day. Footnote **??** in the main text explains this mechanism, but a numerical example follows: if the base bid is N$1.00, the algorithm will submit auction-specific bids that fluctuate around this target, with the daily average constrained not to exceed N$1.00. If the base bid is increased by 10%, the entire distribution of submitted bids shifts accordingly.

A third important feature of the institutional environment is that changes in the base bid are not reflected immediately in the submitted bids observed in the data. There is a delay between modifying the base bid in the platform interface and seeing the corresponding change in winning bids in the auctions. This delay is directly relevant for defining the treatment period. Each day during the experiment, The Company updated the base bid at approximately 11:00 BRT. In the data, this update is visible as a temporary period in which the treated campaigns stop winning auctions, typically lasting between fifteen minutes and two hours. According to The Company, this interruption corresponds to the demand-side platform's system updating its internal bidding parameters after the change in the base bid. When such a gap is detected, I define the start of treatment for that day as the moment auction wins resume. When no clear interruption appears, likely because the update

was processed quickly, I conservatively assume treatment begins at 14:00 BRT, providing a three-hour buffer to ensure that the update has propagated fully. Treatment is assumed to end at 08:00 BRT the following day, unless a second morning gap clearly indicates a later resumption.

The control period for each treated campaign-day is taken to be the identical time interval on the same weekday in the week prior to the experiment. For example, if campaign A is treated on a Tuesday beginning at 11:34 and ending at 08:00 on Wednesday, the control period is defined as Tuesday 11:34 to Wednesday 08:00 of the prior week. This choice of control period has two advantages. First, it keeps the comparison as close as possible in calendar time, minimizing concerns about seasonality or patterns related to the day of the week. Second, using the previous week ensures that the base bid in the control period matches the base bid at the beginning of the treatment period. For instance, suppose a campaign's base bid was N$1.00 on the Monday of the experiment week, and treatment increased it by 10 percent. The corresponding Monday of the previous week also had a base bid of N$1.00, guaranteeing that the comparison is between an N$1.10 bid and the appropriate N$1.00 baseline. Using longer lags, such as the same weekday one month earlier, would risk differences in the underlying base bid, undermining the experimental contrast.

B Pseudo-Code for Solving the Model

The following pseudo-code implements the simulation algorithm that computes the two equilibrium objects in the model: the bidding strategies and the entry cost thresholds.

1. Guess a vector $\boldsymbol{k_0^*}$ of participation cost thresholds. Each entry i, s in this vector corresponds to the entry thresholds used by advertiser i in segment s

    2. Guess a vector $\bar{\boldsymbol{b}}_{\boldsymbol{0}}$ of non-random components of bids. Similarly, each entry i, s in this vector corresponds to the non-random component of the bid used by advertiser i in segment s

3. Simulate entry decisions by comparing drawn entry costs $\boldsymbol{\kappa}$ to the (guessed) entry thresholds $\boldsymbol{k_0^*}$

4. Simulate bidding decisions by summing drawn idiosyncratic shocks $\boldsymbol{\epsilon}$ to the non-random component of bids: $\boldsymbol{b_0^* = \bar{b}_0 + \epsilon}$ (Equation 5)

5. Entrants bid and payoffs are realized

   - Repeat steps 2 to 5 N_{sim} times to get $\bar{b}_{is} = \mathbb{E}[v|i \text{ wins in } s]\ \forall i, s$ (Equation 4)
   - Update $\boldsymbol{\bar{b}(k_0^*)}$ until convergence. This is the equilibrium non-random component of bids for the initially guessed vector of entry cost thresholds

6. Compute $\boldsymbol{\pi(\bar{b}(k_0^*))}$, advertisers' ex-ante expected profit of entering each segment

7. Update $\boldsymbol{k_1^* = \pi(\bar{b}(k_0^*))}$. This represents the indifference condition that defines the equilibrium entry cost threshold (Equation 6)

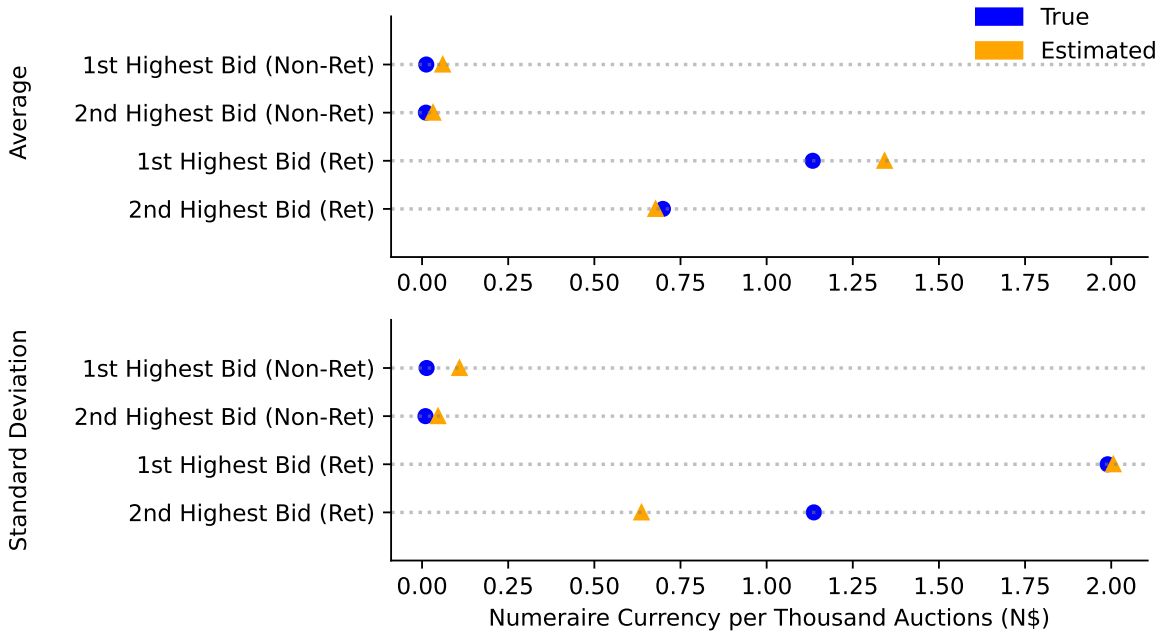
   - Repeat steps 1 to 7 until convergence of participation cost thresholds

C Robustness Checks: Structural Model

This section presents a robustness exercise showing that increasing the number of potential bidders in the model from 10 to 30 does not materially change the model's ability to match key features of the data. That is, I show evidence indicating that the main results are not driven by the choice of $N = 10$. When the model is re-estimated and simulated with $N = 30$, its fit to the data remains stable across all three sets of objects shown here.

To illustrate this, I report three objects analogous to those shown in the main text for the baseline specification: in Figure 14 the comparison between empirical and simulated average and standard deviation moments; in Table 14 the comparison between empirical and simulated winning probabilities; and in Table 15 the equilibrium participation cost thresholds implied by the model.

Figure 14: True and Estimated Moments with 30 Simulated Bidders - Bids



This figure compares the true set of bid moments used in estimation to their simulated counterparts, using 30 simulated bidders instead of 10. The value of moments is displayed on the x-axis in a numeraire currency per thousand auctions (N$). The first subplot shows moments pertaining to the average for highest and second highest bids of retargeting and non-retargeting campaigns, while the last subplot shows the moments for standard deviations. Blue dots represent the true moments, and orange triangles represent the simulated moments.

Table 14: Winning Probability Conditional on Bidding Using 30 Simulated Bidders

Equilibrium object	Data	Model
Average winning probability for retargeting (%)	56.0	53.4
Average winning probability for non-retargeting (%)	1.9	7.3

This table compares the average winning probability conditional on bidding observed in the data to the one predicted by the model, using 30 simulated potential bidders in each targeting setting.

Table 15: Equilibrium Entry Cost Thresholds Using 30 Simulated Bidders

Equilibrium object	Value (N$)
Average equilibrium entry cost threshold for retargeting	0.025
Average equilibrium entry cost threshold for non-retargeting	0.001

This table shows the average equilibrium entry cost thresholds predicted by the model with 30 simulated potential bidders in each targeting setting. Values are multiplied by a constant and displayed in a numeraire currency per thousand auctions (N$).

Figure 14 compares the observed moments in the data with those generated by the model when $N = 30$. This first set of moments includes averages and standard deviations of the highest and second-highest bids. The model continues to closely reproduce these empirical patterns, and the magnitude of the discrepancies is similar (and often smaller) to those obtained under the baseline specification with $N = 10$.

Table 14 compares observed and simulated winning probabilities for each campaign type under $N = 30$. The fit is again comparable to the baseline: the simulated winning probabilities track the empirical ones across the retargeting and non-retargeting campaigns, with no systematic deviations arising from the larger number of bidders If anything, a larger number of potential bidders allows the model to more accurately predict the winning probability for non-retargeting bidders.

Finally, Table 15 reports the estimated equilibrium entry cost thresholds for retargeting and non-retargeting segments when $N = 30$. These thresholds are similar in magnitude to those obtained in the baseline estimation, and the relative ordering between segments remains unchanged. This confirms that the participation decisions implied by the model are not sensitive to the increase in the number of potential bidders.

Online Appendix

OA.1 Regulatory Landscape of Third-Party Cookies

While first-party cookies are generally regarded as beneficial for users, third-party cookies are not. Although they facilitate the delivery of more relevant ads to consumers, third-party cookies are often deemed invasive for several reasons. These cookies are frequently set without explicit user knowledge or consent, enabling companies to track user behavior across websites without the user's awareness or approval. By doing so, companies can create detailed user profiles by monitoring users across multiple websites, violating consumers'

online privacy. Furthermore, these companies often share or sell these data to other entities, again without the user's consent. Such concerns have encouraged a wave of data privacy legislation across Europe and the United States in recent years.

The legislation that set the standards for this regulatory surge was the General Data Protection Regulation (GDPR). It was implemented in Europe in 2018 and has become the main legislation for data privacy, regarded as the standard for subsequent rules. The GDPR sets standards for data collection, processing, and storage for websites offering goods or services to European citizens, mandating that websites acquire users' consent to collect and use their personal data. Following the GDPR, several countries have enacted their own data privacy laws, such as the California Consumer Privacy Act (CCPA) in the United States, China's Personal Information Protection Law (PIPL), and Brazil's General Data Protection Law (LGPD).

In anticipation of stricter regulations concerning consumer privacy, major industry players have initiated movements to phase out the use of third-party cookies. For example, Safari and Firefox have already blocked third-party cookies by default on their browsers. In 2020, Google Chrome, which accounts for approximately two-thirds of the global browser market share,[15] announced a plan to remove third-party cookies as well. This plan has garnered substantial attention from both advertisers and regulatory bodies in Europe, in particular the Competition and Markets Authority (CMA) in the UK. The primary concern among legislators was that the removal of third-party cookies from Chrome could solidify Google's already dominant market position. Advertisers would need to explore alternative methods to reach targeted audiences, potentially relying on Google's proprietary solutions,[16] further strengthening Google's market power. In 2025, Google decided to abandon most of the development around its privacy initiative, largely in response to regulatory concerns

[15] `https://gs.statcounter.com/browser-market-share` (Archive.org, accessed on 06/21/2024)

[16] Google's proposal to replace third-party cookies is the Privacy Sandbox. It is a set of solutions aimed at allowing advertisers to still run targeted ads while enhancing consumer privacy. For further details, see `https://privacysandbox.com/` (Archive.org, accessed on 06/21/2024)

and oversight from the CMA.[17] However, many advertisers and industry participants had already begun shifting toward a more privacy-oriented ecosystem,[18] meaning the transformation toward fewer third-party cookies was underway, regardless of the results of Google's efforts.

OA.2 Additional Data Details

The data used in this paper comes from a Brazilian company (henceforth, The Company) that acts as an intermediary between advertisers and demand-side platforms.[19] Their primary service is assisting advertisers in identifying the most suitable target audiences for their products, with a strong emphasis on retargeting strategies. Advertisers seeking to implement retargeting campaigns must integrate a code snippet into their websites, commonly referred to as a *pixel*. This pixel facilitates the exchange of cookies between the advertiser's domain and the ad servers. The Company possesses the technological infrastructure to support advertisers in this implementation process. Additionally, The Company helps advertisers in determining the most profitable segments within an advertiser's website for executing effective retargeting strategies.

I use two main datasets. First, I have access to auction-level data encompassing all ad auctions won by The Company's clients between November 2021 and March 2022. For each auction, I observe the winning bid, the advertiser's payment,[20] and identifiers for the advertiser, publisher, and user. Additionally, the dataset includes the date and time of the auction, indicators for whether the user viewed[21] or clicked the ad, and variables describing

[17] `https://privacysandbox.com/news/update-on-plans-for-privacy-sandbox-technologies/` (Archive.org, accessed on 11/03/2025)

[18] `https://www.adexchanger.com/programmatic/most-chrome-privacy-sandbox-testers-see-a-silver-lining-in-their-investment/`(Archive.org, accessed on 05/05/2025)

[19] For confidentiality reasons, I cannot disclose the name of the company from which I sourced the data.

[20] If the auction is first-price, those two quantities are the same.

[21] According to the Interactive Advertisement Bureau, an ad banner is considered to be viewed by a user if at least 50% of the ad was visible on the user's screen for at least one continuous second. See `https://www.iab.com/news/viewability-has-arrived-what-you-need-to-know-to-see-through-this-sea-change/` (Archive.org, accessed on 06/21/2024)

the user's device type, model, operating system, and geographical location. Notably, I also observe a variable called *campaign ID*. An ad campaign represents a configuration of targeting settings and bid strategies defined by the advertiser to achieve a specific goal. Advertisers can simultaneously run multiple campaigns to reach different target audiences. For instance, a shoe retailer may create one campaign targeting consumers who visited its "running shoes" section, with a maximum bid of $3.00 per thousand auctions. Concurrently, it could run another campaign targeting users who added items to their shopping cart, with a maximum bid of $10 per thousand auctions, or a broad campaign aiming to reach all possible online consumers (i.e. a no-targeting campaign), with a maximum bid of $0.05 per thousand auctions. The presence of the campaign ID in the dataset allows me to observe all targeting configurations for each auction won by the advertisers.

I augment the auction-level data with three unique pieces of information for each campaign. I observe (i) all targeting settings chosen by advertisers in their campaigns; (ii) the resulting audience sizes each campaign reached; and (iii) a discretized CDF of each campaign's winning probability as a function of bids. These data come from the *auction planner*, a tool provided by the demand-side platform to advertisers. The auction planner is a platform that allows advertisers to input their campaign settings and simulate the campaign's performance for a given bidding strategy based on historical data gathered by the demand-side platform.[22] The data I gather from the auction planner complement the auction-level dataset, which does not contain the number of active bidders in each auction. Therefore, I use data on winning probabilities to infer the intensity of competition faced by each campaign.

[22]The auction planner is a common feature that demand-side platforms offer to advertisers. See, for example, the case of Yahoo (`https://www.advertising.yahooinc.com/our-dsp/plan`) (Archive.org, accessed on 06/21/2024)

OA.3 Selected Data for Estimation

Only a subset of the data described in Section 3 is used to estimate the structural model. Since the focus is on the decision to either employ retargeting or no targeting at all, I select a part of the dataset that allows for the cleanest possible comparison between retargeting and no-targeting campaigns.

In the dataset, all 2,121 retargeting campaigns are run on browsers, either on desktops/laptops or mobile devices. However, only 10.3% of these campaigns exclude mobile devices, while 49.5% of them are run exclusively on mobile devices. For this reason, I only use campaigns that are run on mobile devices. I select mobile campaigns that either (i) only do retargeting; or (ii) do not target any other specific settings. This selection serves two purposes. First, it ensures that no other targeting settings confound the bids used to estimate advertisers' valuations. Second, by requiring all campaigns to be run on the same device type, it standardizes the maximum number of consumers reachable by the auctions. In the model, this number was J, the number of consumers. Since the number of consumers in each advertiser's retargeted audience (J_i) is observed, it is possible to compute the unconditional probability $\frac{J_i}{J}$ that a random auction for advertiser i in a non-retargeting setting is for a consumer in i's audience. The data shows that, for mobile campaigns, $J = 80$ million consumers.

From the pool of mobile campaigns, only those that appear for at least 7 days in the data are used. It is common practice in the online advertisement industry for advertisers to test various targeting strategies and keep those that they believe are profitable. By only using campaigns that are mature enough, I ensure that no consider experimental campaigns are used in estimation, which could still be in their initial learning stages. Those campaigns could have bidding behavior not compatible with an optimal bidding policy. Including these campaigns in the estimation could bias the results.

Finally, I only use second-price auctions. In my dataset, the auction format is a mix between first- and second-price auctions (Zeithammer 2019). This is decided at the

moment of the impression, either by the publisher or by the exchange. I simplify the optimal bidding function in the structural model by only considering second price auctions.

OA.4 Details on the Simulation Process

Ideally, the estimation of the structural model described in section 5 would be done with thousands of potential bidders. However, using a large number of potential entrants substantially increases the computational complexity of the algorithm. Thus, the current version of this paper performs an approximation procedure based on the assumption that bidders are identical up to their audience size.

I approximate the structure of the retargeting market by reducing the number of simulated potential bidders. In the data, advertisers running retargeting campaigns typically reach very small audiences: the median retargeting campaign reaches about $J_i = 78,500$ users out of a total mobile device population of $J = 80,000,000$. This implies that a randomly drawn consumer belongs to the median advertiser's retargeting pool only with probability 0.1%. If I were to simulate the industry with, for example, $2,000$ potential bidders each with probability 0.1% of having a random consumer in their retargeting pool, then the probability that no advertiser recognizes that consumer is $(1 - 0.001)^{2000} \approx 13.5\%$. Instead of simulating thousands of bidders with extremely small audiences, which is computationally prohibitive, I simulate a smaller set of representative bidders, each assigned a proportionally larger audience. For instance, with $N = 10$ representative bidders, a probability of roughly 20% per bidder yields $(1 - 0.2)^{10} \approx 10.7\%$, a very similar probability that no advertiser recognizes the consumer. This approximation preserves the key equilibrium object governing market thickness while dramatically reducing computational costs. The only assumption required is that bidders are identical up to their audience size, so that rescaling the number of bidders and their corresponding audience probabilities leaves the relevant aggregate structure intact. In addition, one drawback of this procedure is that with a small simulated N,

the maximum number of potential competitors in non-retargeting auctions is mechanically limited, which makes it impossible for the model to reproduce very low winning probabilities like the 1.9% observed in the planner data. In Appendix C, I show robustness checks relaxing this approximation on the number of potential bidders N. The current estimation uses $N = 10$ representative bidders.

References

Ada, Sıla, Nadia Abou Nabout, and Elea McDonnell Feit (2022). Context Information Can Increase Revenue in Online Display Advertising Auctions: Evidence from a Policy Change. *Journal of Marketing Research* 59.5, pp. 1040–1058.

Ahmadi, Iman, Nadia Abou Nabout, Bernd Skiera, Elham Maleki, and Johannes Fladenhofer (2024). Overwhelming Targeting Options: Selecting Audience Segments for Online Advertising. *International Journal of Research in Marketing* 41.1, pp. 24–40.

Alcobendas, Miguel, Shunto Kobayashi, and Matthew Shum (2023). The Impact of Privacy Measures on Online Advertising Markets. *Working Paper*.

Alcobendas, Miguel and Robert Zeithammer (2021). Adjustment of Bidding Strategies after a Switch to First-Price Rules. *SSRN Electronic Journal*.

Aridor, Guy, Yeon-Koo Che, Brett Hollenbeck, Maximilian Kaiser, and Daniel McCarthy (2025). Evaluating the Impact of Privacy Regulation on E-Commerce Firms: Evidence from Apple's App Tracking Transparency. *Management Science, forthcoming*.

Athey, Susan, Jonathan Levin, and Enrique Seira (2011). Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions. *The Quarterly Journal of Economics* 126.1, pp. 207–257.

Balseiro, Santiago R, Omar Besbes, and Gabriel Y Weintraub (2015). Repeated Auctions with Budgets in Ad Exchanges: Approximations and Design. *Management Science* 61.4, pp. 864–884.

Balseiro, Santiago R. and Yonatan Gur (2019). Learning in Repeated Auctions with Budgets: Regret Minimization and Equilibrium. *Management Science* 65.9, pp. 3952–3968.

Bergemann, Dirk and Alessandro Bonatti (2015). Selling Cookies. *American Economic Journal: Microeconomics* 7.3, pp. 259–294.

Bergemann, Dirk, Tibor Heumann, and Stephen Morris (2021). Selling Impressions: Efficiency vs. Competition. *Working Paper*.

Bergemann, Dirk and Martin Pesendorfer (2007). Information Structures in Optimal Auctions. *Journal of economic theory* 137.1, pp. 580–609.

Budish, Eric, Robin S Lee, and John J Shim (2024). A Theory of Stock Exchange Competition and Innovation: Will the Market Fix the Market? *Journal of Political Economy* 132.4, pp. 1209–1246.

Chen, Jianqing and Jan Stallaert (2014). An economic Analysis of Online Advertising Using Behavioral Targeting. *Mis Quarterly* 38.2, 429–A7.

Choi, Hana, Carl F Mela, Santiago R Balseiro, and Adam Leary (2020). Online Display Advertising Markets: A Literature Review and Future Directions. *Information Systems Research* 31.2, pp. 556–575.

Choi, Hana and Carl F. Mela (2019). Monetizing Online Marketplaces. *Marketing Science*.

Choi, W. Jason and Amin Sayedi (2019). Learning in Online Advertising. *Marketing Science* 38.4, pp. 584–608.

Deisenroth, Daniel, Utsav Manjeer, Zarak Sohail, Steven Tadelis, and Nils Wernerfelt (2024). Digital Advertising and Market Structure: Implications for Privacy Regulation. *Working Paper*.

Despotakis, Stylianos, R. Ravi, and Amin Sayedi (2021). First-Price Auctions in Online Display Advertising. *Journal of Marketing Research* 58.5, pp. 888–907.

Dubé, Jean-Pierre, John G Lynch, Dirk Bergemann, Mert Demirer, Avi Goldfarb, Garrett Johnson, Anja Lambrecht, Tesary Lin, Anna Tuchman, and Catherine E Tucker (2025). The Intended and Unintended Consequences of Privacy Regulation for Consumer Marketing: A Marketing Science Institute Report. *Working Paper*.

Goldberg, Samuel G, Garrett Johnson, and Scott K Shriver (2024). Regulating Privacy Online: An Economic Evaluation of the GDPR. *American Economic Journal: Economic Policy* 16.1, pp. 325–358.

Goldfarb, Avi and Catherine Tucker (2011). Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science* 30.3, pp. 389–404.

Hummel, Patrick and R Preston Mcafee (2016). When Does Improved Targeting Increase Revenue? *ACM Transactions on Economics and Computation* 5.1, pp. 1–29.

Johnson, Garrett (2013). The Impact of Privacy Policy on the Auction Market for Online Display Advertising. *Working Paper*.

Johnson, Garrett, Scott K Shriver, and Shaoyin Du (2020). Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry? *Marketing Science* 39.1, pp. 33–51.

Kobayashi, Shunto J, Garrett Johnson, and Zhengrong Gu (2024). Privacy-Enhanced versus Traditional Retargeting: Ad Effectiveness in an Industry-Wide Field Experiment.

Korganbekova, Malika and Cole Zuber (2023). Balancing User Privacy and Personalization. *Working Paper*.

Krishna, Vijay (2009). *Auction Theory*. Elsevier Inc.

Lambrecht, Anja and Catherine Tucker (2013). When Does Retargeting Work? Information Specificity in Online Advertising. *Journal of Marketing Research* 50.5, pp. 561–576.

Levin, Dan and James L Smith (1994). Equilibrium in Auctions with Entry. *The American Economic Review* 84.3, pp. 585–599.

Levin, Jonathan and Paul Milgrom (2010). Online Advertising: Heterogeneity and Conflation in Market Design. *American Economic Review* 100.2, pp. 603–607.

Li, Tong and Xiaoyong Zheng (2009). Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions. *Review of Economic Studies* 76.4, pp. 1397–1429.

Lu, Shijie and Sha Yang (2020). Targeting Breadth in Behaviorally Targeted Display Advertising. *Working Paper*.

Marotta, Veronica, Vibhanshu Abhishek, and Alessandro Acquisti (2019). "Online Tracking and Publishers' Revenues: An Empirical Analysis". *Workshop on the Economics of Information Security*, pp. 1–35.

Peukert, Christian, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer (2022). Regulatory Spillovers and Data Governance: Evidence from the GDPR. *Marketing Science* 41.4, pp. 746–768.

Rafieian, Omid and Hema Yoganarasimhan (2021). Targeting and Privacy in Mobile Advertising. *Marketing Science* 40.2, pp. 193–218.

Ravichandran, Deepak and Nitish Korula (2019). Effect of Disabling Third-Party Cookies on Publisher Revenue. *Google Report*.

Rayo, Luis and Ilya Segal (2010). Optimal information disclosure. *Journal of Political Economy* 118.5, pp. 949–987.

Roberts, James W and Andrew Sweeting (2010). Entry and Selection in Auctions. *Working Paper*.

— (2016). Bailouts and the Preservation of Competition: The Case of the Federal Timber Contract Payment Modification Act. *American Economic Journal: Microeconomics* 8.3, pp. 257–288.

Samuelson, William F (1985). Competitive Bidding with Entry Costs. *Economics Letters* 17.1–2, pp. 53–57.

Sousa, Tulio (2026). Testing Independent Private Values in Online Ad Auctions With a Field Experiment. *Working Paper, Indiana University*.

Sun, Tianshu, Zhe Yuan, Chunxiao Li, Kaifu Zhang, and Jun Xu (2024). The Value of Personal Data in Internet Commerce: A High-Stakes Field Experiment on Data Regulation Policy. *Management Science* 70.4, pp. 2645–2660.

Vulkan, N, AE Roth, and Z Neeman (2013). *The Handbook of Market Design*. OUP Oxford.

Wernerfelt, Nils, Anna Tuchman, Bradley Shapiro, and Robert Moakler (2022). Estimating the Value of Offsite Data to Advertisers on Meta. *Working Paper*.

Zeithammer, Robert (2019). Soft Floors in Auctions. *Management Science* 65.9, pp. 4204–4221.