Essays in Empirical Industrial Organization

Thesis by
Shunto Jerry Kobayashi

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2024
Defended April 12, 2024

Shunto Jerry Kobayashi
ORCID: 0000-0002-1418-9153

ACKNOWLEDGEMENTS

I would like to extend my sincere gratitude to my advisors, Matthew Shum, Yi Xin, and Gabriel Lopez-Moctezuma. Matthew Shum's invaluable mentorship and deep insights have been instrumental in shaping my research journey, and I was very fortunate to have a seasoned economist like him as my primary advisor. Yi Xin and Gabriel Lopez-Moctezuma offered exceptional guidance and consistent encouragement, which greatly helped me in overcoming hardships. Their belief in my abilities has been a constant source of motivation. I am also deeply grateful to Thomas Palfrey for his guidance and feedback as the committee chair.

Special recognition goes to my co-authors, each of whom has played a significant role in the development of my research. Miguel Alcobendas's collaboration was essential in the first two chapters of this dissertation, and his vast knowledge and strong expertise as an industry expert significantly influenced my research direction. Aldo Lucia's research agenda shaped the development of the third chapter. I am proud to have navigated through challenging times and graduate alongside him. Ke Shi's partnership and contributions were invaluable for the second chapter. Collaborating with him, a fellow student in the same field, significantly boosted my knowledge base and overall morale.

Furthermore, I wish to extend my heartfelt appreciation to my fellow graduate students in the same program, particularly Danny Ebanks and Meng-Zhang Fong, who, along with myself, constituted our cohort and have since graduated. Interacting with Danny and Meng-Zhang has enriched my understanding of political science and game theory, respectively, and I take great pride in being part of the same cohort.

In addition, I am deeply indebted to my family and friends for their constant support and encouragement throughout my journey. I am especially grateful to my parents, Hiroko and Junji, who have provided unwavering love and comfort that were crucial for my achievements. Finally, I am immensely grateful to my beloved wife, Wendy, for her constant support, understanding, and patience. Since I met her in the second year of my PhD, she has been my source of strength throughout this journey.

ABSTRACT

This dissertation comprises three essays related to the field of Empirical Industrial Organization. Chapter 1 and 2 contribute to the economic literature on online advertising auctions, and Chapter 3 contributes to the study of decision-making under risk using structural methods.

In Chapter 1, co-authored with Miguel Alcobendas, we provide a novel empirical analysis of a large-scale sequential market employing auctions to allocate objects to firms with budget constraints. Leveraging a unique proprietary dataset of online ad auctions, we examine the trade-off participants face due to short-run budget constraints. We develop and estimate a finite-horizon dynamic game among bidders with heterogeneous budgets, and we find that dynamic incentives significantly influence their participation and bidding strategies. We conduct a counterfactual simulation comparing first-price and second-price formats, illustrating how dynamics lead to significant disparities in competitive outcomes.

In Chapter 2, co-authored with Miguel Alcobendas, Matthew Shum, and Ke Shi, we investigate the impact of removing third-party cookies on the online advertising market. Utilizing a proprietary dataset of online ad auctions, we document stylized facts about the value of third-party cookies to advertisers. Adopting a structural approach, we simulate counterfactual scenarios to quantify the impact of Google's plan to phase out third-party cookies from Chrome. Our analysis suggests a 54% reduction in publisher revenue and a 40% reduction in advertiser surplus under an outright ban. Introduction of alternative tracking technologies under Google's Privacy Sandbox initiative would mitigate some of the loss. We find big tech firms can leverage their informational advantage to gain a larger surplus from the ban.

In Chapter 3, co-authored with Aldo Lucia, we explore the limited ability of prominent economic models in explaining multiple behavioral patterns. Conducting an experiment with 500 participants, we study two classical behaviors inconsistent with Expected Utility: the common ratio effect and preferences for randomization. We illustrate the lack of generalizability of existing models across these behaviors. Motivated by this, we introduce a novel empirical approach that does not commit on specific decision models. Our method offers more accurate out-of-sample predictions about behaviors under risk, both inside and outside laboratory settings, compared to leading economic models and machine learning algorithms.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

Economists have employed structural econometric methods to analyze the decision-making processes of firms and consumers, especially in the field of Empirical Industrial Organization. These methods enable us to connect data with economic models, generating new insights and allowing for the simulation of counterfactual scenarios that can inform businesses and policymakers. Broadly speaking, this dissertation provides contributions to two different areas by formulating and utilizing structural econometric methods. Chapter 1 and 2 study the strategic interaction between advertising technology firms in online advertising auctions. Chapter 3 analyzes individuals' decision-making under risk by integrating experimental and structural methods.

In Chapter 1, co-authored with Miguel Alcobendas, we provide a novel empirical analysis of a large-scale sequential market that employs auctions to allocate objects to firms with budget constraints, leveraging a unique proprietary dataset of the online advertising market. In this market, because of their short-run budget constraints, participants face a trade-off between winning auctions immediately or holding out for later opportunities. This dynamic incentive prompts them to adjust their entry rates and bidding strategies accordingly. We develop and estimate a finite-horizon dynamic game between bidders with heterogeneous budgets facing a sequence of simultaneous auctions to quantify this incentive and analyze its implication in competition and auction design. We find that a substantial markdown occurs due to the dynamic incentives arising from budget constraints, and this markdown varies significantly among bidders with different budgets. Using the estimated structural model, we provide a counterfactual simulation comparing the first-price and second-price formats. Unlike the standard environment, we find that dynamics and heterogeneous budgets lead to a significant disparity in the welfare distributions under them. This highlights that even a seemingly simple mechanism choice can have competitive implications in such a dynamic environment.

In Chapter 2, co-authored with Miguel Alcobendas, Matthew Shum, and Ke Shi, we study the effect of removing third-party cookies on the online advertising market by leveraging a proprietary dataset of online ad auctions. Online privacy protection has gained momentum in recent years and spurred both government regulations and private-sector initiatives. A centerpiece of this movement is the removal of third-party cookies, which are widely employed to track online user behavior and implement targeted ads, from web browsers. We first document stylized facts about the value of third-party cookies to advertisers. Adopting a structural approach to recover advertisers' valuations from their bids in these auctions, we simulate a few counterfactual scenarios to quantify the impact of Google's plan to phase out third-party cookies from Chrome, its market-leading browser. Our counterfactual analysis suggests that an outright ban would reduce publisher revenue by 54% and advertiser surplus by 40%. The introduction of alternative tracking technologies under Google's Privacy Sandbox initiative would recoup part of the loss. In either case, we find that big tech firms can leverage their informational advantage over their competitors and gain a larger surplus from the ban.

In Chapter 3, co-authored with Aldo Lucia, we illustrate the limited ability of prominent economic models in explaining multiple behavioral patterns. We conduct an experiment with 500 participants that studies two classical behaviors inconsistent with Expected Utility: the common ratio effect and preferences for randomization. The lack of generalizability of leading economic models across these two behaviors calls for the development of new empirical strategies to make predictions. Motivated by this observation, we introduce a novel empirical approach that enables us to predict behavior under risk without leaning on specific decision models. We further demonstrate that this method offers more accurate out-of-sample predictions about behaviors under risk, both inside and outside laboratory settings, than leading economic models and machine learning algorithms.

Chapter 1

DYNAMIC AUCTIONS WITH BUDGET-CONSTRAINED BIDDERS: EVIDENCE FROM THE ONLINE ADVERTISING MARKET

1.1 Introduction

Auctions are employed in many real-world contexts, leading to extensive theoretical and empirical research that provides valuable insights for shaping policy decisions and mechanism design.[1] However, much of this prior research has primarily focused on analyzing auction models where bidders face only one auction. In practice, bidders almost always make bid decisions in the presence of multiple auctions, often conducted sequentially. For instance, sequential auctions are prevalent in procurement, gas and oil lease, wholesale electricity, treasury, art, online retail, and online advertising markets. Such multi-object, sequential auction scenarios have received comparatively less attention in the literature. In particular, there is a notable gap in our understanding regarding sequential auctions in which participating bidders face budget constraints. When auctions are held sequentially, intertemporal budget constraints can strategically link these auctions, influencing competitive dynamics. Financial constraints are pervasive, affecting both consumers with budget limitations and firms operating as buyers, who may face restricted purchasing power due to financial frictions or institutional constraints. Given the prevalence of sequential auctions and budget constraints in real-world settings, research in this domain holds significant promise for informing policy decisions and influencing mechanism design across diverse markets.

We propose a novel structural model of dynamic auctions with budget-constrained bidders and empirically analyze the online display advertising market, where intertemporal budget constraints play a crucial role. Our model offers both tractability and flexibility, enabling predictions of strategic behavior across various auction mechanisms. By estimating model primitives using a proprietary dataset of dynamic first-price auctions for ad opportunities, we find that dynamic incentives significantly affect markdown, varying across bidders with different budgets. We provide a coun-

[1] Refer to Krishna (2009) and Milgrom (2004) for comprehensive introductions to auction theory. See Paarsch et al. (2006), Hickman et al. (2012), Gentry et al. (2018), and Perrigne and Vuong (2019) for overviews of econometric methods and empirical studies on auctions.

terfactual analysis comparing the first-price and second-price mechanisms, which have the same revenue and welfare considerations under the conventional auction model. We discover that dynamics and heterogeneous budgets lead to substantial differences in the surplus distribution. Intermediate and smaller budget bidders fare better in the first-price format due to reduced price variance, which allows for more aggressive bidding. This heightened competition prompts larger bidders to spend quickly, leading to diminished competition in later periods and ultimately benefiting smaller bidders overall. This novel finding highlights the significance of price volatility in shaping competitive outcomes in dynamic mechanism design with budget-constrained buyers.

Our empirical setting is the online display advertising market. This is the market behind online banner and video advertisements, generating more than $100 billion annually in the US. We use a novel proprietary dataset of auctions hosted on Yahoo's ad exchange. In this market, a significant proportion of advertising opportunities are allocated through real-time auctions. When a user visits a website, it triggers an instantaneous auction where the user's characteristics are revealed to bidders. Currently, the market predominantly uses the first-price auction mechanism. The highest bidder secures the privilege of displaying their ad on the user's screen once the page fully loads. These real-time auctions enable advertisers to effectively target users and ensure their ads are presented before users navigate away from the webpage. In this market, advertisers typically hire bidding agents who participate in these auctions on their behalf, and advertisers frequently impose specific campaign budgets on these agents, typically allocated on a daily basis. These bidding agents are frequently affiliated with major tech firms such as Google and Amazon, which tend to attract numerous advertisers, including those with substantial campaign budgets. One responsibility of these bidding agents is to strategize on how to effectively participate and bid within the continuous stream of instantaneous auctions while adhering to the daily budgets assigned to them.

We first document dynamic patterns in the data that are consistent with daily budget constraints. First, we observe a declining trend in both the entry rate and bid levels throughout the day, from morning to evening. This trend is in line with diminishing demand, likely caused by bidders exhausting their daily budgets. In fact, theoretical research has shown that sequential auctions can exhibit such a decreasing price pattern with unit-demand bidders (Engelbrecht-Wiggans, 1994; Bernhardt and Scoones, 1994; Gale and Hausch, 1994). Second, we also find that the entry rate and bid level are negatively affected by the frequency of auctions. In other words, when there is an increase in the number of auctions (from high supply), both the entry rate and bid levels decrease. This suggests that bidders may exercise caution by submitting less competitive bids when faced with a higher volume of auctions, aiming to preserve their future spending capacity and prevent exceeding their budget constraints by winning too many auctions. This relationship is robust to controlling for numerous bidder and auction characteristics, including time fixed effects.

Motivated by the institutional features and the dynamic pattern in the data, we develop a structural model for dynamic auctions with budget-constrained bidders. In this model, each day in the market is represented as a finite-horizon dynamic game where bidders face numerous auctions in each period. The number of auctions per period follows a stochastic daily supply pattern. Bidders have independent private entry costs and valuations for auctions, and they are penalized at the end of the game if their total expenditure exceeds their budget. We analyze the best-response problem under the first-price auction mechanism, which is the current mechanism used in our empirical context. Our analysis reveals that the dynamic constraint introduces another force to depress their bids in addition to the force from being in the first-price auction. There is a tradeoff between allocating their budget toward current auctions versus preserving it for later opportunities. This opportunity cost manifests as an additional markdown whose magnitude depends on the number of remaining periods, the frequency of auctions, and the bidder's remaining budget. In addition, the dynamic tradeoff similarly influences the entry strategy.

Solving our model poses significant computational challenges due to its dynamic nature, exacerbated by several factors. These include the presence of continuous choice variables without closed-form expressions, a finite time horizon resulting in non-stationarity, a relatively large number of players (around thirty), and, most critically, a high-dimensional continuous state space with a continuous state variable (remaining budget) associated to each player. To tackle these complexities, we leverage the fact that bidders do not have access to information about their rivals' spending behaviors. In light of this information asymmetry, we adopt a large-market solution concept in which bidders rely on the equilibrium distribution of players' states for each period as their belief. Hence, each bidder decides his entry and bid strategy conditional on the time period, public state (number of auctions), and his own remaining budget. By employing this approach, we effectively reduce the problem's dimensionality, enabling estimation and counterfactual simulation while still allowing for meaningful analysis of dynamic bidding behaviors.

Following the literature on structural estimation of dynamic games, we adopt a two-step approach to estimate our structural model (Bajari et al., 2007; Aguirregabiria and Mira, 2007). Assuming that the market is in equilibrium under our solution concept, we estimate our structural model by leveraging bidders' best-response problem given their rivals' equilibrium behavior. In the first step, we estimate the time-dependent distribution of the number of auctions, along with the reduced-form entry probability and bid distribution. In the second step, we solve for bidders' entry and bid strategies as best responses to rivals' estimated behaviors from the first stage and estimate the structural parameters through maximum likelihood estimation. This sequential approach allows us to avoid the need for solving the equilibrium and simulating the equilibrium state distribution during the estimation process.

The identification of our structural parameters, such as bidders' budgets and the budget constraint parameter, relies on the exclusion restriction that bidders' valuations are independent of the state variables, which are the frequency of auctions and their remaining budgets. This assumption is required for disentangling the effect on bids from valuations and intertemporal budget constraints. For instance, when bids are low, we must determine whether this is due to low valuations or increased dynamic tradeoffs. The exclusion restriction enables us to identify the parameters relevant to budget constraints by using the correlation between bids and the state variables, which impact only the dynamic tradeoffs. This exogeneity assumption is plausible for our market environment because advertisers and their bidding agents typically compute their valuations for impressions based on a combination of the probability of clicking/making a sale and their willingness to pay for such events. This probability is computed based on that user's contextual and behavioral data alone.

Applying our structural estimation method on a large-scale proprietary dataset of online banner-ad auctions from Yahoo reveals significant dynamic incentives arising from budget constraints. The markdown, representing the gap between valuations and bids, averages 83.5% of expected valuations. Our estimated model demonstrates that first-price auctions induce 59.4% shading, and dynamic budget constraints add an extra 24.2% shading. Our model also quantifies the impact on entry decisions, with an average entry probability of 19.4%, compared to an unconstrained entry probability of 45.1%. These findings underscore the significance of dynamic incentives in this context. Additionally, we observe a notable concentration of estimated budgets among bidders, which has a substantial impact on the heterogeneity in their bids and entry decisions.

Using our estimated structural model, we simulate a counterfactual scenario motivated by a significant institutional change that took place around 2018. During this period, the predominant auction mechanism in the online ad market shifted from the second-price format to the first-price auction. This transition was prompted by concerns within the industry that ad exchanges, serving as intermediaries between publishers and advertisers, were not actually implementing the second-price auction as claimed, leading to a loss of trust among market participants. In response to this industry-wide credibility crisis, market participants advocated for the first-price auction due to its transparency in revealing what winners pay. Motivated by this shift, which happened years before our sample period, we simulate the second-price auction format using the estimated structural model as a counterfactual scenario to analyze the revenue and welfare consequences.

Our counterfactual simulation reveals that the first-price auction yields slightly higher total revenue and total bidder surplus compared to the second-price auction. More importantly, we observe a substantial disparity in welfare distribution between these two auction formats. We find that bidders below the two largest budget holders face more favorable outcomes under the first-price auction. This suggests that, in addition to its transparency benefits, the first-price auction may offer a more robust competition in the presence of market concentration. This outcome can be attributed to the reduced price volatility under the first-price auction. Lower price volatility allows bidders to bid more aggressively, as it enables better control over their spending patterns. While bidders with intermediate-size and small budgets lower their entry rates in response to this increase in competition, the two bidders with the largest budgets keep a similar entry rate since they can afford to. Nevertheless, this increase in competition induces these top bidders to spend more rapidly and leads to decreased competition in later periods. Then, smaller bidders can enjoy this smaller competition and earn more surplus during this period. This difference leads bidders other than the top two bidders to be better off under the first-price auction. This result underscores that even a seemingly simple choice of first-price or second-price can have competitive implications when auctions are conducted sequentially and participated by bidders with heterogeneous budgets.

While prior research has empirically explored dynamic aspects in various auction contexts, our study offers a novel perspective by investigating online display ad auctions with intertemporal budget constraints, contributing the first empirical analysis of dynamic auctions with budget-constrained bidders. Our work builds on the growing empirical literature utilizing structural models to study repeated auctions. In contrast to the studies focusing on procurement auctions with increasing marginal costs (Jofre-Bonet and Pesendorfer, 2003; Groeger, 2014; Raisingh, 2022), eBay auctions with single-unit demand (Hendricks and Sorensen, 2018; Bodoh-Creed et al., 2021; Backus and Lewis, 2023), or oil and gas lease auctions with synergy effects (Kong, 2021), our paper highlights the unique dynamics arising from intertemporal budget constraints in online ad auctions. The common theme in the previous studies is that a bidder's future payoffs depend on whether they win the current auction. Meanwhile, in our environment, the price they would pay also impacts future payoffs by affecting the future spending ability. In essence, the current bid not only determines the probability of winning but also impacts future payoffs through the potential payment. Consequently, our dynamic bidding problem presents an additional layer of complexity, increasing the relevance of price on competitive dynamics, as illustrated in our counterfactual simulation.

In the theoretical auction literature, several papers analyze simultaneous or sequential auctions participated by budget-constrained bidders (Palfrey, 1980; Benoît and Krishna, 2001; Pitchik and Schotter, 1988; Pitchik, 2009; Ghosh and Liu, 2019). Our environment and structural model are substantially different from these theoretical works. We examine an environment that is well approximated by a finite sequence of simultaneous auctions, and it has a large number of auctions in every period (at least thousands) and a relatively large number of bidders (around thirty). In contrast, theoretical studies on simultaneous or sequential auctions often focus on a small number of auctions and bidders (typically two for each) to investigate equilibrium existence and theoretical properties. Consequently, the assessment of revenue and welfare implications for various auction formats in our environment remains theoretically ambiguous. Our counterfactual exercise makes a novel finding that using the first-price format in this environment benefits bidders with smaller budgets by increasing the spending rate of bidders with large budgets.

This paper also contributes to the empirical economic literature on online advertising markets (Yao and Mela, 2008; Athey and Nekipelov, 2011; Celis et al., 2014; Decarolis and Rovigatti, 2021; Ostrovsky and Schwarz, 2023).[2] While Yao

[2]The computer science and operation research literature has studied the theoretical and algo-

and Mela (2008) and Athey and Nekipelov (2011) highlight the presence of intertemporal budget constraints in the sponsored-search ad market and their potential significance, they do not incorporate these constraints in their structural models. Our contribution lies in developing a structural model of dynamic auctions that explicitly incorporates such intertemporal budget constraints. Furthermore, our paper aligns with 2023, which investigates the competition effects of privacy protection measures in the online display ad market, considering firm heterogeneity in their information on consumers. This paper reinforces the importance of accounting for firm heterogeneity when analyzing competition in this market, as it evaluates the competitive implications arising from heterogeneous financial capabilities among bidding firms.

1.2 Institutional Background

Display ad market

The recent online advertisement market employs the real-time bidding process to trade a large portion of impressions, which is the industry term for opportunities to display ads to visitors of websites. As the name suggests, through the real-time bidding (RTB) process, publishers of websites and advertisers trade impressions via auctions in real-time as consumers visit these websites. Each auction typically lasts only milliseconds. Hence, under the RTB process, impressions are sold impression-by-impression rather than via signing contracts in advance for bulks of impressions. Advertisers may display clickable banners or videos after purchasing impressions, and the content of these advertisements may reflect various characteristics of the impressions. For example, a retailer may attempt to retarget consumers by displaying products the consumers viewed in the past. One advantage of the RTB process is that it provides granularity to advertisers for targeting a specific audience. Rather than buying media or ad slots to a loosely targeted audience, the RTB process allows advertisers to target a particular audience directly.

---

rithmic aspects of online ad auctions with budget constraints. See Agarwal et al. (2014), Xu et al. (2015), Balseiro et al. (2015), Balseiro et al. (2020), Conitzer et al. (2022b), Conitzer et al. (2022a), and Gaitonde et al. (2022). Some of these papers focus on developing bidding algorithms with budget constraints. Others use stylized models to study revenue considerations of auctions participated by such algorithms. In contrast, our structural model takes a more general approach by not being tailored to a specific algorithm. Instead, it captures the broader features of any algorithm that could be deployed in the market, providing a robust framework for analysis. Furthermore, our model accounts for the critical non-stationarity of the market, whereas many theoretical studies in this literature assume a stationary environment. This enables our model to reflect the dynamic nature of the market better. Finally, we contribute a novel empirical analysis of online ad auctions with budget-constrained bidders by using a real-world dataset and our structural model.

Generally, the RTB process involves publishers, ad exchanges, demand-side platforms, and advertisers.[3] An *ad exchange* is an online server that hosts auctions. These auctions can have various formats, such as first-price and second-price auctions. Advertisers typically bid in ad auctions through demand-side platforms because it is technologically complex to target individual impressions and optimally bid for them. A *demand-side platform (DSP)* is an intermediary that assists advertisers in targeting and bidding for impressions in ad exchanges, and it typically uses optimized bidding algorithms because of the fast-paced nature of ad auctions. The sequence of the RTB process roughly works as follows:

1. A user visits a webpage of a publisher and triggers an impression.
2. The publisher sends an ad request to an ad exchange containing the user information.
3. The ad exchange starts an auction for the impression and forwards the ad request to demand-side platforms (DSPs).
4. Each DSP decides whether to participate and which advertiser to allocate this impression among its clients, and then it bids on behalf of the chosen advertiser in the auction held in the ad exchange.
5. The advertiser represented by the winning DSP gets the impression.
6. Finally, the corresponding advertisement is displayed to the user.

Marketing campaign and budget settings

When an advertiser wants to start advertising a banner or video ad, they register a marketing campaign with a DSP. The advertiser sets various key marketing campaign parameters, such as performance goals (number of clicks, conversions, or impressions), targeting audience, campaign length, payment scheme, and budget. Advertisers may choose to pay DSPs a fee proportional to spending or a fee per click/conversion. The budget specifies how much the DSP can spend during the campaign period to purchase ad opportunities, and generally this budget is evenly split over days during the campaign. See Figure 1.1 for an example of a daily budget configuration of a marketing campaign. In practice, this daily budget constraint

[3] In reality, there are also *supply-side platforms (SSPs)* that support publishers, but we omit them in our explanation for brevity. See Yuan et al. (2013) and Choi et al. (2020) for more detail.

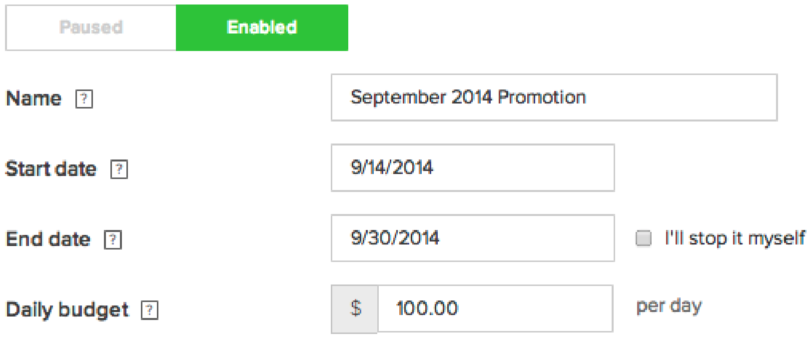


Figure 1.1: Settings of a marketing campaign on a major demand-side platform.

is soft since DSPs often underspend or overspend by a bit; however, it provides a method for advertisers to discipline the spending behavior of the DSPs they hire.

Advertisers have several reasons for setting daily budgets. First, it serves as a safeguard against erroneous spending by the bidding agent. Given the rapid nature of the online ad market, a mistake can be costly, potentially depleting the advertiser's entire budget within a mere hour. Second, advertisers often want to sustain their ad campaign for a longer period than just one day, and the daily budget constraint is a way to ensure that they advertise roughly evenly during the campaign period.

1.3 Data and Stylized Facts

Data description

This paper employs data of ad auctions held at the Yahoo Ad Exchange for ad opportunities on Yahoo's websites. Like other exchanges, the Yahoo Ad Exchange is a clearinghouse that facilitates transactions between publishers and advertisers (represented by DSPs), and it runs first-price auctions. This dataset is suitable for our study for two reasons. First, Yahoo is one of the most popular websites in the US[4], so the data provides a representative sample for our study. Yahoo is one of the most popular publishers that sell banner ad opportunities, and it also provides a diverse range of websites, such as Mail, News, and Finance. Therefore, although DSPs may be bidding for ad opportunities on multiple publishers and even in multiple exchanges, the data provides us a representative sample of ad auctions faced by advertisers who use banner ads.

[4] As of May 2021, Yahoo is ranked fourth in the US popularity by Alexa Rank (`https://www.alexa.com/topsites/countries/US`), which is an industry-standard website ranking.

variable	mean	std	min	median	max
timestamp (PDT)	Thu 00:06:05		Mon 00:00:04	Wed 18:01:42	Sun 23:59:57
Bid	1.000	1.682	0.061	0.577	369.070
Winning bid	2.294	3.441	0.061	1.182	369.070
Number of bidders	7.205	4.732	1.000	7.000	25.000
computer	0.953	0.212	0.000	1.000	1.000
optout	0.066	0.248	0.000	0.000	1.000
match_cookie_prop	0.628	0.336	0.000	0.778	1.000

Table 1.1: Summary statistics.

Summary Statistics

Our dataset contains auctions for impressions generated in the US on Yahoo's websites during a week in the second quarter of 2021. Because there can easily be tens of millions of impressions on just one website per day, we sample our data at a rate of 0.08%. We restrict our attention to a specific popular banner format for simplification.

Table 1.1 provides summary statistics on the key variables in our dataset. We observe data on 8,856,603 bids from 1,229,300 auctions, each of which is for an impression triggered by a user[5]. For confidentiality reasons, we normalize bids to have a sample mean equal to 1, but we may use dollar signs for variables relating to bids in this paper. For each auction, we have the auction outcomes, winning bid (revenue), number of participants (DSPs), and impression characteristics. There are 33 unique DSPs bidding on behalf of 71,011 advertisers in the data; note that each DSP has at most one bid per auction in our dataset. The statistics for the number of participants indicate that these bidders (DSPs) enter only a subset of auctions, which suggests entry is an important behavior to investigate.

The `timestamp` variable provides the time in Eastern Daylight Time when Yahoo held the auction; this variable is central in our analysis as we use this variable to determine the temporal proximity between auctions. We have seven days' worth of data (Monday to Sunday) collected during the second quarter of 2021. The variable `computer` indicates whether the user is accessing from a computer or phone/tablet; it suggests that about 95% of impressions are from computers. Two variables correspond to the availability of the user's information. The variable `optout` is an indicator function of the user opting out from behavioral targeting. When a user opts out, advertisers can no longer target the user based on their personal information; nevertheless, they can still use the user's geographic location

[5]This indicates there were roughly 1.5 billion auctions on Yahoo during this period.

and contextual information. The variable match_cookie_prop is the proportion of DSPs that successfully matched the user with records in their databases via third-party cookies.[6] Thus, it is harder for DSPs to track a user with a small match_cookie_prop. Note that mechanically, we have match_cookie_prop = 0 when the user opted out. Our data show that opt-out users trigger about 6.6% of impressions, and the average proportion of match_cookie_prop is about 63%.[7]

In addition, although we do not report their summary statistics for confidentiality, our dataset contains user characteristics drawn from Yahoo's database of user profiles, which are constructed based on users' cookies and Yahoo accounts (if they exist). Although bidders do not directly observe the content of this database, these variables are good proxies for user information bidders have access to. We have users' gender and age information. The gender variable is either Unknown, Male, or Female, and the age variable is either Unknown, 25 to 44, or 45 plus. The variable seg_size gives the number of market segments that the user belongs to; these segments predict the user's interests in particular topics, such as automobiles and sports. The variables total_rev, num_month_sold, and avg_month_sold summarize the past monetization of impressions generated by the user. The variable total_rev is Yahoo's total revenue from selling the user's past impressions, which is standardized to have mean zero and variance one for confidentiality. The variable num_month_sold counts the number of months when Yahoo monetized the user, and avg_month_rev is the average revenue per month calculated with the two former variables. Finally, the variable profile_length measures in days how long the user profile existed in the database.

Table 1.2 provides the frequency table for two key categorical variables: browser and site name (anonymized for confidentiality). We observe that most impressions come from Chrome, followed by Edge, Safari, and Firefox. The table also shows the range of websites within the Yahoo domain, and we see that although each website has a significant number of observed impressions, there is a considerable variation in their total visits.

Figure 1.2 shows the geographic distribution of impressions graphed based on their geographic coordinates. We observe users accessing from a variety of regions, and many impressions come from cities with high population density, such as New York and Los Angeles. In addition to geographic coordinates, we observe the state and

[6] We use an aggregate measure of cookie match since cookie-match information is unavailable for two DSPs.

[7] See 2023 for how privacy protection measures impact ad auctions in this market.

variable	value	n
browser	Chrome	729484
	Edge	220771
	Safari	138794
	Firefox	103956
	(27 other browsers)	36295
sitename	Site-1	495951
	Site-2	243124
	Site-3	165875
	Site-4	113272
	Site-5	71922
	Site-6	35325
	Site-7	32163
	Site-8	22518
	Site-9	21725
	Site-10	9825
	Site-11	9275
	(5 other sites)	8325

Table 1.2: Frequency table for categorical variables.

city where impressions originate from; these variables are used as control variables in our reduced-form results.

Stylized Facts

We summarize some stylized facts in this market that are consistent with how bidders in this market dynamically participate in ad auctions with budget constraints.

Figure 1.3 shows the time-series plots from an average weekday of the number of auctions, average bid, average number of participants, and average winning bid per 5 minutes on a weekday. We find that these statistics show similar patterns on each day in our data. Since the supply of advertising opportunities is directly tied to online traffic, the frequency of auctions is the highest around noon and the lowest around 3 AM.

Observation 1: Declining Price

In Figure 1.3, we observe that the average bid, average number of bidders, and average winning bid (price) have a declining pattern. This is consistent with bidders having less purchasing power from spending their daily budgets. Because they have less remaining budgets as time goes on, they enter auctions at a lower rate and

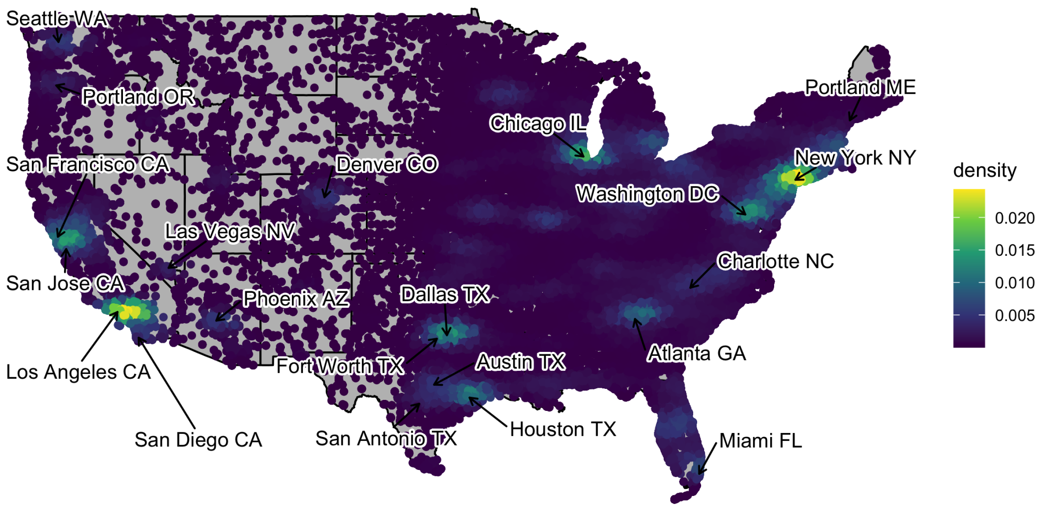


Figure 1.2: Geographic distribution of impressions. Impressions from Alaska and Hawaii are excluded from the figure. The labels are for the top 20 cities with the highest number of impressions.

submit more conservative bids.

The literature on sequential auctions has studied declining price patterns in other settings extensively. This phenomenon might initially appear as an anomaly, seemingly presenting an arbitrage opportunity, but prior studies have identified several mechanisms that can lead to declining prices in sequential auctions. In particular, Engelbrecht-Wiggans (1994), Bernhardt and Scoones (1994), and Gale and Hausch (1994) find that a declining price can manifest in sequential auctions participated by single-unit demand bidders whose valuations are random across objects. Our market environment shares some similarities with these studies, as bidders operate under daily budget constraints, limiting their demand, and there is a significant level of heterogeneity across impressions, which are also horizontally differentiated.

The theoretical studies find that a declining price can occur in sequential auctions due to specific factors. First, there is less competition as time progresses due to diminishing demand. Second, bidders with high valuations also face high delay costs. These costs arise because, as time progresses, there is a chance they will encounter worse objects due to randomness, and they are not guaranteed to win in later periods. Hence, these high-valuation bidders find bidding worthwhile even in earlier periods, where competition is more intense.

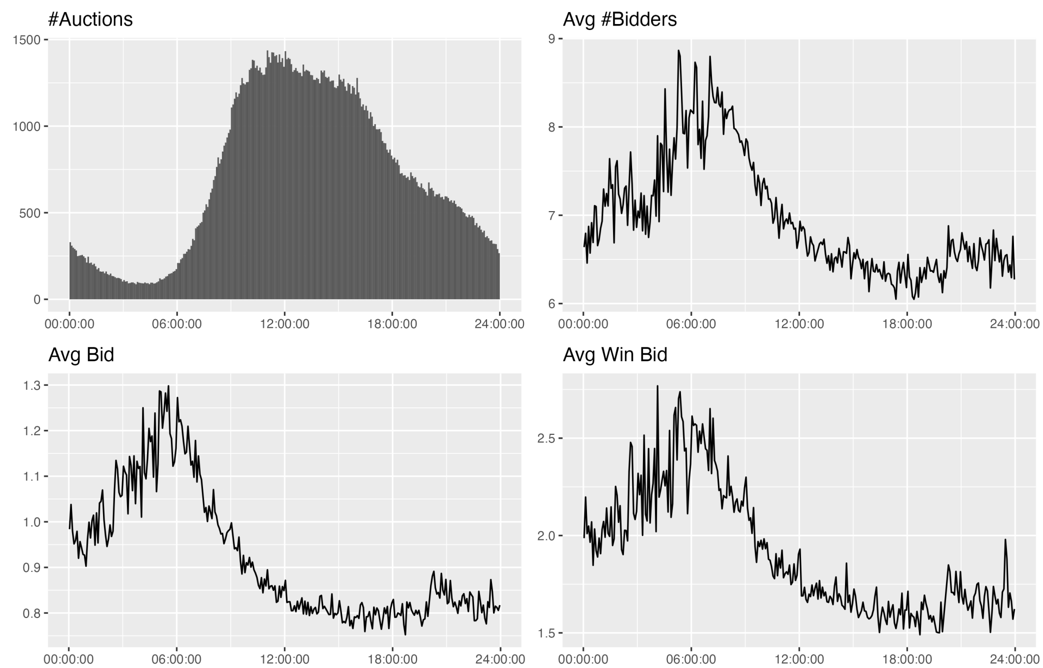


Figure 1.3: Time-series plots from an average weekday for the number of auctions (impressions), average bid, average number of participants, and average winning bid per 5-minute interval. The horizontal axis is the time in Eastern Daylight Time.

Observation 2: Price jumps when budgets are renewed

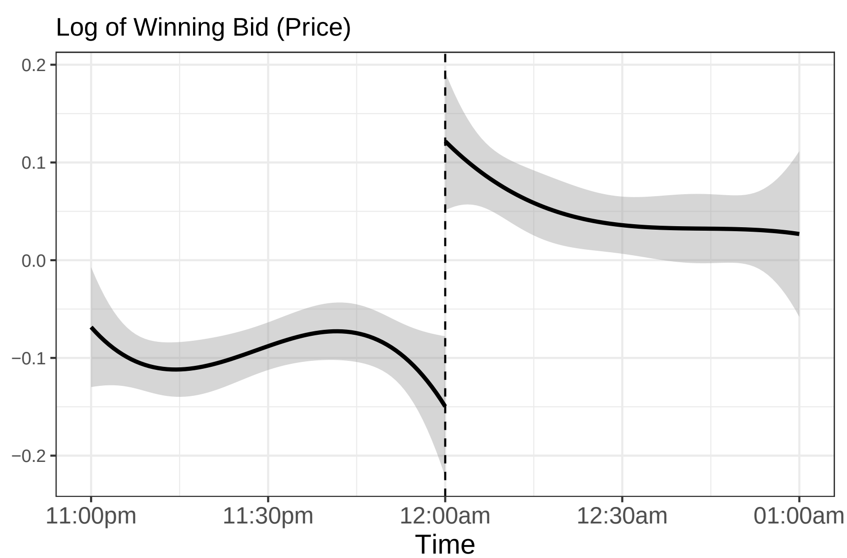


Figure 1.4: Regression discontinuity plot of the log of winning bid (price). We fit cubic polynomials before and after the budget renewal time.

The intra-day declining price pattern is accompanied by a jump in bidders' bidding behavior when their daily budgets are renewed. Using 12AM in the eastern US time as a discontinuity point[8], we perform a regression discontinuity analysis to examine the effect of daily budget renewal on the bidding behavior. We focus on the two-hour time interval around the discontinuity point. In Figure 1.4, we find that there is a significant effect on the average price when the daily budget constraints are reset. We find that bidders become more aggressive in their bids when they are supplied with new budgets from advertisers.[9] Table A.1 in Appendix A.1 shows the regression discontinuity results where we control for a rich set of observed auction characteristics, which show that ceteris paribus, the price in ad auctions jumps by about 40% on average when budgets are renewed.

Observation 3: Price declines when the number of auctions is high

In Figure 1.5, we aggregate auctions in each 5-minute time interval, and we plot the number of auctions within the interval versus the average bid, average number of participants, and average winning bid within the interval. Figure 1.5 shows that when the number of auctions increases, the average bid, average number of entrants, and average winning bid decrease, and vice versa. This inverse relationship is consistent with budget constraints. When there is a large number of auctions, bidders risk hurting their future spending ability or violating budget constraints by winning too many auctions if they submit competitive bids. Thus, they need to depress their bids to mitigate this risk.

This relationship is robust to controlling for the rich observed heterogeneity of impressions and various fixed effects. Table A.2 in Appendix A.1 shows the results from regressing bids and entry decisions on the number of auctions and control variables. As control variables, we include numerous impression characteristics and fixed effects for the websites, browsers, cities, day-hour, DSPs, and advertisers. We include Day-Hour FE to remedy any time-variant unobserved quality of impressions. The reduced-form results show that the coefficient of log of the number of auctions per 5-min interval is negative and statistically significant in both the bid and entry

[8]We find that most bidders' daily budgets are renewed at 12AM in the eastern US time since typically the renewal time is set at 12AM in advertisers' local timezone and most advertisers are in the eastern region. In principle, we should see a change in bidding behavior at 12am in the western US time, for example; however, we do not observe such a consistent significant jump in our data at that time.

[9]As detailed in the preceding section, advertisers enforce daily budget constraints on their bidding agents to mitigate the risk of erroneous overspending and to maintain a consistent presence throughout their ad campaign period. Moreover, the practice of resetting these constraints at the end of each day offers convenience in terms of billing and accounting processes.

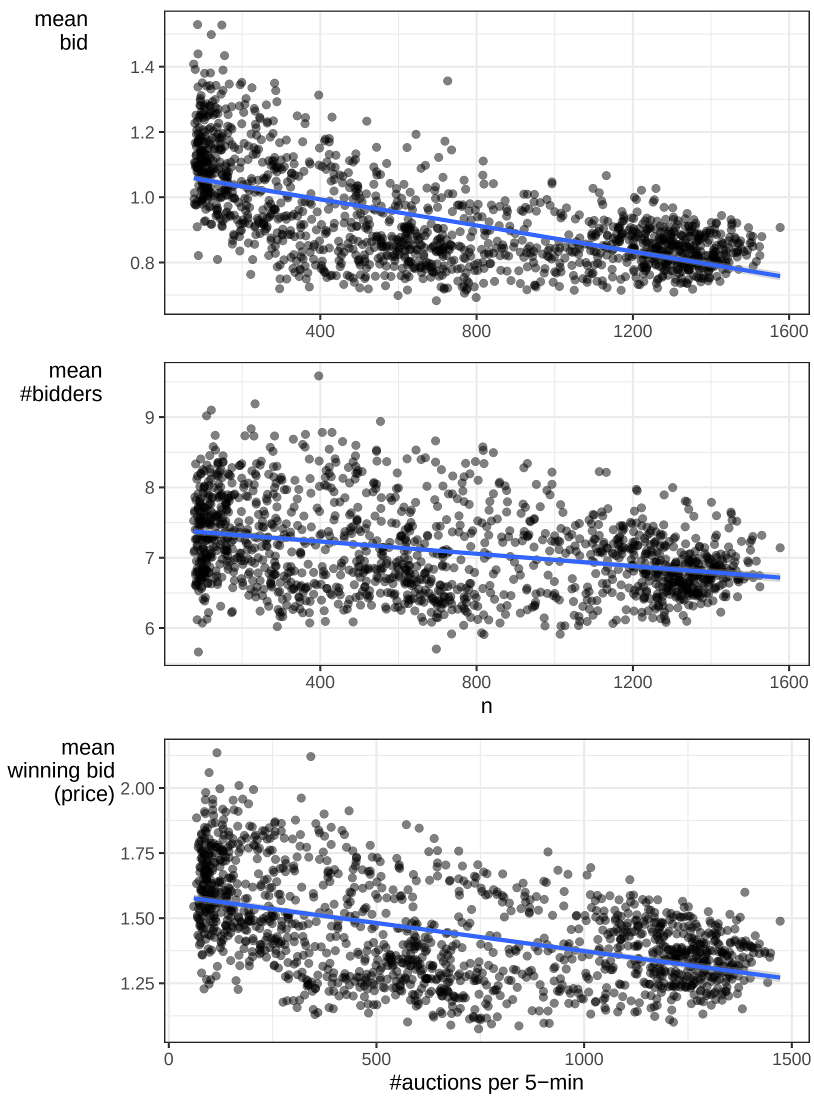


Figure 1.5: Scatter plots of the average bid, average number of participants, and average winning bid (price) versus the number of impressions per 5-minute interval. The blue curves correspond to linear regressions.

regressions. Hence, bidders become more conservative in their entry and bids when there are more auctions.

1.4 Structural Model

Motivated by the institutional settings and the stylized facts from the data, we formulate a structural model of forward-looking bidders making entry and bid decisions dynamically while facing a stream of auctions.

Model Setup

To focus on the intra-day dynamics coming from the daily budget constraint and cyclical supply, we model the market on each day as an isolated strategic environment. On each day, there are $i = 1, \ldots, N$ bidders, and each bidder's initial budget $w_{i1} = w_i$ is independently and privately drawn from F_w at the beginning of the day. These bidders face a sequence of simultaneous auctions for $t = 1, \ldots T$ periods. The last period T is determined and common knowledge; it corresponds to the final period before the end of the day.

At the beginning of each period t, bidders observe the number of auctions K_t, which is drawn from $F_K^{(t)}$, which is time specific. Before his entry costs and valuations are realized for these auctions, each bidder i commits to an entry threshold strategy $\tau_{it} \geq 0$ and bid strategy $b_{it} : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ that are used for each auction $k = 1, \ldots, K_t$. For each auction k, an entry cost C_{ikt} is independently and privately drawn from F_C, and bidder i enters if $C_{ikt} \leq \tau_{it}$. If he enters, then valuation X_{ikt} is independently and privately drawn from F_X, and i submits $b_{it}(X_{ikt})$. To make the optimization problem tractable, we assume that bid strategies take the flexible form $b^\gamma(x) = \sum_{j=1}^{J} \gamma_j h_j(x)$, where $h_j(x)$ are a set of basis functions (such as polynomials or splines) and $\gamma \in \mathbb{R}^J$. Note that this assumption imposes minimal restrictions beyond ensuring smoothness while accommodating a wide range of bidding strategies.[10]

Given the submitted bids, the spot auction rule determines the winner and price for each auction k. With our institutional environment in mind, we suppose that the spot auction follows the first-price auction. Hence, for each auction k, the winner is the highest bidder, and the price is his bid $\max_i B_{ikt}$ where $B_{ikt} = b(X_{ikt} \mid \gamma_{it})$. Then, each bidder receives the goods they won and earns $\sum_{k=1}^{K_t} \mathbb{1}\{C_{ikt} \leq \tau_{it}\}\mathbb{1}\{B_{ikt} > B_{-ikt}\}X_{ikt}$ and pays $S_{it} = \sum_{k=1}^{K_t} \mathbb{1}\{C_{ikt} \leq \tau_{it}\}\mathbb{1}\{B_{ikt} > B_{-ikt}\}B_{ikt}$. In sum, as the

[10]In our empirical application, we use cubic spline basis functions for h_j.

stage payoff, the bidder receives

$$\sum_{k=1}^{K_t} \mathbb{1}\{C_{ikt} \leq \tau_{it}\} \left(\mathbb{1}\{B_{ikt} > B_{-ikt}\} \left(X_{ikt} - B_{kt} \right) - C_{ikt} \right).$$

In addition, the bidder's budget for period $t+1$ is updated as $w_{it+1} = w_{it} - S_{it}$.

After period T, bidders suffer a penalty of $\eta Q(w_{iT+1})$, where $\eta > 0$. $Q(w)$ is a differentiable function such that it is zero when $w > 0$, meaning that the budget constraint is satisfied. This penalty captures any negative consequences associated with violating the budget constraint. In our empirical application, we set $Q(w) = \min(0, w)^2$ to capture the reputation damage that the bidding agent suffers by violating the daily budget constraints imposed by its advertisers.[11]

The following summarizes the sequence of the game:

1. Each bidder i's initial budget $w_{i1} = w_i$ is independently and privately drawn from F_w.
2. For each $t = 1, \ldots, T$,
   a) Bidders observe the number of auctions K_t, which is drawn from $F_K^{(t)}$.
   b) Each bidder i chooses bid strategy $b_{it} : \mathbb{R}_+ \to \mathbb{R}_+$ and entry threshold strategy $\tau_{it} \geq 0$.
   c) For each auction $k = 1, \ldots, K_t$,
      i. i 's entry cost C_{ikt} is independently and privately drawn from F_C.
      ii. i enters if $C_{ikt} \leq \tau_{it}$
      iii. Each entrant's valuation X_{ikt} is independently and privately drawn from F_X.
      iv. Each entrant submits bid $B_{ikt} = b_{it}^{\gamma}(X_{ikt})$
      v. The highest bidder gets the good and pays his own bid.
   d) Each bidder's remaining budget is subtracted by his spending, $w_{it+1} = w_{it} - S_{it}$.
3. Bidders suffer penalty $\eta Q(w_{iT+1})$

[11] The penalty also accommodates other contexts. For example, if there is a borrowing cost from spending beyond the available cash, we can set $Q(w) = \min(0, w)$ with η representing the interest rate. In addition, setting $\eta = \infty$ allows us to incorporate hard budget constraints.

Best-Response Analysis

We begin our analysis by assuming the form of the bidder's belief over the strategic behavior of each other and analyzing the best-response problem. When bidders strategize, the key object that matters is the distribution of competing bids in each auction (accounting for entry). This distribution is influenced by the time period, the number of auctions, and the remaining budgets of competing players. In our empirical application of online banner-ad auctions held at Yahoo's ad exchange, the remaining budgets of players are not public, and when a bidder loses in an auction, they do not get the information of the identity of the winner and the price he paid. Hence, there is very little information a bidder has about other players' spending. Leveraging this asymmetric information, we assume that in period t with K_t auctions, bidders believe that the highest rival bid in each auction is independently drawn from distribution $\Psi_t(\cdot|K_t)$. In the next subsection, we endogenize this distribution by formulating our equilibrium concept.

The idea is that although bidders do not learn each other's private state variable (remaining budget), they can use the interactions from previous games to forecast the competition in each period t and how it changes with respect to the supply level (number of auctions). Not only is this assumption reasonable, but it also greatly improves the tractability of the dynamic game. As our application has a relatively large number of bidders (around thirty), a fully rational belief with complete information over rivals' budgets would lead to a high-dimensional state space, which would make both solving the best-response problem and solving for equilibrium computationally infeasible. The assumption we make over the belief over competing bids turns the best-response problem into a finite-time horizon dynamic problem with two state variables: the number of auctions and the bidder's own remaining budget.

Now, given the belief, we look at a generic bidder's strategic problem while taking other bidders' strategies as given. This best-response problem provides us insights into the tradeoffs bidders face, and it also forms the basis of our estimation method. We proceed by backward induction and analyze the Bellman formulation. Given the number of auctions K_T and bidder i's remaining budget w_{iT}, his objective is

$$\begin{aligned}
&\max_{\gamma,\tau} E\left[\sum_{k=1}^{K_T} \mathbb{1}\{C_{ikT} \leq \tau_{iT}\}\left(\mathbb{1}\{b^{\gamma}(X_{ikT}) > B_{-ikT}\}\left(X_{ikT} - b^{\gamma}(X_{ikT})\right) - C_{ikT}\right)\right] \\
&\qquad - E\left[\eta Q(w_{iT+1}) \mid \gamma, \tau\right] \\
&= \max_{\gamma,\tau} K_T F_C(\tau)\left(E\left[\Psi_T(b^{\gamma}(X) \mid K_T)(X - b^{\gamma}(X))\right] - E[C \mid C \leq \tau]\right) \\
&\qquad\qquad - E\left[\eta Q(w_{iT} - S_{iT}) \mid \gamma, \tau\right].
\end{aligned}$$

Denoting this maximized value as $V_T(K, w)$, the Bellman formulation of the objective of period $t = 1, \ldots, T-1$ is given by

$$\begin{aligned}
&\max_{\gamma,\tau} K_t F_C(\tau)\left(E\left[\Psi_t(b^{\gamma}(X) \mid K_t)(X - b^{\gamma}(X))\right] - E[C \mid C \leq \tau]\right) \\
&\qquad\qquad + E\left[EV_{t+1}(w_{it} - S_{it}) \mid \gamma, \tau\right],
\end{aligned} \tag{1.1}$$

where $EV_{t+1}(w) = E[V_{t+1}(K_{t+1}, w)]$ is the ex-ante value function in which the number of auctions is averaged out with distribution $F_K^{(t)}$. Given its similarity to the last period's objective, with a bit of abuse of notation, we denote $EV_{T+1}(w) = -\eta Q(w)$ for the rest of our analysis.

Now, we analyze the first-order necessary conditions for the bidding problem while assuming differentiability. The one with respect to the bid function parameter γ is given by

$$\begin{aligned}
&E\Bigg[\underbrace{\left(X - \frac{\Psi_t(b^{\gamma}(X) \mid K_t)}{\Psi_t'(b^{\gamma}(X) \mid K_t)} - b^{\gamma}(X)\right)}_{\text{Static FOC}} \Psi_t'(b^{\gamma}(X) \mid K_t)\nabla_{\gamma} b^{\gamma}(X)\Bigg] \\
&\qquad\qquad + \underbrace{\frac{1}{K_t F_C(\tau)}\nabla_{\gamma} E\left[EV_{t+1}(w_{it} - S_{it}) \mid \gamma, \tau\right]}_{\text{Dynamic Tradeoff}} = 0.
\end{aligned}$$

Meanwhile, the first-order condition with respect to the entry threshold τ is

$$\tau = \underbrace{E\left[\Psi_t(b^{\gamma}(X) \mid K_t)(X - b^{\gamma}(X))\right]}_{\text{Static Threshold}} + \underbrace{\frac{1}{K f_C(t)}\frac{\partial}{\partial \tau} E\left[EV_{t+1}(w_{it} - S_{it}) \mid \gamma, \tau\right]}_{\text{Dynamic Tradeoff}}.$$

The first-order conditions mainly consist of the static component and dynamic component. The optimality condition for the bid strategy contains an expression that is typically found in static first-price auction models (Guerre et al., 2000), along with an additional element resulting from the dynamic budget constraint.

In particular, we observe that the state variables (K_t, w_{it}) directly affect only the latter component. Similarly, the first-order condition for the participation strategy implies that the optimal threshold equals the static entry threshold determined by the expected payoff from an auction (Li and Zheng, 2009) and an additional dynamic component.

The dynamic component illustrates how entry and bids are influenced by dynamic tradeoffs. Increasing the likelihood of participation and bids impacts the continuation value by 1) making it more likely to win more auctions and 2) increasing the realized spending. Because having less remaining budget negatively affects his future surplus, the bidder must internalize this tradeoff and adjust his entry rate and bids further down from the statically optimal ones. Dynamic markdowns are a common feature in structural models that deal with sequential auctions, as demonstrated in prior research (Jofre-Bonet and Pesendorfer, 2003; Bodoh-Creed et al., 2021; Kong, 2021; Backus and Lewis, 2023). However, a notable departure in this study is the consideration that in earlier models, the option value was primarily influenced by whether one won an auction or not, while in our model, the option value is also impacted by the amount paid for a win.[12]

Now, we consider how the optimal strategies react to changes in the state variables under this model. When the number of auctions K_t changes, it impacts the dynamic component. Specifically, a higher K_t tends to result in larger spending from this period while keeping the strategies the same. Consequently, ceteris paribus, an increase in K_t introduces a force to make the strategies more conservative, consistent with the empirical pattern that bidders tend to be less aggressive when more auctions are present.

The current remaining budget w_{it} also appears in the first-order conditions only through the dynamic component. Ceteris paribus, decreasing w_{it} directly shifts down w_{it+1}. The ex-ante value function $EV_{t+1}(\cdot)$ typically exhibits a concave increasing pattern, as having a larger budget aids the bidder in securing more future opportunities, albeit at a diminishing rate. Consequently, the reduction in w_{it} amplifies the sensitivity of the continuation value to current-period spending, prompting the bidder to adopt more conservative strategies. The relationship between the optimal strategies and the remaining budget w_{it} highlights that the variability across bidders in their entry and bid decisions results from factors beyond random entry

[12] As demonstrated in our counterfactual analysis, this feature makes the price distribution relevant in shaping strategic behavior within our environment.

costs and valuations; it is also influenced by bidders' remaining budgets. This suggests that the size of each bidder's budget introduces heterogeneity in their behaviors.

The first-order necessary conditions highlight how this environment differs from the standard auction environment. If the budget constraint does not matter (i.e., if $\eta = 0$), the dynamic problem collapses into a series of static bidding problems, and the state variables K_t and w_{it} become irrelevant. However, with the budget constraint, the bidder needs to weigh the stage payoff and the option value from having more budget for the next period, and this tradeoff is influenced by the state variables.

Equilibrium

We now establish our solution concept by formalizing how a bidder's belief over competing bids is constructed. A pure strategy equilibrium of our model consists of time-dependent strategies $(\gamma_t(K,w), \tau_t(K,w))$ and bidders' beliefs regarding competing bids in each auction $\Psi_t(b \mid K)$ that satisfy the following conditions:

1. (Optimality) For each period t and state variables (K_t, w_{it}), $(\gamma_t(K,w), \tau_t(K,w))$ are a best response given the belief $\Psi_t(b \mid K)$, meaning they solve the problem specified in (1.1).

2. (Consistency)

$$\begin{aligned}\Psi_t(b \mid K) &= E\left[\prod_{j\neq i} \Pr(j \text{ does not enter, or } j \text{ enters and submits } B_{jt} \leq b \mid K, w_{jt})\right] \\ &= E\left[\prod_{j\neq i}\Big(1 - F_C(\tau_t(K,w_{jt})) + F_C(\tau_t(K,w_{jt}))F_X(b^{-1}(b \mid \gamma_t(K,w_{jt})))\Big)\right],\end{aligned}$$

where the distribution of state variables (remaining budgets) $(w_{jt})_{j\neq i}$ is determined by the initial budget distribution F_W, the distribution of the number of auctions $\{F_K^{(s)}\}_{s=1,\dots,t-1}$, the strategies employed by bidders $\{(\gamma_s(K,w), \tau_s(K,w))\}_{s=1,\dots,t-1}$, and the state transition rule.

The first condition requires that bidders are acting optimal given their belief, and the second condition ensures that the current belief is consistent with the optimal strategies they have employed in previous periods.[13] If bidder i had the knowledge about other bidders' remaining budgets $(w_{jt})_{j\neq i}$, his fully rational belief over competing bids would be the expression inside the expectation in the second condition.

[13]Our solution concept is similar to the large market equilibrium concepts used in prior works that study dynamic games (Hopenhayn, 1992; Krusell and Smith, 1998; Weintraub et al., 2008; Bodoh-Creed et al., 2021; Backus and Lewis, 2023).

However, given that bidders' initial budgets are private and they do not observe each other's spending, we require that their belief is averaged out with respect to the equilibrium state distribution of $(w_{jt})_{j \neq i}$. The rationale for this is that since a separate game occurs each day, bidders can rely on historical data regarding the intraday pattern of competition to formulate their participation and bidding strategies.[14]

In our analysis, we assume that a pure strategy equilibrium of the dynamic game exists and is unique given the model primitives, and we assume that each day in our dataset is independently sampled from this equilibrium.[15] In Appendix A.4, we present a computational algorithm for solving the dynamic game using our solution concept. This algorithm alternates between two steps: first, obtaining the belief Ψ_t by simulating the path of state variables using the given strategies, and second, obtaining the best response strategies given the belief through backward induction. Importantly, we find that our algorithm converges to the same equilibrium from various initial points, providing some support for our assumption of equilibrium existence and uniqueness. The formal proof of equilibrium existence and uniqueness is left for future research, as the conventional approach of backward induction does not apply to our environment. This is due to the fact that the equilibrium strategy in period t depends on the strategies in periods 1 through $t-1$, given that these strategies determine the belief regarding competing bids in period t.

Model Discussion

In our empirical application of the model to the online display advertising market, our analysis of dynamic strategic behavior abstracts away from two features of the market. First, we treat each demand-side platform (DSP) representing multiple advertisers as one budget-constrained bidder. In reality, each advertiser has a separate campaign budget, and DSPs need to make sure the constraint of each advertiser it represents is satisfied. However, modeling this relationship between DSPs and advertisers goes beyond the scope of this paper, and we leave it as a potential avenue for future research and extension of our model.

[14] One possible way to allow for firms learning each other's state within a game is to adapt the moment based Markov equilibrium proposed by Ifrach and Weintraub (2017) to our model. Their solution concept permits firms to track the state variables of a few dominant firms and form beliefs on the state variables of other firms conditional on their aggregate statistics. See also Fershtman and Pakes (2012) and Asker et al. (2020) for tractable ways to model players learning each other's state when there is persistent private information.

[15] Our assumption of equilibrium existence and uniqueness parallels the existing empirical works on non-standard auction games (Fox and Bajari, 2013; Kim et al., 2014; Saini, 2012; Gentry et al., 2020).

Second, while display ad auctions occur continuously in the real market, we discretize time and assume that multiple auctions happen simultaneously in each period. Hence, our model features a sequence of simultaneous first-price auctions. We make this assumption because of tractability, and often these auctions can even occur at the same time or within a very short time frame. This approximation aligns with the practices of DSPs, which also employ a discrete-time framework and assume that auctions within the same time interval happen simultaneously.

1.5 Estimation

Using our novel proprietary dataset of online display ad auctions, we estimate the primitives of the structural model of dynamic auctions with budget-constrained bidders. The model primitives are the distribution of the number of auctions $F_K^{(t)}$ for each period t, distribution of entry costs F_C, distribution of valuations F_X, budget constraint parameter η, and bidders' budgets $(w_i)_{i=1}^N$.

Each day in our dataset corresponds to one independent game in our structural model, and we consider hourly periods, giving us $t = 1, \ldots, T = 24$ periods in a day. Our dataset includes a total of $N = 33$ bidders. For each day d, we observe the number of auctions K_{td} for each hour t, as well as the spending per period S_{itd} and bids $(B_{iktd})_{k=1}^{K_{td}}$ for bidder i during each hour t. Note that we have $B_{iktd} = \emptyset$ if bidder i did not enter auction k in period t on day d. We suppress the day index when there is no confusion.

Our structural model, being a dynamic game, faces a common challenge in structural estimation that direct estimation requires solving for the equilibrium for every set of structural parameters. To circumvent this computational burden during the estimation process, we adopt a two-step approach, following the literature of structural estimation of dynamic games (Bajari et al., 2007; Aguirregabiria and Mira, 2007; Jofre-Bonet and Pesendorfer, 2003). Assuming that the market is in equilibrium under our solution concept, we estimate our structural model by leveraging bidders' best-response problem given their rivals' equilibrium behavior. In the first step, we estimate the time-dependent distribution of the number of auctions $F_K^{(t)}$, along with the reduced-form entry probability and bid distribution.[16] In the second step, we solve for bidders' entry and bid strategies $(\gamma_t(K, w), \tau_t(K, w))$ as best responses to rivals' estimated behaviors from the first stage and estimate the structural parame-

[16]These objects essentially serve as the conditional choice probability (CCP), using the terminology commonly employed in dynamic structural models. One key distinction from conventional approaches is that we have continuous actions instead of discrete actions.

ters through maximum likelihood estimation. This sequential approach allows us to avoid the need for solving the equilibrium and simulating the equilibrium state distribution during the estimation process.

First Stage

In the first stage, for each hour t, we estimate the distribution of the number of auctions $F_K^{(t)}$ and the belief on competition bids $\Psi_t(\cdot \mid K)$, which is the central equilibrium object in our structural model. We assume that the number of auctions comes from the negative binomial distribution[17] with time-specific parameters, capturing the daily pattern seen in Figure 1.3. We obtain the belief on competing bids $\Psi_t(\cdot \mid K)$ by deriving it from the entry probability and bid distribution estimated from the data. For tractability and numerical convenience, we use parametric forms to estimate them.[18] For the entry probability, we use logistic regression of entry outcome on the number of auctions K_t with time-period fixed effect to estimate it, meaning that entry probability is logistic$(\alpha_t^{entry} + \beta_K^{entry} K_{td})$. For the bid distribution, we assume $B_{iktd} \sim \text{LogNormal}(\alpha_t^{bid} + \beta_K^{bid} K_{td}, \sigma_t)$. Then, we derive the distribution on competing bids $\Psi_t(\cdot \mid K)$ by using these estimated objects.

Second Stage

In the second stage, we estimate the distribution of entry costs F_C, distribution of valuations F_X, budget constraint[19] parameter $\eta \in \mathbb{R}$, and bidders' budgets $(w_i)_{i=1}^N$. To facilitate our estimation, we introduce parametric assumptions for F_C and F_X. We assume $C \sim \text{TruncatedNormal}(\mu_C, \sigma_C)$ and $X \sim \text{LogNormal}(\mu_X, \sigma_X)$. Our set of structural parameters is denoted as $\theta = (\mu_C, \sigma_C, \mu_X, \sigma_X, \eta, (w_i)_{i=1}^N)$. We estimate these structural parameters through maximum likelihood estimation with an inner loop solving for bidders' entry and bid strategies $(\gamma_t(K, w), \tau_t(K, w))$ as best responses to the estimated belief on competing bids $\widehat{\Psi}_t(\cdot \mid K)$ from the first stage. We use the best-response problem to estimate the model as if it is a single-agent continuous-choice dynamic problem with a finite horizon. This avoids the computational burden of computing for equilibrium during estimation.

For each set of structural parameters θ, the inner-loop solves the best-response

[17]The negative binomial distribution is more flexible than the Poisson distribution since it allows the mean and variance to be different.

[18]Using parametric assumptions on the entry probability and bid distribution follows prior empirical works on one-shot auctions with entry (Athey et al., 2011; Krasnokutskaya and Seim, 2011).

[19]To review, when bidders violate their budget constraints, they suffer the penalty $\eta Q(w_{iT+1})$ where $Q(w) = \min(0, w)^2$ and w_{iT+1} is the remaining budget at the end of the game.

problem in (1.1) for $(\gamma_t(K, w; \theta), \tau_t(K, w; \theta))$ to evaluate our likelihood function. The best-response problem is solved as dynamic programming with a finite horizon via backward induction. The state variables are the number of auctions K_t and bidder's remaining budget w_{it}. Note that we do not observe budgets in our dataset, so we obtain w_{it} by setting $w_{it} = w_i - \sum_{s=1}^{t-1} S_{is}$ using the observed spending per period S_{it}. The central object of our dynamic programming is the ex-ante value function $EV_t(w) = E_{K_t}[V_t(K_t, w)]$. We approach this by solving the Bellman formulation in (1.1) over a grid of the state variables K and W for $t = T, \ldots, 1$.[20] We create the grid of the first state variable by taking random draws from the estimated distribution $\widehat{F}_K^{(t)}$ from the first stage; this grid is time-dependent. For the second state variable w, because this is a continuous variable, we create a grid by taking points in $[a, b]$ where $a < 0$ is a negative value that is unlikely to happen in equilibrium but nevertheless important for determining the shape of the value function, and $b > 0$ is a value above the maximum observed total spending. To evaluate the ex-ante value function outside of the grid and to obtain its derivative, we use a cubic spline with a monotonicity constraint for interpolation.

For each set of structural parameters θ, the above procedure provides us the entry thresholds $\tau_t(K, w; \theta)$ and bid strategies $b(\cdot \mid \gamma_t(K, w; \theta))$ for each K, w, and t.[21] Note that under the true θ, each observed bid B_{iktd} satisfies $B_{iktd} = b(X_{iktd} \mid \gamma_t(K, w; \theta))$ where X_{iktd} is the valuation, meaning $X_{iktd} = b^{-1}(B_{iktd} \mid \gamma_t(K, w; \theta))$. This relationship allows us to "back out" valuations by applying the inverse bid strategy on observed bids (Guerre et al., 2000). Given the best-response strategies, we derive the entry probabilities and density of the bid distribution to compute the log-likelihood. The entry probability is expressed by

$$\tilde{p}_t(K_t, w_{it}; \theta) = 1 - F_C(\tau_t(K_t, w_{it}; \theta); \mu_C, \sigma_C),$$

and the bid density is

$$\tilde{g}_t(B \mid K_t, w_{it}; \theta) = f_X(b^{-1}(B \mid \gamma_t(K_t, w_{it}; \theta)); \mu_X, \sigma_X)(b'(B \mid \gamma_t(K_t, w_{it}; \theta)))^{-1},$$

where the right-most term comes from the change of variables from valuations to bids. Finally, we can calculate the likelihood from the observed data by using these objects.

[20] Each optimization problem is solved by using the first-order conditions outlined in Section 1.4. We ensure global optimality by using multiple initial points.

[21] Note that in our empirical application, we set $b(x \mid \gamma) = \sum_{j=1}^{J} \gamma_j h_j(x)$ where h_j are a cubic spline basis functions with monotonicity constraints.

Identification

The identification of our structural parameters, especially the budget constraint parameter η and bidders' budgets $(w_i)_{i=1}^N$, relies on the exclusion restriction that bidders' valuations are independent of the state variables, which are the number of auctions and their remaining budgets. This assumption is required for disentangling the effect on bids from valuations and intertemporal budget constraints. For instance, when bids are low, we must determine whether this is due to low valuations or increased dynamic tradeoffs. The exclusion restriction enables us to identify the parameters relevant to budget constraints by using the correlation between bids and the state variables, which impact the dynamic tradeoffs.

For illustration, our assumption implies the following conditional moment conditions:

$$E[b^{-1}(B \mid \gamma_t(K_t, w_{it}; \theta)) - E[X_{ikt} \mid \theta] \mid Z_{it}] = 0,$$

where Z_{it} is remaining budget w_{it} or number of auctions K_t. This essentially means that the valuations backed out via the inverse bid function should not be correlated with our instruments. Such an exclusion restriction with other instrumental variables has been used in the empirical auction literature to test a model (Haile et al., 2003) or identify structural parameters (Guerre et al., 2009; Gentry et al., 2020).

The exogeneity assumption is plausible for our market environment as typically demand-side platforms (bidders) and advertisers compute their valuations for impressions based on a combination of the probability of clicking/making a sale and their value of such events, and this probability is computed based on the contextual and behavioral data of that user alone. Hence, the short-run supply level of impression (the number of auctions) and their current campaign budgets should not directly influence how they value advertising opportunities.

Our structural framework can accommodate potential sources of unobserved heterogeneity that might interfere with our identification strategy above. First, valuations could be correlated with the number of auctions through a time-varying unobserved heterogeneity. For instance, the average user browsing the internet during the daytime could be different from the average online user at night, and the supply levels are different across these time periods. One possible remedy for this issue is to allow time-dependent valuation distribution $F_X^{(t)}$, analogous to introducing time fixed effects in standard econometric models. The second potential confounder is that valuations could be correlated with bidders' budgets through a bidder unobserved heterogeneity. Advertisers with larger budgets may also happen to have higher valuations for impressions. We can alleviate this issue by classifying bidders into groups and estimating structural parameters for each group separately. For both of these situations, identification is possible by using the variation across different days (or games).

1.6 Estimation Results

First-Stage Estimates

First, we present the parameter estimates for the distribution of the number of auctions $F_K^{(t)}$ and the estimated reduced-form entry probability and bid distribution. In Figure 1.6, we illustrate the probability mass function of $\widehat{F}_K^{(t)}$, which reflects the daily supply pattern that is also shown in Figure 1.3.[22] The estimated distribution also illustrates how the variance of K_t changes over time.

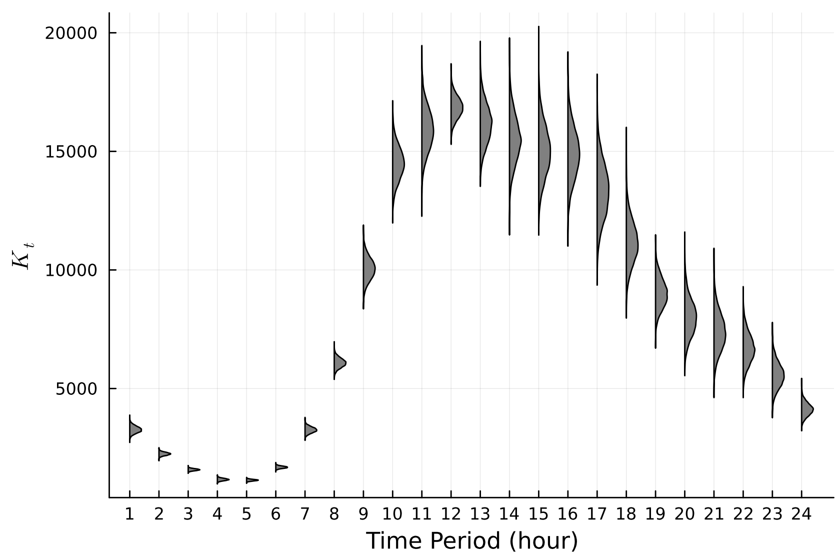


Figure 1.6: The probability mass function of the estimated distribution of the number of auctions $\widehat{F}_K^{(t)}$ for each time period t.

Table 1.3 shows the estimated parameters of the reduced-form entry probability and bid distribution. Notably, in line with the second stylized fact reported in Section 1.3, the estimated coefficient on K_t indicates that the entry probability and bid distribution are negatively impacted by an increase in the number of auctions, all else being equal. Furthermore, consistent with the first stylized fact, the estimated intercepts illustrate a declining trend in the entry probability and bids.

[22]The table containing the estimated parameters are in Appendix A.2.

	Entry Probability	Bid Distribution	
β_K	-3.627e-5 (5.485e-7)	-3.417e-5 (3.852e-7)	
	α_t^{entry}	α_t^{bid}	σ_t
$t = 1$	-1.354 (0.003)	-0.365 (0.003)	0.900 (0.0019)
$t = 2$	-1.278 (0.004)	-0.325 (0.003)	0.893 (0.0022)
$t = 3$	-1.263 (0.004)	-0.291 (0.003)	0.902 (0.0027)
$t = 4$	-1.283 (0.005)	-0.224 (0.004)	0.947 (0.0033)
$t = 5$	-1.246 (0.005)	-0.196 (0.004)	0.944 (0.0033)
$t = 6$	-1.158 (0.004)	-0.128 (0.003)	0.950 (0.0026)
$t = 7$	-1.124 (0.003)	-0.146 (0.002)	0.915 (0.0018)
$t = 8$	-1.117 (0.004)	-0.226 (0.002)	0.872 (0.0012)
$t = 9$	-1.156 (0.005)	-0.287 (0.004)	0.838 (0.0009)
$t = 10$	-1.239 (0.008)	-0.347 (0.005)	0.823 (0.0008)
$t = 11$	-1.294 (0.008)	-0.383 (0.006)	0.808 (0.0008)
$t = 12$	-1.322 (0.009)	-0.410 (0.006)	0.800 (0.0007)
$t = 13$	-1.354 (0.009)	-0.441 (0.006)	0.790 (0.0008)
$t = 14$	-1.363 (0.008)	-0.453 (0.006)	0.785 (0.0008)
$t = 15$	-1.357 (0.008)	-0.452 (0.005)	0.790 (0.0008)
$t = 16$	-1.367 (0.008)	-0.470 (0.005)	0.782 (0.0008)
$t = 17$	-1.376 (0.007)	-0.465 (0.005)	0.782 (0.0008)
$t = 18$	-1.365 (0.006)	-0.458 (0.004)	0.785 (0.0009)
$t = 19$	-1.380 (0.005)	-0.455 (0.003)	0.793 (0.0010)
$t = 20$	-1.371 (0.005)	-0.452 (0.003)	0.793 (0.0011)
$t = 21$	-1.335 (0.004)	-0.387 (0.003)	0.831 (0.0012)
$t = 22$	-1.340 (0.004)	-0.431 (0.003)	0.825 (0.0012)
$t = 23$	-1.347 (0.004)	-0.450 (0.002)	0.830 (0.0014)
$t = 24$	-1.359 (0.003)	-0.464 (0.002)	0.837 (0.0016)

Table 1.3: Reduced-form estimates of entry probability logistic$(\alpha_t^{entry} + \beta_K^{entry}(K_{td} - \overline{K}_t))$ and bid distribution LogNormal$(\alpha_t^{bid} + \beta_K^{bid}(K_{td} - \overline{K}_t), \sigma_t)$ where $\overline{K}_t$ is the sample average per period. We de-mean K_{td} for an illustration purpose.

Second-Stage Estimates

We present the estimates for the structural parameters $\theta = (\mu_C, \sigma_C, \mu_X, \sigma_X, \eta, (w_i)_{i=1}^N)$ from the second stage. Table 1.4 presents the estimates of $(\mu_C, \sigma_C, \mu_X, \sigma_X, \eta)$, along with their estimated standard errors.

First, note that the estimate for η is positive and statistically significantly different from 0. Since bidders' dynamic bidding problem collapses to a series of static bidding problems if $\eta = 0$, this confirms that bidders care about the budget constraint and hence act dynamically. The estimated model reveals that, within the dataset,

Parameters	Estimate	SE
μ_C	-11.3776	0.0091
σ_C	7.2533	0.0062
μ_X	0.9046	0.0007
σ_X	1.0950	0.0006
η	0.6457	0.0084

Table 1.4: Estimates of structural parameters. The standard errors are computed using the White (sandwich) estimator using the numerical Hessian and Jacobian.

the typical bidder exceeds their budget approximately 26% of the time. When such overspending occurs, it amounts to an average of around 8% of their budgets, indicating that these bidding agents occasionally exceed the budgets set by their clients (advertisers) in pursuit of maximizing their payoffs while avoiding excessive violations.

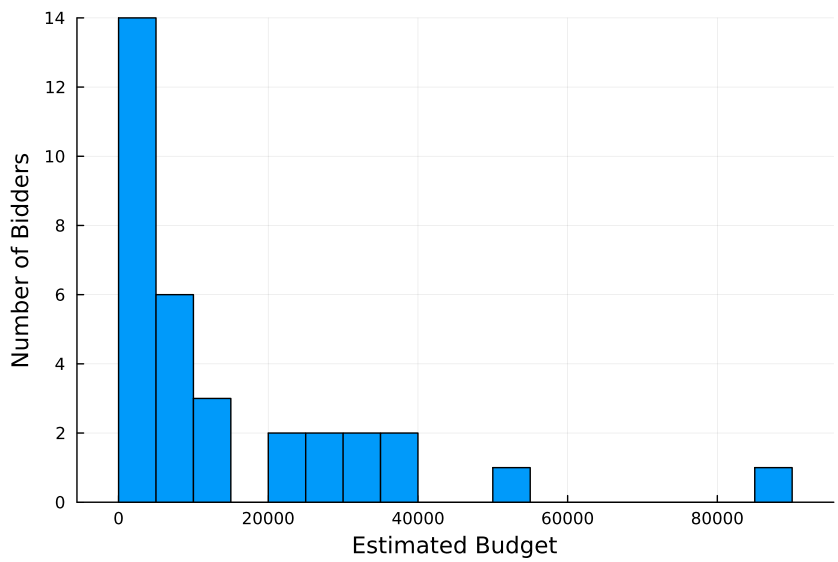


Figure 1.7: Histogram of the estimated budgets $(\widehat{w}_i)_{i=1}^N$.

Figure 1.7 shows the histogram of the estimated daily budgets $(\widehat{w}_i)_{i=1}^N$ of bidders. The budget distribution is significantly skewed, and it reflects how the online advertising market is concentrated. In particular, the distribution shows that there are a few large players and many smaller players. The former type includes large tech companies like Google and Amazon.

Dynamic Incentive

Given the estimated structural model, we can analyze the magnitude of the dynamic incentives created by the budget constraint. First, we look at the markdown (valuation minus bid) obtained by the model. Averaging across time periods, bidders, and days, we find that the markdown is 3.76, which is 83.5% of the expected valuation (4.5). This markdown reflects both the fact that bidders face first-price auctions and dynamic incentives. To decompose these two different incentives, we simulate counterfactual static bids while taking the probability of winning as in the data but removing the dynamics created by the budget constraint. We find that the counterfactual static markdown is 2.67 on average, and it is 59.4% of the expected valuation. This highlights that facing the first-price format for each auction leads bidders to shade their bids by 59.4% from valuations, and the dynamic budget constraint leads them to shade further by 24.2% on average. This demonstrates that dynamic incentives in this market are significant for the bidders. Figure 1.8 shows the relationship between the daily budget and the average markdown. We see that the heterogeneity in budgets, in turn, leads to heterogeneity in how aggressive bidders are. We find in our counterfactual analysis (Section 1.7) that this competitive variation has a substantial welfare implication.

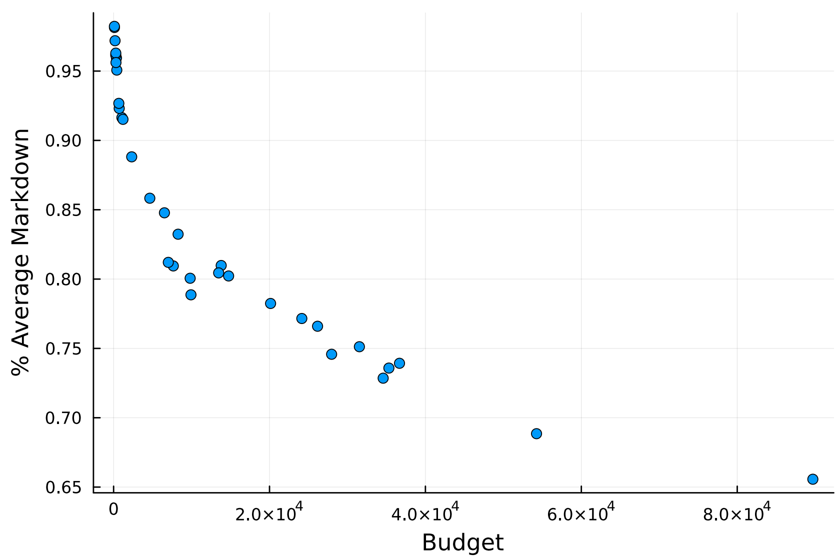


Figure 1.8: Budget vs %Average Markdown. Each point represents a bidder.

Our structural model also endogenizes entry, so it also allows us to quantify the effect of budget constraints on bidders' entry decisions. The static simulation above also provides us the counterfactual static entry probabilities, purely coming from stochastic entry costs. The average entry probability fitted by the model is 19.4%, and the average static entry probability is 45.1%, which again illustrates the importance of capturing the dynamic budget constraint to analyze bidders' behavior in this market. Figure 1.9 shows the relationship between the daily budget and the average entry probability, and again it shows that heterogeneity in budgets leads to heterogeneity in entry behaviors.

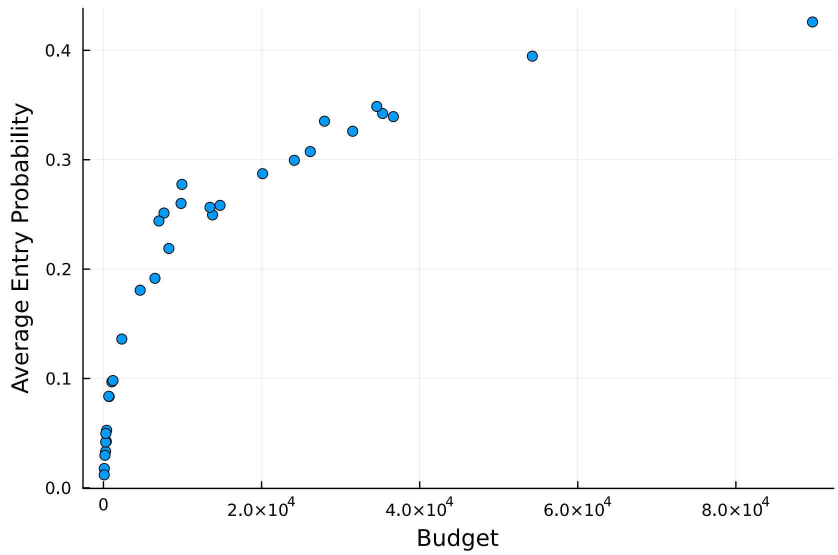


Figure 1.9: Budget vs Average Entry Probability. Each point represents a bidder.

Finally, we utilize the estimated structural model to decompose the dynamic incentive into two components: one arising from the diminishing budget and the other from approaching the terminal period. In the data, we observe that the average bid decreases over time due to diminishing budgets. However, as time progresses, bidders have fewer opportunities remaining, which should in principle make them less constrained. Therefore, what we observe in the data results from the interplay of these two effects: the diminishing budget effect and the diminishing remaining opportunities effect. In Figure 1.10, we illustrate this by considering a bidder with a median budget (approximately $8000) and comparing their average bid as fitted to the data with the model-predicted average bid when their remaining budget is held constant, thereby isolating the effect of having fewer opportunities as time elapses. We observe that the bidder becomes more aggressive with a constant budget as time progresses. However, the diminishing budget effect ultimately dominates, leading to a declining bid path, as indicated by the fitted model.

1.7 Counterfactuals

Using the estimated structural model, we simulate a counterfactual motivated by an institutional change that occurred several years ago. Although the current online ad market primarily uses the first-price auction mechanism, ad exchanges (auctioneers mediating publishers and advertisers) used the second-price format until around

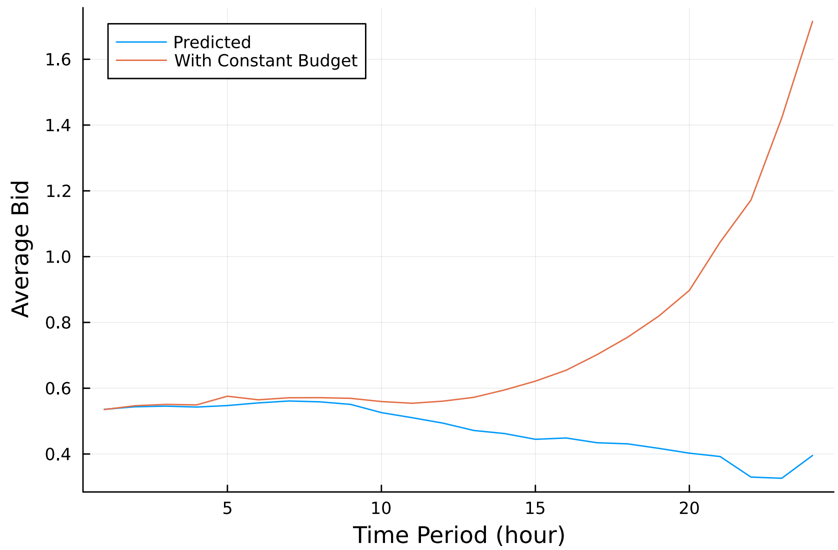


Figure 1.10: The median-budget bidder's predicted mean bid trajectory compared to the predicted path when keeping their remaining budget constant.

2018. The shift from the second price to the first price was spurred by an industry-wise outcry that ad exchanges are charging something other than the second-price even though claiming to be running the second price auction. Hence, this industry-wide credibility loss of market makers led participants to demand the first-price auction for its transparency over what winners pay. Motivated by this shift, we simulate the second-price auction format using the estimated structural model as a counterfactual scenario to analyze the revenue and welfare consequences.

Although the theoretical auction literature has established that the first-price auction and the second-price auction provide the same revenue and welfare for the standard auction environment, it is ambiguous whether this holds for our environment. Alcobendas and Zeithammer (2023) and Goke et al. (2022) provide event-based analyses of this transition, and a prominent finding in their research is that bidding agents required an extended period, often spanning several months, to adapt their bidding strategies for the first-price auction format. As market conditions can drastically change in such a time span, it highlights that event-based approaches may be inadequate to provide an equilibrium analysis of the comparison between the two formats. Our structural framework provides a way to compare the long-run equilibrium outcomes from the first-price and second-price auctions.

Using the best-response iteration algorithm described in Appendix A.4, we solve both the benchmark scenario with the first-price auction (FPA) and the counterfactual scenario with the second-price auction (SPA) as continuous-action dynamic games with a finite-time horizon. The best-response formulation for the second-price format is given in Appendix A.3.

Auction Format	First Price	Second Price
Price Average	\$2.364	\$2.362
Price Variance	1.1246	3.565
Expected Total Revenue	\$480,427.33	\$480,073.49
Expected Total Bidder Surplus	\$1,191,000	\$1,185,000

Table 1.5: Aggregate statistics of the simulated results under the first-price auction (status quo) and the second-price auction.

	First Price	Second Price	%(FPA - SPA)
Large Bidders	\$191,362.15	\$197,345.68	-3.03%
Medium Bidders	\$473,791.65	\$467,448.36	1.36%
Small Bidders	\$297,439.98	\$291,925.99	1.89%

Table 1.6: Total bidder surplus of each type of bidders.

Table 1.5 shows aggregate statistics of the two auction mechanisms. It shows that the total revenue and total bidder surplus are slightly better under the first-price format on average. We find that expected daily (total) revenue and expected total bidder surplus are slightly higher under the first-price format than the second-price format. They are both about 0.1% higher under the first-price format.

We find a more substantial difference when we analyze the difference in the welfare distribution among bidders. First, we classify bidders based on their estimated budgets. Based on the distribution of budgets in Figure 1.7, we classify two bidders with budgets ranging from \$50,000 to \$90,000 as 'Large,' eleven bidders with budgets between \$10,000 and \$50,000 as 'Medium,' and twenty bidders with budgets below \$10,000 as 'Small.' Table 1.6 shows the expected total utility obtained by each type of bidders under the two mechanisms, and it shows that the top two bidders with the largest budgets are better off under SPA while other bidders with smaller budgets are worse off. The combined welfare of the two large bidders is 3% higher in SPA than in FPA, and the one for the other bidders is 1% lower in SPA. This suggests FPA has an interesting property in this environment that, when compared to SPA, it redistributes welfare from bidders with large budgets to those with smaller budgets. This suggests that the transition from SPA to FPA in the online display ad market was a welfare improvement event for smaller players, in addition to the fact that they can enjoy the transparency of FPA.

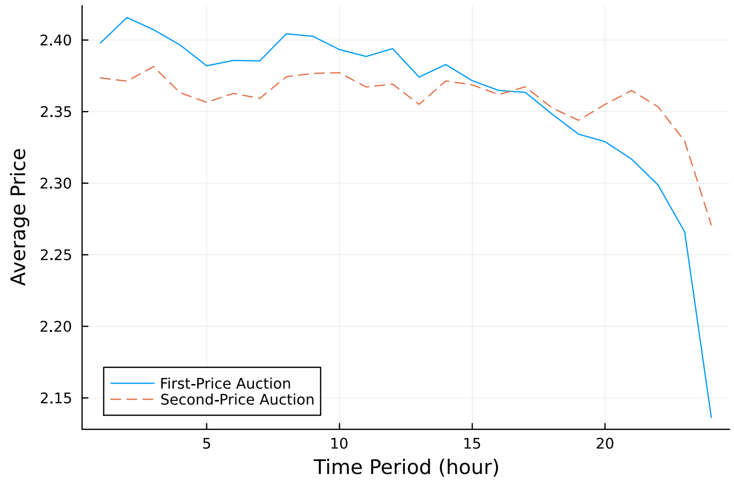


Figure 1.11: Average price per period.

Now we compare the dynamic outcomes of FPA and SPA. Figure 1.11 shows the average price per period. Note that the price is the highest bid under FPA and the second-highest bid under SPA. The average price under FPA, which is the mechanism used in the data, shows a declining pattern as we see in our descriptive results (Figure 1.3). As explained before, this is coming from bidders becoming conservative from decreasing budgets. Although the price path from SPA also shows a declining pattern, there is some distinctive difference between them. The figure shows that the average price from FPA is systematically higher than SPA until around 3 PM, and then the relationship switches. This suggests that FPA generates more revenue until 3 PM, and then SPA generates more revenue after that.

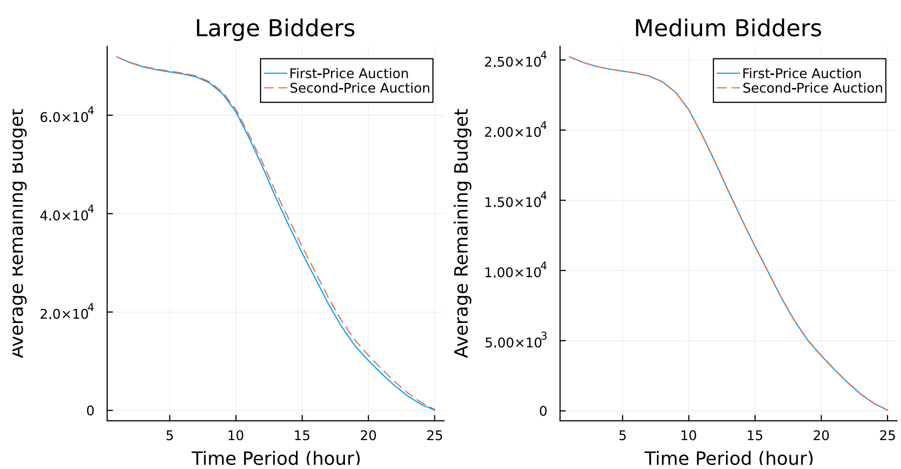


Figure 1.12: Average remaining budget per period for Large and Medium bidders.

Analyzing the spending path of players reveals the critical difference that is driving the dynamic difference. Figure 1.12 shows the average remaining budget of Large and Medium bidders for each period. The spending path of Large bidders shows that in the afternoon, their remaining budgets tend to be lower under FPA. Meanwhile, the spending path of Medium bidders is relatively similar across the two auction formats. This suggests that the price difference between FPA and SPA after 3 PM is primarily driven by the large bidders having tighter budgets in the afternoon under FPA.

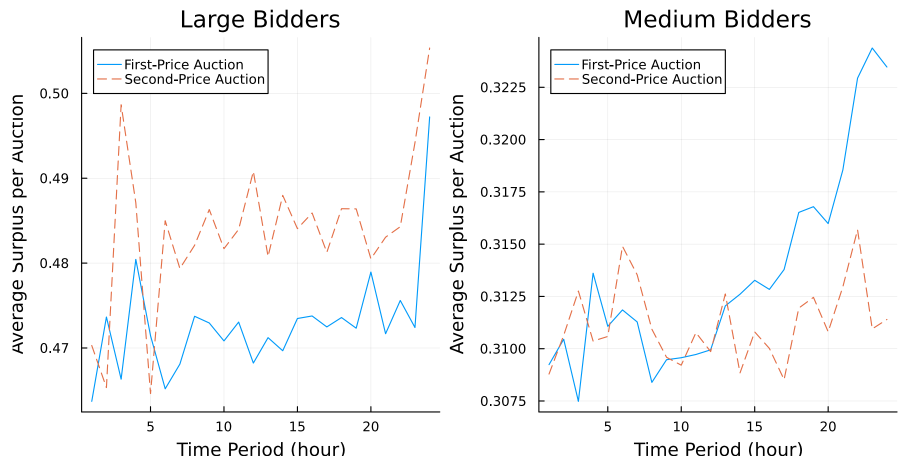


Figure 1.13: Average surplus per period for Large and Medium bidders.

The difference in the spending speed of the large bidders across the two formats has a significant welfare consequence for smaller bidders. Figures 1.13 show the time series of average surplus per auction for Large and Medium bidders. They show that Medium bidders experience a larger surplus under FPA after around 3 PM, when FPA becomes less competitive than SPA as shown in 1.11. Meanwhile, we do not see such a pattern for Large bidders. This dynamic difference suggests that the contrast in the welfare distribution is driven by smaller bidders enjoying less competition in the afternoon from the large bidders under FPA.

What is driving the large bidders to spend more rapidly under FPA? We analyze the entry and bidding behavior to understand this. First, in Figure 1.14, we find that FPA has a higher expected spending per auction conditional on entering most of the time. This demonstrates that players bid more aggressively under FPA. However, in Figure 1.15, we also find that while Large bidders have similar entry patterns across the two formats, Medium and Small bidders enter auctions at lower rates under FPA. Since the number of entrants is not public, entry rates affect the probability of winning an auction. Hence, borrowing the terminology of Li and Zheng (2009), we find that for Large bidders' spending, the competition effect coming from all

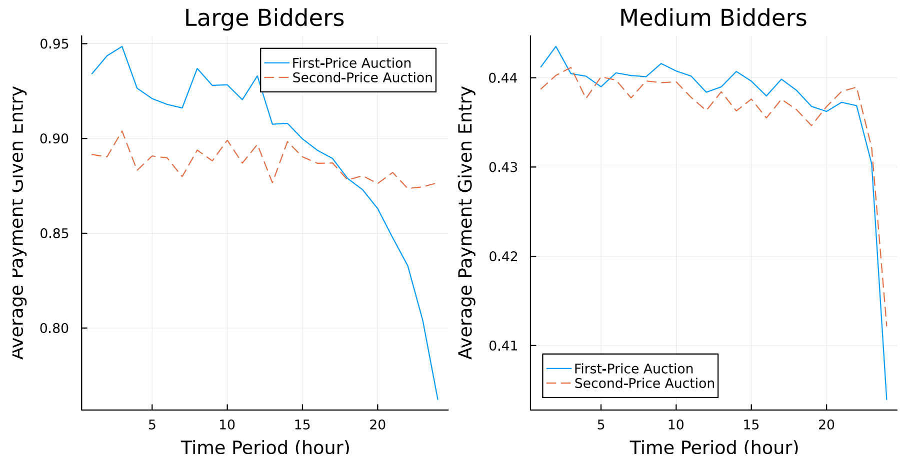


Figure 1.14: Average spending from entered auctions for Large and Medium bidders.

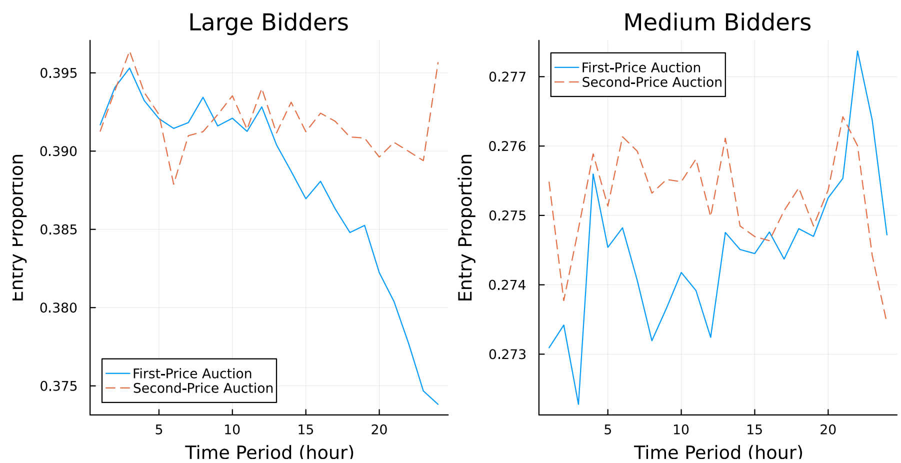


Figure 1.15: Average proportion of auctions entered for Large and Medium bidders.

bidders submitting more aggressive bids under FPA dominates the entry effect from Medium and Small bidders entering less frequently.

Finally, we analyze why bidders submit more aggressive bids when they enter auctions under FPA. In the standard auction environment, one key difference between the payment for FPA and SPA is that the variance is higher for SPA.[23] Similarly, Table 1.5 shows that this is also the case in our simulation. This difference can be crucial for bidders in our environment since it affects their ability to control their spending dynamically. In particular, there can be more "accidents" in which they end up paying more than they expected. Looking back at the bidder's bidding problem under FPA in (1.1) and SPA in (A.1), we see that the spending enters nonlinearly in

[23]In fact, the revenue from the second-price auction is a mean-preserving spread of the one from the first-price auction Krishna (2009).

the objective function through the continuation value,

$$E\left[EV_{t+1}(w_{it} - S_{it}) \mid \gamma, \tau\right],$$

where S_{it} is the total spending from the current period, given bid strategy γ and entry threshold τ. Note that $EV_t(\cdot)$ exhibits a concave increasing pattern in our estimated structural model because having a larger budget aids the bidder in securing more future opportunities, albeit at a diminishing rate. Intuitively, the continuation value makes bidders effectively risk averse; ceteris paribus, they dislike having a higher variance in their payment because of its concavity in the spending from the current period. Hence, bidders are more conservative under the second-price auction, resulting from the willingness to sacrifice some gain with a reduction in the variance.

The finance literature has extensively documented that financial constraints tend to induce risk aversion in firms (Froot et al., 1993; Opler et al., 1999). In particular, theoretical studies by Milne and Robertson (1996), Holt (2003), and Rochet and Villeneuve (2005) investigate the dynamic problem of a financially-constrained firm determining dividends and investment policies, and they consistently find that the concavity in the value function with respect to the cash holding leads the firm to exhibit risk aversion, which is in line with the findings in our model.

1.8 Conclusion

When price discovery is necessary for time-sensitive goods, it is common practice to conduct an auction for each item sequentially. These dynamic settings may lead to behaviors distinct from static environments and affect the revenue and welfare outcomes of various auction formats. This paper investigates how intertemporal budget constraints affect competition in the online advertising market. Furthermore, we examine how bidders with varying budgets face disparate welfare outcomes under different auction mechanisms.

We develop a finite-horizon dynamic game between bidders with heterogeneous budgets facing numerous auctions in each period. We estimate the model using a proprietary dataset of online ad auctions from Yahoo. Our estimation results show that bidders indeed exhibit behavior consistent with dynamic budget constraints, and there is a significant disparity in daily budgets among players, contributing to the heterogeneity observed in participation and bidding behaviors.

To gain insights into the strategic implications of dynamic incentives arising from intertemporal constraints, we conduct two counterfactual exercises. First, we simulate bidders' counterfactual entry and bidding behaviors if they were unconstrained. This exercise reveals that, on average, approximately 30% of the markdown can be attributed to dynamic constraints, which also lead to a reduction in participation probability by around 25 percentage points.

As our second counterfactual exercise, we compare first-price (the status quo) and second-price auction outcomes. Although both auction formats yield equivalent revenue and welfare outcomes in the standard auction environment with symmetric bidders, we discover that dynamics and heterogeneous budgets lead to substantial welfare differences between them. Intermediate and smaller budget bidders fare better in the first-price format due to reduced price variance, which allows for more aggressive bidding. This heightened competition prompts larger bidders to spend quickly, leading to diminished competition in later periods and ultimately benefiting smaller bidders overall. This highlights that even a seemingly simple mechanism choice can have competitive implications in such a dynamic environment.

The main contribution of this paper is to empirically analyze how budget constraints shape competition in auctions when held sequentially. Our approach involves introducing a novel structural framework for analyzing such an environment. The relevance of our findings and framework extends beyond the online advertising market. Sequential auctions are prevalent in various settings, encompassing online retail platforms, financial markets, and energy markets, where buyers often face financial constraints. Traditionally, these scenarios have been examined by treating individual auctions as isolated static events. However, our work reveals how dynamic constraints can interlink these sequential auctions, introducing nuanced insights into competition dynamics.

References

Agarwal, Deepak et al. (Aug. 2014). "Budget Pacing for Targeted Online Advertisements at LinkedIn". In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York New York USA: ACM, pp. 1613–1619. ISBN: 978-1-4503-2956-9. DOI: 10.1145/2623330.2623366.

Aguirregabiria, Victor and Pedro Mira (2007). "Sequential Estimation of Dynamic Discrete Games". In: *Econometrica* 75.1, pp. 1–53. ISSN: 1468-0262. DOI: 10.1111/j.1468-0262.2007.00731.x.

Alcobendas, Miguel and Robert Zeithammer (2023). "Slim Shading in Ad Auctions: Adjustment of Bidding Strategies to First-Price Rules". In.

Alcobendas, Miguel et al. (Oct. 2023). *The Impact of Privacy Protection on Online Advertising Markets*. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.3782889.

Asker, John et al. (Sept. 2020). "A Computational Framework for Analyzing Dynamic Auctions: The Market Impact of Information Sharing". In: *The RAND Journal of Economics* 51.3, pp. 805–839. ISSN: 0741-6261, 1756-2171. DOI: 10.1111/1756-2171.12341.

Athey, Susan, Jonathan Levin, and Enrique Seira (Feb. 2011). "Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions*". In: *The Quarterly Journal of Economics* 126.1, pp. 207–257. ISSN: 0033-5533, 1531-4650. DOI: 10.1093/qje/qjq001.

Athey, Susan and Denis Nekipelov (2011). "A Structural Model of Sponsored Search Advertising Auctions". In: p. 72.

Backus, Matthew and Gregory Lewis (Mar. 2023). *Dynamic Demand Estimation in Auction Markets*. Tech. rep., w22375. DOI: 10.3386/w22375.

Bajari, Patrick, C. Lanier Benkard, and Jonathan Levin (2007). "Estimating Dynamic Models of Imperfect Competition". In: *Econometrica* 75.5, pp. 1331–1370. ISSN: 1468-0262. DOI: 10.1111/j.1468-0262.2007.00796.x.

Balseiro, Santiago et al. (2020). "Budget Management Strategies in Repeated Auctions". In: p. 110.

Balseiro, Santiago R., Omar Besbes, and Gabriel Y. Weintraub (2015). "Repeated Auctions with Budgets in Ad Exchanges: Approximations and Design". In: *Management Science* 61.4, pp. 864–884. ISSN: 0025-1909.

Benoît, Jean-Pierre and Vijay Krishna (2001). "Multiple-Object Auctions with Budget Constrained Bidders". In: *The Review of Economic Studies* 68.1, pp. 155–179. ISSN: 0034-6527.

Bernhardt, Dan and David Scoones (1994). "A Note on Sequential Auctions". In: *The American Economic Review* 84.3, pp. 653–657. ISSN: 0002-8282.

Bodoh-Creed, Aaron L, Jörn Boehnke, and Brent Hickman (Feb. 2021). "How Efficient Are Decentralized Auction Platforms?" In: *The Review of Economic Studies* 88.1, pp. 91–125. ISSN: 0034-6527, 1467-937X. DOI: 10.1093/restud/rdaa017.

Celis, L. Elisa et al. (Dec. 2014). "Buy-It-Now or Take-a-Chance: Price Discrimination Through Randomized Auctions". In: *Management Science* 60.12, pp. 2927–2948. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.2014.2009.

Choi, Hana et al. (June 2020). "Online Display Advertising Markets: A Literature Review and Future Directions". In: *Information Systems Research* 31.2, pp. 556–575. ISSN: 1047-7047, 1526-5536. DOI: 10.1287/isre.2019.0902.

Conitzer, Vincent et al. (Mar. 2022a). "Multiplicative Pacing Equilibria in Auction Markets". In: *Operations Research* 70.2, pp. 963–989. ISSN: 0030-364X, 1526-5463. DOI: 10.1287/opre.2021.2167.

Conitzer, Vincent et al. (Dec. 2022b). "Pacing Equilibrium in First Price Auction Markets". In: *Management Science* 68.12, pp. 8515–8535. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.2022.4310.

Decarolis, Francesco and Gabriele Rovigatti (Oct. 2021). "From Mad Men to Maths Men: Concentration and Buyer Power in Online Advertising". In: *American Economic Review* 111.10, pp. 3299–3327. ISSN: 0002-8282. DOI: 10.1257/aer.20190811.

Engelbrecht-Wiggans, Richard (Jan. 1994). "Sequential Auctions of Stochastically Equivalent Objects". In: *Economics Letters* 44.1-2, pp. 87–90. ISSN: 01651765. DOI: 10.1016/0165-1765(93)00327-K.

Fershtman, Chaim and Ariel Pakes (Nov. 2012). "Dynamic Games with Asymmetric Information: A Framework for Empirical Work*". In: *The Quarterly Journal of Economics* 127.4, pp. 1611–1661. ISSN: 1531-4650, 0033-5533. DOI: 10.1093/qje/qjs025.

Fox, Jeremy T and Patrick Bajari (Feb. 2013). "Measuring the Efficiency of an FCC Spectrum Auction". In: *American Economic Journal: Microeconomics* 5.1, pp. 100–146. ISSN: 1945-7669, 1945-7685. DOI: 10.1257/mic.5.1.100.

Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein (Dec. 1993). "Risk Management: Coordinating Corporate Investment and Financing Policies". In: *The Journal of Finance* 48.5, pp. 1629–1658. ISSN: 0022-1082, 1540-6261. DOI: 10.1111/j.1540-6261.1993.tb05123.x.

Gaitonde, Jason et al. (May 2022). *Budget Pacing in Repeated Auctions: Regret and Efficiency without Convergence*.

Gale, Ian L. and Donald B. Hausch (Nov. 1994). "Bottom-Fishing and Declining Prices in Sequential Auctions". In: *Games and Economic Behavior* 7.3, pp. 318–331. ISSN: 0899-8256. DOI: 10.1006/game.1994.1054.

Gentry, Matthew, Tatiana Komarova, and Pasquale Schiraldi (2020). "Preferences and Performance in Simultaneous First-Price Auctions: A Structural Analysis". In: p. 76.

Gentry, Matthew L. et al. (2018). "Structural Econometrics of Auctions: A Review". In: *Foundations and Trends® in Econometrics* 9.2-4, pp. 79–302. ISSN: 1551-3076, 1551-3084. DOI: 10.1561/0800000031.

Ghosh, Gagan and Heng Liu (Jan. 2019). "Sequential Second-Price Auctions with Private Budgets". In: *Games and Economic Behavior* 113, pp. 611–632. ISSN: 08998256. DOI: 10.1016/j.geb.2018.11.005.

Goke, Shumpei et al. (Jan. 2022). *Bidders' Responses to Auction Format Change in Internet Display Advertising Auctions*.

Groeger, Joachim R. (2014). "A STUDY OF PARTICIPATION IN DYNAMIC AUCTIONS". In: *International Economic Review* 55.4, pp. 1129–1154. ISSN: 0020-6598.

Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong (2000). "Optimal Nonparametric Estimation of First-Price Auctions". In: *Econometrica* 68.3, pp. 525–574.

– (2009). "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions". In: *Econometrica* 77.4, pp. 1193–1227. ISSN: 1468-0262. DOI: 10.3982/ECTA7028.

Haile, Philip A, Han Hong, and Matthew Shum (2003). "Nonparametric Tests for Common Values In First-Price Sealed-Bid Auctions". In: *National Bureau of Economic Research*, p. 57.

Hendricks, Kenneth and Alan Sorensen (Sept. 2018). *Dynamics and Efficiency in Decentralized Online Auction Markets*. Tech. rep. w25002. Cambridge, MA: National Bureau of Economic Research, w25002. DOI: 10.3386/w25002.

Hickman, Brent R., Timothy P. Hubbard, and Yiğit Sağlam (Jan. 2012). "Structural Econometric Methods in Auctions: A Guide to the Literature". In: *Journal of Econometric Methods* 1.1. ISSN: 2156-6674. DOI: 10.1515/2156-6674.1019.

Holt, Richard W. P. (Jan. 2003). "Investment and Dividends under Irreversibility and Financial Constraints". In: *Journal of Economic Dynamics and Control* 27.3, pp. 467–502. ISSN: 0165-1889. DOI: 10.1016/S0165-1889(01)00057-4.

Hopenhayn, Hugo A. (Sept. 1992). "Entry, Exit, and Firm Dynamics in Long Run Equilibrium". In: *Econometrica* 60.5, p. 1127. ISSN: 00129682. DOI: 10.2307/2951541.

Ifrach, Bar and Gabriel Y. Weintraub (July 2017). "A Framework for Dynamic Oligopoly in Concentrated Industries". In: *The Review of Economic Studies* 84.3, pp. 1106–1150. ISSN: 0034-6527. DOI: 10.1093/restud/rdw047.

Jofre-Bonet, Mireia and Martin Pesendorfer (2003). "Estimation of a Dynamic Auction Game". In: *Econometrica* 71.5, pp. 1443–1489. ISSN: 1468-0262. DOI: 10.1111/1468-0262.00455.

Kim, Sang Won, Marcelo Olivares, and Gabriel Y. Weintraub (2014). "Measuring the Performance of Large-Scale Combinatorial Auctions: A Structural Estimation Approach". In: *Management Science* 60.5, pp. 1180–1201. ISSN: 0025-1909.

Kong, Yunmi (Jan. 2021). "Sequential Auctions with Synergy and Affiliation across Auctions". In: *Journal of Political Economy* 129.1, pp. 148–181. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/711402.

Krasnokutskaya, Elena and Katja Seim (Oct. 2011). "Bid Preference Programs and Participation in Highway Procurement Auctions". In: *American Economic Review* 101.6, pp. 2653–2686. ISSN: 0002-8282. DOI: 10.1257/aer.101.6.2653.

Krishna, Vijay (2009). *Auction Theory*. 2nd ed. Burlington, MA: Academic Press/Elsevier. ISBN: 978-0-12-374507-1.

Krusell, Per and Anthony A. Smith Jr. (Oct. 1998). “Income and Wealth Heterogeneity in the Macroeconomy”. In: *Journal of Political Economy* 106.5, pp. 867–896. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/250034.

Li, Tong and Xiaoyong Zheng (Oct. 2009). “Entry and Competition Effects in First-Price Auctions: Theory and Evidence from Procurement Auctions”. In: *Review of Economic Studies* 76.4, pp. 1397–1429. ISSN: 00346527, 1467937X. DOI: 10.1111/j.1467-937X.2009.00558.x.

Milgrom, Paul Robert (2004). *Putting Auction Theory to Work*. Cambridge University Press.

Milne, Alistair and Donald Robertson (Aug. 1996). “Firm Behaviour under the Threat of Liquidation”. In: *Journal of Economic Dynamics and Control* 20.8, pp. 1427–1449. ISSN: 0165-1889. DOI: 10.1016/0165-1889(95)00906-X.

Opler, Tim et al. (Apr. 1999). “The Determinants and Implications of Corporate Cash Holdings”. In: *Journal of Financial Economics* 52.1, pp. 3–46. ISSN: 0304-405X. DOI: 10.1016/S0304-405X(99)00003-3.

Ostrovsky, Michael and Michael Schwarz (May 2023). “Reserve Prices in Internet Advertising Auctions: A Field Experiment”. In: *Journal of Political Economy*, pp. 000–000. ISSN: 0022-3808. DOI: 10.1086/725702.

Paarsch, H.J., H. Hong, and M.R. Haley (2006). *An Introduction to the Structural Econometrics of Auction Data*. Mit Press. MA. ISBN: 978-0-262-16235-7.

Palfrey, Thomas R. (1980). “Multiple-Object, Discriminatory Auctions with Bidding Constraints: A Game-Theoretic Analysis”. In: *Management Science* 26.9, pp. 935–946. ISSN: 0025-1909.

Perrigne, Isabelle and Quang Vuong (Aug. 2019). “Econometrics of Auctions and Nonlinear Pricing”. In: *Annual Review of Economics* 11.1, pp. 27–54. ISSN: 1941-1383, 1941-1391. DOI: 10.1146/annurev-economics-080218-025702.

Pitchik, Carolyn (July 2009). “Budget-Constrained Sequential Auctions with Incomplete Information”. In: *Games and Economic Behavior* 66.2, pp. 928–949. ISSN: 08998256. DOI: 10.1016/j.geb.2008.10.001.

Pitchik, Carolyn and Andrew Schotter (1988). “Perfect Equilibria in Budget-Constrained Sequential Auctions: An Experimental Study”. In: *The RAND Journal of Economics* 19.3, pp. 363–388. ISSN: 0741-6261. DOI: 10.2307/2555662.

Raisingh, Diwakar (Oct. 2022). “The Effect of Pre-announcements on Participation and Bidding in Dynamic Auctions”. In: p. 65.

Rochet, Jean-Charles and Stéphane Villeneuve (Aug. 2005). “Corporate Portfolio Management”. In: *Annals of Finance* 1.3, pp. 225–243. ISSN: 1614-2446, 1614-2454. DOI: 10.1007/s10436-005-0018-7.

Saini, Viplav (2012). “Endogenous Asymmetry in a Dynamic Procurement Auction”. In: *The RAND Journal of Economics* 43.4, pp. 726–760. ISSN: 0741-6261.

Weintraub, Gabriel Y., C. Lanier Benkard, and Benjamin Van Roy (2008). “Markov Perfect Industry Dynamics With Many Firms”. In: *Econometrica* 76.6, pp. 1375–1411. ISSN: 1468-0262. DOI: 10.3982/ECTA6158.

Xu, Jian et al. (Aug. 2015). “Smart Pacing for Effective Online Ad Campaign Optimization”. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 2217–2226. DOI: 10.1145/2783258.2788615.

Yao, Song and Carl F. Mela (Sept. 2008). “Online Auction Demand”. In: *Marketing Science* 27.5, pp. 861–885. ISSN: 0732-2399, 1526-548X. DOI: 10.1287/mksc.1070.0351.

Yuan, Shuai, Jun Wang, and Xiaoxue Zhao (June 2013). “Real-Time Bidding for Online Advertising: Measurement and Analysis”. In: *arXiv:1306.6542 [cs]*.

Chapter 2

THE IMPACT OF PRIVACY PROTECTION ON ONLINE ADVERTISING MARKETS

2.1 Introduction

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

In this paper, we investigate the welfare consequences of Google's plan to remove third-party cookies and introduce alternative tracking technologies under its "Privacy Sandbox" initiative.[3] Our key contribution is to quantify the unequal distributional effects on the demand side of the online advertising market, which

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

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

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

encompasses advertisers and their intermediaries who purchase advertising opportunities and match them with advertisers. Although potentially beneficial to consumer privacy, the proposed plans could have negative spillovers in terms of information monopoly and anti-competitive practices of large companies. Removing third-party cookies will undermine firms' ability to target consumers and reduce the surplus of advertisers and their intermediaries. Notably, certain intermediaries, such as major tech companies like Google, can directly obtain users' behavioral information from its widely popular online products (the Google search engine, Gmail, YouTube, etc.), while other smaller intermediaries have no such recourse. Although the proposed new technology might partially offset the loss, we demonstrate that this is insufficient to diminish the information advantage enjoyed by large players.

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

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

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

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

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

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

the ban.

Related literature

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

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

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

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

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

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

[8]See Acquisti et al. (2016) and Brown (2016) for reviews of the economics of privacy and Goldfarb and Que (2023) for a review of the economics of digital privacy. Several authors (Goldberg et al., 2019; Aridor et al., 2020) study the impact of the European Union's General Data Protection

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

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

2.2 Market background

Online ad auctions

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

The process of online ad auctions can involve many parties interacting automatically in real time. The auction is triggered when the user opens the web page through her browser. The publisher packages the offer of an ad space along with information about the user and sends it to the ad exchange.[9] The ad exchange then sends out a

---

Regulation (GDPR) on web traffic and ad revenue. See Johnson (2022) for a survey of studies on the economic consequences of GDPR.

[9]The offer is usually made through a supply-side platform server that acts on behalf of the publisher. This step is not relevant to our purpose. A data management platform could also be

bid request to potential bidders (DSPs), inviting them to submit a bid. Given the large volume of auctions and the complexity of online bidding, advertisers do not participate directly in these auctions, but rather via *demand-side platforms (DSPs)*, which bid on behalf of their advertiser clients.[10] Using information about the user ready to view the ad, the DSP selects the most suitable advertiser for that impression and calculates the optimal bid for the ad space, considering competition from other DSPs. In any auction, DSPs typically submit only one bid on behalf of one of their advertiser clients.[11] In what follows, we use the terms advertisers and DSPs interchangeably and abstract away from the distinction between the DSPs and their advertiser clients.

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

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

---

involved to retrieve stored information of the user that may be of interest to the advertisers. The supply-side platform packages the ad space offer with all relevant information and sends it to the ad exchange.

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

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

Cookies and behavioral targeting

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

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

Privacy protection

Given the controversial nature of third-party cookies and the growing concern over privacy breaches, many internet entities have either eliminated or curtailed third-party cookies in recent years. Web browsers have been at the forefront of this move.

Safari and Firefox (which we refer to as the *blocked* browsers) have already blocked third-party cookies for their users and effectively shut down behavioral targeting by blocking the execution of scripts embedded by third-party servers. Third-party cookies are mostly unavailable for users of blocked browsers. On the other hand, as of 2022, Chrome, together with a few other browsers including Microsoft Edge (the *allowed* browsers), still enables third-party cookies by default. Third-party cookies are generally available on these browsers but could still be absent for a host of reasons.[12]

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

2.3 Data and Descriptive Statistics

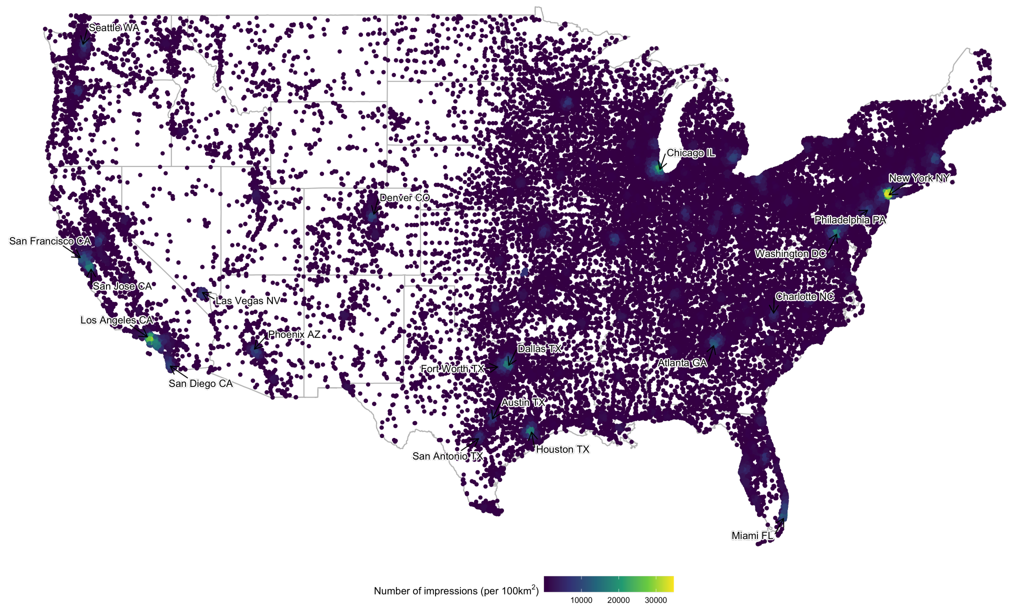
We employ bidding data from banner ad auctions on sixteen websites of Yahoo, including Homepage, News, Finance, etc. We focus on a specific display ad format known as medium rectangular (MREC) units, which has the dimension 300×250 and is displayed to the right of the main content. This is one of the most popular ad formats, and the fixed size and position help us eliminate potential heterogeneity arising from these aspects. We consider a sample of user impressions from the United States during one week in May 2022. Figure 2.1 shows the geographical distribution of our sample, which roughly coincides with the population density of the US. The dataset consists of over 5.5 million bids from about 740,000 auctions.

Table 2.1 presents summary statistics of key variables in the dataset. The variable *bid* is the submitted bid price of an individual DSP. For reasons of confidentiality, we normalize the submitted bids to have a sample mean equal to 1. For every auction, we observe the *number of bidders* (out of a total of 33 DSPs) who entered

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

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

Figure 2.1: Geographical distribution of impressions.



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

the auction and submitted a bid, as well as the *winning (highest) bid*. There is substantial variation in the number of actual bidders for each auction, with a mean of 7.5 bidders and a standard deviation of 4.7. Our empirical model will factor in this important behavioral pattern and account for bidders' entry decisions.

Two key variables describe the availability of third-party cookies for each impression. The variable *percentage of cookie matched* is the percentage of DSPs in each auction who matched the user with a profile in the bidders' database constructed with third-party cookies. Small percentages of cookie matched indicate that less information is available for the user.[14] The variable *cookie matched* is a binary variable indicating whether the percentage of cookie matched is nonzero for the impression. In other words, it indicates whether at least one bidder has a cookie identifier for the user. For ease of interpretation, our empirical analysis will primarily focus on this variable. In what follows, we refer to impressions with cookie matched = 1 as "cookie impressions" and those with cookie matched = 0 as

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

Table 2.1: Summary statistics.

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

"cookieless impressions."

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

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

Table 2.2: Summary statistics of impression characteristics by browser.

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

revenue derived from the user, respectively, where total revenue is normalized with mean 0 and standard deviation 1. Finally, the variable *days in database* (in 10,000s) is the number of days for which the user profile has existed in Yahoo's database. A smaller number of days may imply that less information is available for the user.[15]

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

Because our analysis focuses on the impact of Google's plan to terminate third-party

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

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

cookies on Chrome, in Table 2.2, we show the mean statistic of key impression characteristics, broken down by browser. Importantly, Chrome accounts for almost 60% of the impressions in our data and dominates the browser industry, suggesting a substantial impact of Google's plan on the market. Impressions from Safari and Firefox, the two browsers that ban third-party cookies by default, account for roughly 20% of impressions. Accordingly, impressions from Safari and Firefox are missing third-party cookie information, i.e., *cookie matched* = 0, and gender and age information is unavailable for these impressions as well.

Bidding patterns

Table 2.3 presents a comparison of summary auction statistics for impressions with and without cookies. For either category, we calculate the averages and standard deviations (in parentheses) of the bid, the winning bid, the number of bidders, and the entry probability. Impressions with third-party cookie identifiers fare better for all variables of interest. In particular, submitted bids on average are about 25% higher (1.0 versus 0.76) for cookie impressions, and winning bids for cookie impressions are over two times higher (2.5 versus 1.2) than cookieless impressions. The difference arises from both higher submitted bids and a larger number of participating bidders, with bidders more than twice as likely to enter auctions for cookie impressions. Finally, the standard deviation of bids and winning bids are higher for cookie impressions. This is expected because DSPs have the most information on these users, which increases the targeting opportunities and hence the variation in advertisers' bids.

Figure 2.2a shows the empirical CDFs of submitted bids in the dataset for five categories of DSPs in the dataset by their type and size (as discussed in Section 2.2): 5 large general-purpose, 10 small general-purpose, 9 rebroadcaster, 3 large specialized, and 6 small specialized DSPs. Consistent with the results in Table 2.3, DSPs tend to bid higher for cookie impressions. In fact, the distribution of bids for cookie impressions first-order stochastically dominates that for cookieless impressions. Figure 2.2a also shows heterogeneity in submitted bid distributions among different groups of DSPs. The differences are driven by a few factors: Large DSPs generally have better access to user information, have more budget and experience, and are more sophisticated in matching advertisers with impressions. Specialized DSPs could focus on some areas of advertising, such as retailing or reconnecting with existing customers (e.g. retargeting). In terms of auction participation, Figure 2.2b displays the frequencies with which the five groups of DSPs participate in

Table 2.3: Comparison between auctions with and without third-party cookies.

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

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

auctions for impressions with and without third-party cookies, and it also highlights heterogeneity in entry behavior across DSPs. The observed heterogeneity among the bidding DSPs motivates us to adopt an auction model with asymmetric bidders in the structural estimation exercise discussed below.

Evidence of the value of third-party cookies

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

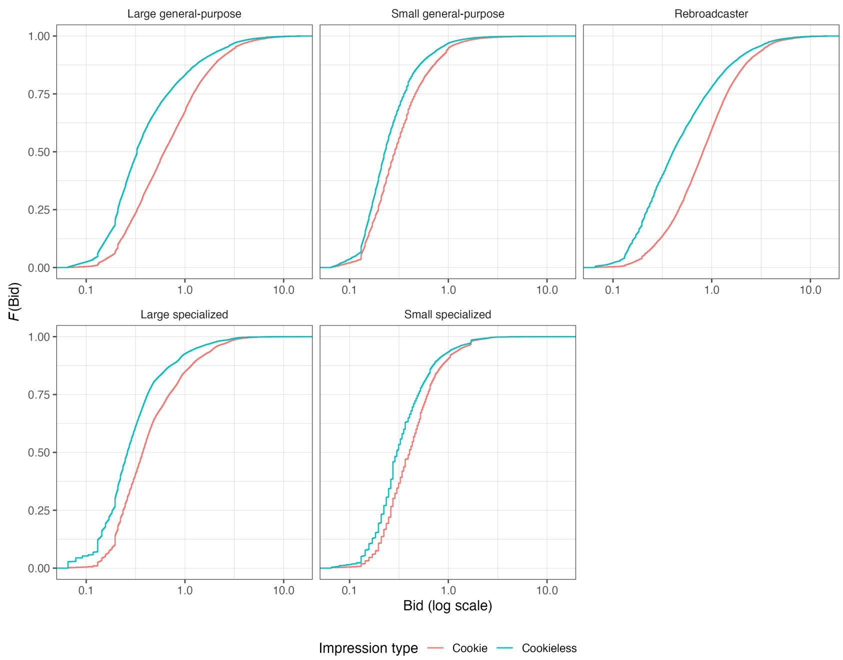
$$y_i = \beta_c \text{Cookie}_i + \mathbf{x}_i' \boldsymbol{\beta} + \alpha_i + \epsilon_i \tag{2.1}$$

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

We first analyze the effect of cookie availability on submitted bids by taking the outcome variable $y_i = \log(\text{Bid}_i)$ for every bid i in equation 2.1. Table 2.4 columns

Figure 2.2: Cookie vs. cookieless: observed bidders' behavior by DSP group.

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



(b) Average entry frequencies.

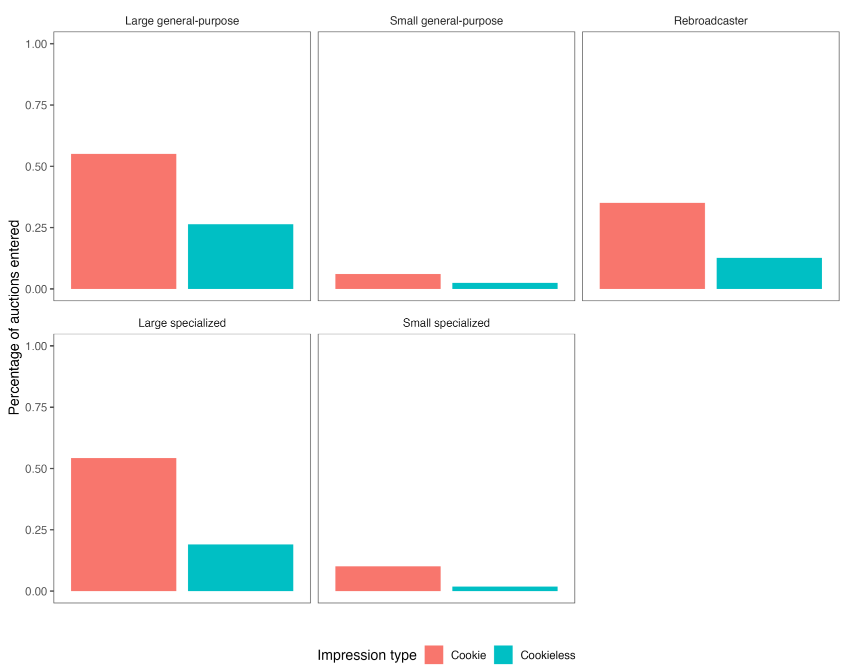


Table 2.4: Regression results for submitted bids and winning bids.

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

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

Table 2.5: Regression results for the number of bidders and entry decision.

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

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

(1)-(3) report the results of three alternative specifications. Column (1) includes only cookie availability and fixed effects; column (2) adds additional covariates; column (3) further adds a DSP fixed effect to account for bidder heterogeneity. We find quantitatively similar results in these models: having third-party cookies increases submitted bids by around 30%.

Next, we take the outcome variable $y_i = \log(\text{Winning bid}_i)$ for each auction i in equation 2.1 to examine the effect of cookie availability on the highest bid, which translates to the revenue for the publisher (Yahoo). Table 2.4 columns (4) and (5) report the results of two alternative specifications. We find that having third-party cookies increases the highest bids, and consequently Yahoo's revenue, by a substantial 75%. Observe that this effect is more than double the effect in columns (1)-(3). The difference can be attributed to the fact that the bid regression does not account for entry; it only captures submitted bids.

An important feature of the online ad market is that bidders participate in auctions selectively. Recall that table 2.1 showed substantial variation in the number of bidders for different auctions, with a mean of 7.5 bidders and a standard deviation of 4.7. Therefore, we run regression 2.1 with the outcome variable y_i as the number of bidders in each auction i. Table 2.5 columns (1) and (2) report the results of two alternative specifications. We find that, on average, an auction with third-party cookie identifiers induces about 5 more bidders (out of 33) to participate in the auction compared to an impression without. This is broadly consistent with some DSPs' strategies who simply only enter auctions with third-party cookie identifiers.

Lastly, we examine the effect of cookie availability on the entry decision of bidders in the auctions. In model 2.1, the outcome variable y_i is Entry_i, a binary variable constructed for each auction-bidder pair that is equal to 1 if the bidder submitted a bid in the auction. Table 2.5 columns (3)-(5) report the results of three alternative specifications of such a linear probability model. We find that, on average, bidders are about 14% more likely to participate and submit a bid if the impression has third-party cookie identifiers. Assuming independence between the 33 DSPs, the increase in entry probability translates to an average increase in the number of bidders by $33 \times 0.14 \approx 5$, which is consistent with the estimation above. As a robustness check, we estimate a logit model for auction participation, $\text{Entry}_i = \mathbf{1}\{\beta_c \text{Cookie}_i + \mathbf{x}_i' \boldsymbol{\beta} + \alpha_i + \epsilon_i \geq 0\}$, where ϵ_i follows the standard logistic distribution. Table 2.5 column (6) reports the estimated marginal effects at the mean or mode values of the explanatory variables. The magnitude of the effect of cookies is comparable to those of the linear models. In the appendix, we report the point estimates of the logit model. The estimated coefficient on cookie availability translates into an odds ratio of $e^{1.19} = 3$; that is, the probability that a bidder participates in an auction for a cookie impression is three times higher than that for an auction for a cookieless impression.

2.4 Structural Estimation

Auction model and equilibrium characterization

Our empirical model is an independent private-value auction model with asymmetric bidders and binding reserve price (Krishna, 2009; Hubbard and Paarsch, 2014). We adopt the independent private-value assumption to reflect how users' impressions are horizontally differentiated; for instance, an impression from a male consumer is more valuable for male fashion brands but less valuable for female fashion brands. As our descriptive evidence (Figure 2.2a) shows that there is significant heterogeneity in bidding behavior across bidders, we allow for bidder heterogeneity in valuation distributions. Finally, our descriptive evidence shows that bidders enter only a fraction of auctions, and auctions in our data have reserve prices that vary across different websites.[16]

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

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

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

Otherwise, the expected profit of bidder i given his bid b_i is

$$\pi_i(b_i) = (v_i - b_i) \prod_{j \neq i} F_j\big(\varphi_j(b_i)\big), \tag{2.2}$$

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

$$\frac{1}{\varphi_i(b_i) - b_i} = \sum_{j \neq i} \frac{f_j\big(\varphi_j(b_i)\big)}{F_j\big(\varphi_j(b_i)\big)} \varphi_j'(b_i) \tag{2.3}$$

for $i = 1, 2, \cdots, N$. Equation (2.3) is a system of nonlinear ordinary differential equations in the inverse bid functions $\varphi_1, \cdots, \varphi_N$ that characterizes the Bayes-Nash equilibrium.[18]

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

$$\varphi_i(r) = r. \tag{2.4}$$

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

$$\varphi_i(\overline{b}) = \overline{v}. \tag{2.5}$$

Specifications and estimation procedure

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

---

[17]Observe that the probability of winning is

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

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

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

independent of one another, abstracting away from potential dynamic considerations of the DSPs.

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

There are two features in our specification that are integral to online ad auctions. First, we account for bidder heterogeneity by allowing asymmetric bidder valuation distributions through α_i and σ_i. We let each bidder i fall into five distinct groups according to their type and size: large general-purpose, small general-purpose, rebroadcaster, large specialized, and small specialized (see section 2.2). With slight abuse of notation, the subscript i of the parameters α_i and σ_i denotes the group to which the bidder belongs. As explained, different types of DSPs cater to advertisers of different budgets, objectives, and targeted consumers, which may lead to an ex-ante difference in their valuations for impressions. The size of DSPs is a key dimension that captures their experience and expertise in matching advertisers with impressions.[21]

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

We adopt a nested estimation procedure in which the inner loop solves for the inverse bidding strategies $\varphi_{it}(b)$ using the equilibrium characterization (2.3) and the outer

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

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

loop estimates the valuation parameters using maximum likelihood.

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

$$\mathcal{L}_{it}(s_{it}, b_{it}, x_t, u_t; \gamma, \alpha_i, \sigma_i) = \big(F_{it}(r_t)\big)^{1-s_{it}} \big(f_{it}(\varphi_{it}(b_{it}))\varphi'_{it}(b_{it})\big)^{s_{it}}, \tag{2.6}$$

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

$$\mathcal{L}_t(\boldsymbol{s}_t, \boldsymbol{b}_t, x_t; \gamma, \boldsymbol{\alpha}, \boldsymbol{\sigma}, \sigma_u) = \int \Big(\prod_{i=1}^{N} \mathcal{L}_{it}\Big)\phi(u_t)du_t, \tag{2.7}$$

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

The inner loop solves for the inverse bidding functions $\varphi_{it}(b)$ for every auction. Because the equilibrium characterization (2.3) admits no closed-form solutions, we adopt a numerical approach to solve the system. Following Hubbard and Paarsch (2009), Hubbard et al. (2013), and Hubbard and Paarsch (2014), we use Mathematical Programs with Equilibrium Constraints (MPEC) to solve for the equilibrium of the first-price auction model with asymmetric bidders. We approximate the inverse bidding functions $\varphi_{it}(b)$ as a linear combination of the first K Chebyshev polynomials scaled to the interval $[r_t, \overline{b}_t]$:

$$\varphi_{it}(b) = \sum_{k=0}^{K} c_{k,it} T_k(b), \tag{2.8}$$

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

Then, we use the MPEC approach to discipline the Chebyshev coefficients $\boldsymbol{c}_t$ so that the first-order conditions defining the inverse equilibrium bid functions are approximately satisfied, subject to the boundary conditions (2.4) and (2.5). In addition, we impose rationality (players must bid less than their valuation) and monotonicity (bid functions are increasing) as shape constraints on the Chebyshev approximations (Hubbard and Paarsch, 2009; Hubbard et al., 2013). Specifically, from equation (2.3), we define

$$G_{it}(b; \boldsymbol{c}_t, \overline{b}_t) = 1 - (\varphi_{it}(b) - b) \sum_{j \neq i} \frac{f_{jt}\big(\varphi_{jt}(b)\big)}{F_{jt}\big(\varphi_{jt}(b)\big)} \varphi'_{jt}(b), \tag{2.9}$$

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

$$\min_{\boldsymbol{c}_t, \overline{b}_t} \quad \sum_{i=1}^{N} \sum_{b \in \mathcal{B}} \left[G_{it}(b; \boldsymbol{c}_t, \overline{b}_t)\right]^2 \tag{2.10}$$

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

Estimation results

Table 2.6 reports the estimated parameters of the bid distributions. The estimates are of the expected sign and magnitude. In particular, the estimated coefficient of cookie availability is positive and significant, confirming that third-party cookies increase bidders' valuations. An impression with third-party cookies available raises the mean valuation by as much as 129 percent compared to an impression without cookies. Given bid shading in first-price auctions, the estimate is consistent with the reduced-form estimate of the effect on submitted bids. The estimated intercepts α_i and standard deviation σ_i show substantial differences in the mean and variance parameters of valuation distribution across different DSP groups, where both small general-purpose and small specialized DSPs have low valuation distributions, reflecting their resource constraints. Lastly, the estimated variance of unobserved auction heterogeneity σ_u, while smaller in comparison to group-specific variances, remains statistically significant and positive. This suggests the presence of unobserved variations in auctions that are not accounted for by group-specific differences in the data.

We next present each bidder group's bidding pattern in response to third-party cookie availability in terms of their valuation distribution, entry probability, and bidding strategy. Following the group classification outlined in section 2.1, we organize the plots by large general-purpose, small general-purpose, rebroadcaster, large specialized, and small specialized DSPs. Each figure shows the outcome variables for both

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

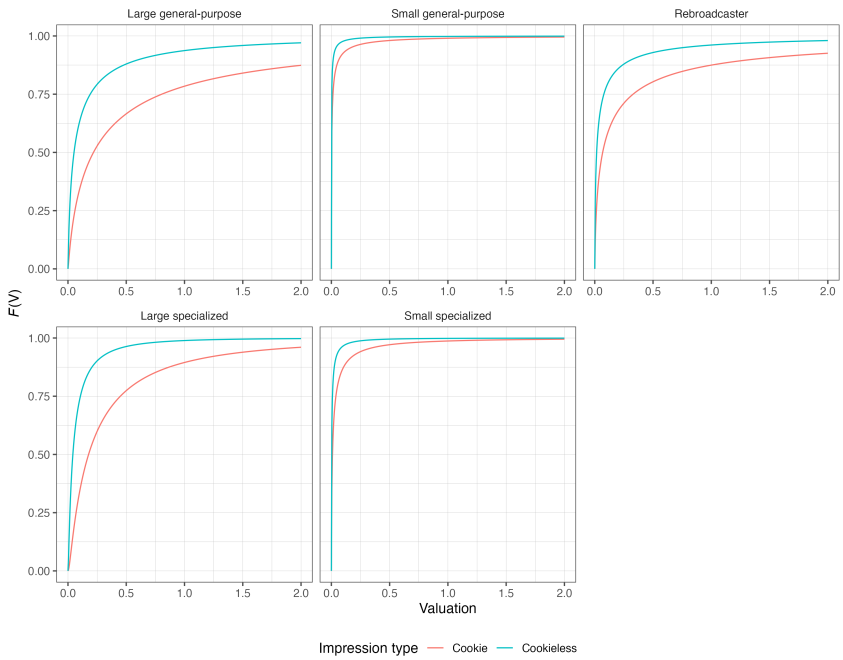
Table 2.6: Estimated parameters of valuation distributions.

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

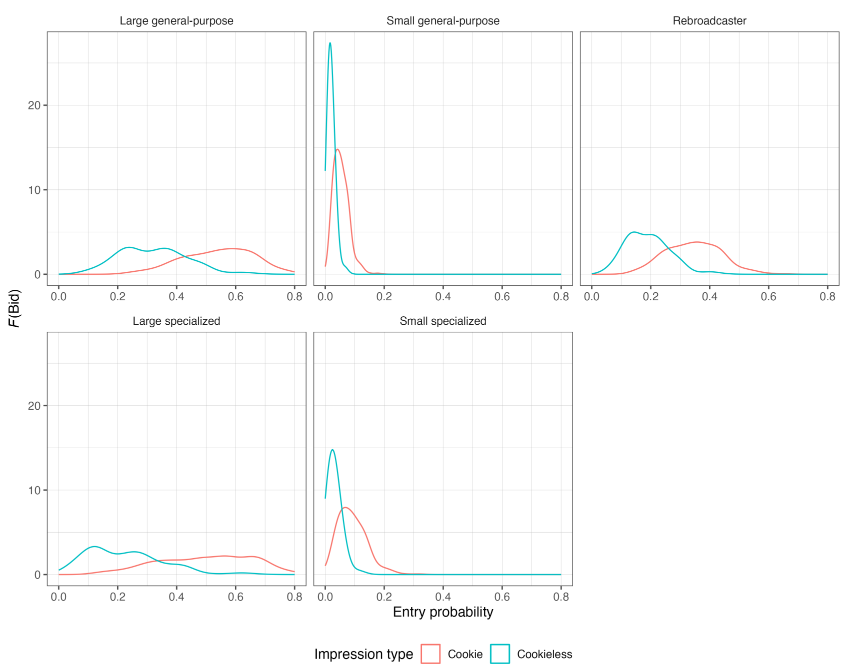
Notes: Parameter estimates of the log of valuation, $\log(v_{it})$, which follows a normal distribution with mean $x_t'\gamma + \alpha_i + u_t$ and variance σ_i^2, where u_t is the unobserved auction heterogeneity that is distributed normally with mean 0 and variance σ_u^2. Estimates of website and browser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

Figure 2.3: Cookie vs. cookieless: estimated bidders' behavior by DSP group.

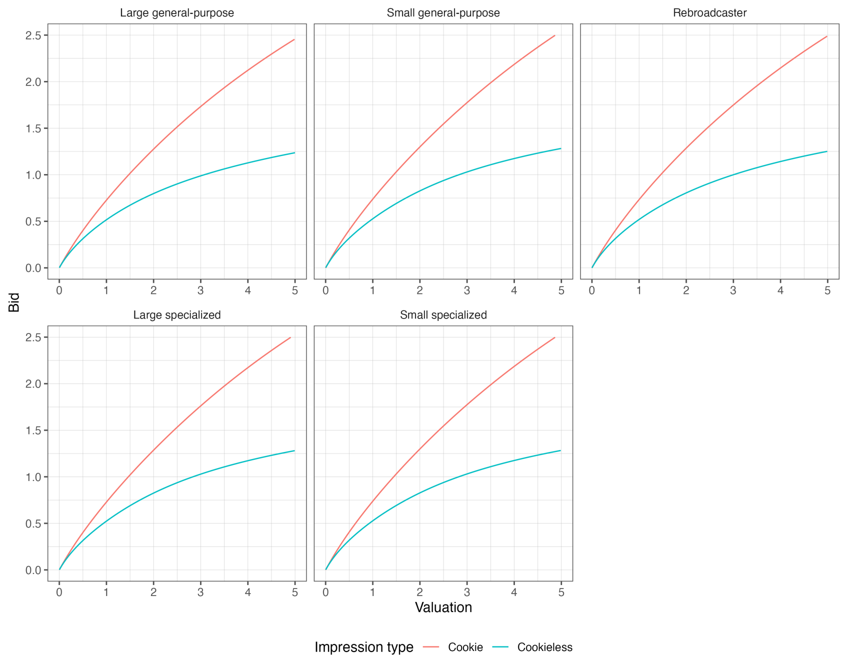
(a) CDFs of valuation distributions.



(b) Density of entry probability.



(c) Bid functions.



Notes: Plots of bidder behavior by DSP groups with estimated parameters. The DSPs are grouped according to their purpose, specialty, and size. See section 2.1 for more details on the classification of DSPs. Subplot (a) shows the cumulative distribution function F_i of valuations at average auction characteristics. (b) shows the empirical density of entry probability, i.e., how likely the valuation exceeds the reserve price and the bidder submits a bid in an auction. (c) shows the bid function β_i at average auction characteristics. See Figure B.1 of the appendix for the bid functions of big and small general-purpose DSPs on the same plot.

impressions with and without cookies. For illustration, the valuation distribution and the bidding strategy are evaluated at the average values of the covariates.

Figure 2.3a shows the cumulative distribution functions (CDFs) of recovered valuation distributions. Figure 2.3b presents the empirical density of fitted entry probability, i.e., for all auctions in the data, the probability that the recovered valuation exceeds the reserve price. For either figure, we observe a clear dominance relationship of cookie impressions over cookieless ones across different DSP groups. Bidders are more likely to place a higher value and submit a bid in an auction with third-party cookies. There is also substantial heterogeneity across bidder groups. Notably, the effect of cookie availability is more pronounced for large DSPs.

Figure 2.3c presents the bidding strategy β_i and shows that bidders bid more aggressively for cookie impressions.[23] Observe that, for the same valuation, bidders

[23]Given the relatively large number of bidders (33 in our data), average bidding strategies appear similar across bidder groups, though they do exhibit differences. See Figure B.1 in the Appendix for

on average place bids on a cookie impression that are about twice as much as those on a cookieless impression. The difference can be attributed to the competition intensity between the two types of auctions, where fewer bidders would participate in auctions for cookieless impressions. Overall, our estimated structural results demonstrate that the difference between the average revenue from the two types of auctions comes from the difference in valuations, entry behavior, and bidding strategies.

2.5 Counterfactual Simulations

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

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

The asymmetric ban mirrors concerns raised by antitrust authorities, whereby certain DSPs may have alternative ways to gather and use ad-relevant information about users even when third-party cookies are blocked. For instance, DSPs affiliated with prominent publishers may have extensive user information through *first-party* cookies, which are typically enabled even by browsers that block third-party cookies by default. They may be able to leverage this rich first-party information about users for placing ads not only on their own websites but also on third-party websites, thereby obtaining a large information advantage over DSPs without similar capabilities.[24]

---

a comparison. In particular, we find that smaller bidders adopt more aggressive bidding strategies to compete against larger bidders, who tend to have higher valuations.

[24]This alternative information collection can be implemented with "digital fingerprinting" methods that track users via IP addresses or device IDs, thus sidestepping cookies altogether. Peukert

A prominent example is Google, which possesses large amounts of first-party information on many internet users via its extensive web ecosystem encompassing the Google search engine, Gmail, YouTube, and more. This unique access to first-party information may allow Google to circumvent the effects of the Chrome third-party cookie ban and perhaps even to benefit from such a ban.[25]

To implement the counterfactual simulations, we draw a random sample of 10,000 auctions of impressions from the data. Importantly, this sample includes impressions from all browsers because we want to investigate the market-wide impact on the advertising market. For Chrome impressions (about 58% in the drawn pool), we manipulate their impression characteristics to emulate scenarios of the Cookiepocalypse. For each Chrome auction and each specification, we draw valuations based on the true user characteristics for each bidder and, depending on the scenario and the bidder, mask any third-party cookie information for each user to simulate the effects of the ban. (That is, the cookie availability variable is set to zero. Other user characteristics associated with third-party cookies are set to either unknown or zero). Given the counterfactual valuation distributions, we compute the bidding strategies by solving the system of ordinary differential equations (2.3) that characterizes the equilibrium.

Cookiepocalypse, blocking third-party cookies on Chrome

We first investigate the effect of Cookiepocalypse on submitted bids, the number of bidders, the winning bid (which translates into the publisher's revenue), and bidders' surplus. The results of this counterfactual simulation are presented in Table 2.7a. We find that the average bid falls from \$0.92 in the benchmark to \$0.56, representing a 39% decrease, and the number of bidders decreases from 7.4 to 4.8. Altogether, this results in about a halving (-54%) of the average publisher revenue from \$2.4 down to \$1.1. This estimate is consistent with several studies investigating the potential effect of removing third-party cookies including industrial studies.[26] On the buyer side, advertisers acquiring impressions through DSPs suffer a substantial

et al. (2022) observe that the drop in third-party cookie requests after the enactment of GDPR in the European Union was accompanied by a rise in first-party cookie requests.

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

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

Table 2.7: Counterfactual simulation of Cookiepocalypse.

(a) Simulated outcome.

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

(b) Welfare distribution.

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

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

40% reduction in their surplus (the difference between valuation and bid), from an average of \$3.7 in the benchmark to \$2.2 in the first counterfactual.

We next investigate the distributional effect among bidders in terms of their winning frequency and surplus to highlight the unequal impact of the Cookiepocalypse. In Table 2.7b, we report the outcome variables in the asymmetric ban counterfactual scenario separately for the Big Tech DSP and the other five bidder groups. (Recall that the Big Tech DSP is drawn from the large general-purpose DSP group in the benchmark.) In terms of winning frequency, the Big Tech DSP wins twice as often (15.4%) in this scenario compared to the benchmark (8.3%), thanks to its informational advantage of having sole access to the behavioral information of Chrome users. Its total surplus also increases accordingly from \$31,800 in the status quo to \$48,900 under the asymmetric ban, a 54% increase. At the same time, all the other bidders are impacted negatively by the ban, winning less frequently and receiving lower surpluses compared to the status quo and symmetric ban scenarios. Our results demonstrate that the third-party cookie ban leads to divergent experiences for the informational advantaged and disadvantaged bidders, where the former benefit from the ban at the cost of the latter.

To further decompose this redistributive effect, we also calculate the "full-information" surplus, that is, the difference between the valuation under cookie availability and the bid in the counterfactual scenario. The gap between the full-information and limited-information surpluses quantifies the loss in bidder welfare due to the inability to make precise matches when DSPs lose the ability to accurately evaluate and target users following the cookie ban. Comparing this difference in Table 2.7b, we see that welfare loss stems primarily from the diminished ability of affected DSP to effectively target users post-cookie ban. The primary factor responsible for the welfare redistribution is the inability of disadvantaged bidders to match with the most appropriate advertisements, rather than the Big Tech DSP monopolizing all the valuable impressions in the market.

Privacy Sandbox, alternative tracking technologies

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

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

The rightmost column of Table 2.8a contains the summary outcomes of the asymmetric ban under the Privacy Sandbox counterfactual. We find that the average bid has fallen from \$0.92 in the benchmark to \$0.82 in the counterfactual, and the number of bidders has decreased from 7.4 to 6.9. Altogether, this results in a 13% drop in the average revenue per auction from \$2.4 to \$2.1, and the bidder (advertiser) surplus drops by 8% from \$3.7 to \$3.4. In a word, the Privacy Sandbox still results in sizable welfare losses for both the publisher and the advertiser—an expected consequence given the coarser information in the market. On the other hand, the impact is a lot more cushioned compared to that of the Cookiepocalypse counterfactual under which the publisher and the advertiser bear a much heavier loss of 54% and 40%, respectively.

Table 2.8b presents the differentiated impact on DSP groups. Compared to the Cookiepocalypse counterfactual in table 2.7b, Privacy Sandbox alleviates the anti-

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

Table 2.8: Counterfactual simulation of Privacy Sandbox.

(a) Simulated outcome.

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

(b) Welfare distribution.

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

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

competitive redistribution as well as the rising market concentration in favor of the Big Tech DSP. For the Big Tech bidder under the asymmetric ban, both its winning frequency (9.4%) and total surplus ($36,000) increase compared to the benchmark (8.3% and $31,800, respectively), representing a more than 10% gain, though the advantage is substantially attenuated compared to that under Cookiepocalypse. The disadvantaged bidders also experience noteworthy improvement compared to Cookiepocalypse. Their metrics under either symmetric or asymmetric ban are much closer to the status quo level: Under the asymmetric ban, for example, large general-purpose DSPs enjoy a surplus of $28,600, below the status quo level of $31,800, but a substantial alleviation compared to $17,600 under Cookiepocalypse. Although still a heavy 10% loss from the advertiser's perspective, this set of results suggests that advertising surplus and user privacy may not necessarily be at great odds. DSPs can rely on privacy-friendly technologies and coarser information to implement targeted ads without severely hurting their bottom lines. The anticompetitive redistribution effect, although much ameliorated compared to Cookiepocalypse, is still present and significant.

2.6 Conclusion

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

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

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

References

Abraham, Ittai et al. (Nov. 2020). "Peaches, lemons, and cookies: Designing auction markets with dispersed information". en. In: *Games and Economic Behavior* 124, pp. 454–477. ISSN: 0899-8256. DOI: 10.1016/j.geb.2020.09.004. URL: https://www.sciencedirect.com/science/article/pii/S0899825620301342 (visited on 11/15/2022).

Acquisti, Alessandro, Curtis Taylor, and Liad Wagman (June 2016). "The Economics of Privacy". en. In: *Journal of Economic Literature* 54.2, pp. 442–492. ISSN: 0022-0515. DOI: 10.1257/jel.54.2.442. URL: https://www.aeaweb.org/articles?id=10.1257/jel.54.2.442 (visited on 12/05/2022).

Aridor, Guy, Yeon-Koo Che, and Tobias Salz (Mar. 2020). *The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from GDPR*. Working Paper. DOI: 10.3386/w26900. URL: https://www.nber.org/papers/w26900 (visited on 02/13/2023).

Athey, Susan and Philip A Haile (2007). "Nonparametric approaches to auctions". In: *Handbook of econometrics* 6. Publisher: Elsevier, pp. 3847–3965.

Athey, Susan, Jonathan Levin, and Enrique Seira (Feb. 2011). "Comparing Open and Sealed Bid Auctions: Evidence from Timber Auctions". In: *The Quarterly Journal of Economics* 126.1, pp. 207–257. ISSN: 0033-5533. DOI: 10.1093/qje/qjq001. URL: https://doi.org/10.1093/qje/qjq001 (visited on 10/18/2022).

Barth, Susanne and Menno D. T. de Jong (Nov. 2017). "The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review". en. In: *Telematics and Informatics* 34.7, pp. 1038–1058. ISSN: 0736-5853. DOI: 10.1016/j.tele.2017.04.013. URL: https://www.sciencedirect.com/science/article/pii/S0736585317302022 (visited on 10/11/2022).

Beales, Howard and Jeffrey A. Eisenach (Jan. 2014). *An Empirical Analysis of the Value of Information Sharing in the Market for Online Content*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.2421405. URL: https://papers.ssrn.com/abstract=2421405 (visited on 02/12/2023).

Brown, Ian (2016). "The economics of privacy, data protection and surveillance". In: *Handbook on the Economics of the Internet*. Edward Elgar Publishing, pp. 247–261. URL: https://econpapers.repec.org/bookchap/elgeechap/14700_5f12.htm (visited on 02/12/2023).

Campbell, James, Avi Goldfarb, and Catherine Tucker (2015). "Privacy Regulation and Market Structure". en. In: *Journal of Economics & Management Strategy* 24.1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jems.12079, pp. 47–73. ISSN: 1530-9134. DOI: 10.1111/jems.12079. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jems.12079 (visited on 02/12/2023).

Celis, L. Elisa et al. (Dec. 2014). "Buy-It-Now or Take-a-Chance: Price Discrimination Through Randomized Auctions". In: *Management Science* 60.12. Publisher: INFORMS, pp. 2927–2948. ISSN: 0025-1909. DOI: 10.1287/mnsc.2014.2009. URL: https://pubsonline.informs.org/doi/abs/10.1287/mnsc.2014.2009 (visited on 08/18/2022).

Decarolis, Francesco, Maris Goldmanis, and Antonio Penta (Oct. 2020). "Marketing Agencies and Collusive Bidding in Online Ad Auctions". In: *Management Science* 66.10. Publisher: INFORMS, pp. 4433–4454. ISSN: 0025-1909. DOI: 10.1287/mnsc.2019.3457. URL: https://pubsonline.informs.org/doi/10.1287/mnsc.2019.3457 (visited on 07/20/2023).

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

Ghose, Anindya and Vilma Todri-Adamopoulos (2016). "Toward a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior". In: *MIS Quarterly* 40.4. Publisher: Management Information Systems Research Center, University of Minnesota, pp. 889–910. ISSN: 0276-7783. URL: https://www.jstor.org/stable/26629681 (visited on 02/14/2023).

Goldberg, Samuel, Garrett Johnson, and Scott Shriver (July 2019). *Regulating Privacy Online: An Economic Evaluation of the GDPR*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.3421731. URL: https://papers.ssrn.com/abstract=3421731 (visited on 02/13/2023).

Goldfarb, Avi (2014). "What is Different About Online Advertising?" In: *Review of Industrial Organization* 44.2. Publisher: Springer, pp. 115–129. ISSN: 0889-938X. URL: https://www.jstor.org/stable/43550450 (visited on 02/09/2023).

Goldfarb, Avi and Verina F. Que (Feb. 2023). *The Economics of Digital Privacy*. Working Paper. DOI: 10.3386/w30943. URL: https://www.nber.org/papers/w30943 (visited on 02/13/2023).

Goldfarb, Avi and Catherine Tucker (Mar. 2019). "Digital Economics". en. In: *Journal of Economic Literature* 57.1, pp. 3–43. ISSN: 0022-0515. DOI: 10.1257/jel.20171452. URL: https://www.aeaweb.org/articles?id=10.1257/jel.20171452 (visited on 02/14/2023).

Haile, Philip A and Yuichi Kitamura (2019). "Unobserved heterogeneity in auctions". In: *Econometrics Journal* 22.1. Publisher: Oxford University Press, pp. C1–C19.

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

Hu, Yingyao, David McAdams, and Matthew Shum (June 2013). "Identification of first-price auctions with non-separable unobserved heterogeneity". en. In: *Journal of Econometrics* 174.2, pp. 186–193. ISSN: 0304-4076. DOI: 10.1016/j.jeconom.2013.02.005. URL: https://www.sciencedirect.com/science/article/pii/S0304407613000407 (visited on 12/07/2022).

Hubbard, Timothy P., René Kirkegaard, and Harry J. Paarsch (Aug. 2013). "Using Economic Theory to Guide Numerical Analysis: Solving for Equilibria in Models of Asymmetric First-Price Auctions". en. In: *Computational Economics* 42.2, pp. 241–266. ISSN: 1572-9974. DOI: 10.1007/s10614-012-9333-z. URL: https://doi.org/10.1007/s10614-012-9333-z (visited on 10/18/2022).

Hubbard, Timothy P. and Harry J. Paarsch (Jan. 2009). "Investigating bid preferences at low-price, sealed-bid auctions with endogenous participation". en. In: *International Journal of Industrial Organization* 27.1, pp. 1–14. ISSN: 0167-7187. DOI: 10.1016/j.ijindorg.2008.05.004. URL: https://www.sciencedirect.com/science/article/pii/S0167718708000672 (visited on 03/01/2023).

– (Jan. 2014). "On the Numerical Solution of Equilibria in Auction Models with Asymmetries within the Private-Values Paradigm". en. In: *Handbook of Computational Economics*. Ed. by Karl Schmedders and Kenneth L. Judd. Vol. 3. Handbook of Computational Economics Vol. 3. Elsevier, pp. 37–115. DOI: 10.1016/B978-0-444-52980-0.00002-5. URL: https://www.sciencedirect.com/science/article/pii/B9780444529800000025 (visited on 03/01/2023).

Jeon, Doh-Shin (2020). *Market power and transparency in open display advertising – a case study*. Tech. rep. EU Observatory on the Online Platform Economy.

Johnson, Garrett (Dec. 2022). *Economic Research on Privacy Regulation: Lessons from the GDPR and Beyond*. Working Paper. DOI: 10.3386/w30705. URL: https://www.nber.org/papers/w30705 (visited on 12/05/2022).

Johnson, Garrett, Scott Shriver, and Samuel Goldberg (Jan. 2022). *Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.3477686. URL: https://papers.ssrn.com/abstract=3477686 (visited on 08/18/2022).

Johnson, Garrett A., Scott K. Shriver, and Shaoyin Du (Jan. 2020). "Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry?" In: *Marketing Science* 39.1. Publisher: INFORMS, pp. 33–51. ISSN: 0732-2399. DOI: 10.1287/mksc.2019.1198. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.2019.1198 (visited on 08/18/2022).

Kesler, Reinhold, Michael Kummer, and Patrick Schulte (2019). *Competition and Privacy in Online Markets: Evidence from the Mobile App Industry*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.3542203. URL: https://papers.ssrn.com/abstract=3542203 (visited on 02/10/2023).

Kong, Yunmi (2020). "Not knowing the competition: evidence and implications for auction design". In: *RAND Journal of Economics* 51.3. Publisher: Wiley Online Library, pp. 840–867.

Krasnokutskaya, Elena (2011). "Identification and Estimation of Auction Models with Unobserved Heterogeneity". In: *The Review of Economic Studies* 78.1. Publisher: [Oxford University Press, Review of Economic Studies, Ltd.], pp. 293–327. ISSN: 0034-6527. URL: https://www.jstor.org/stable/23015856 (visited on 10/19/2022).

Krasnokutskaya, Elena and Katja Seim (Oct. 2011). "Bid Preference Programs and Participation in Highway Procurement Auctions". en. In: *American Economic Review* 101.6, pp. 2653–2686. ISSN: 0002-8282. DOI: 10.1257/aer.101.6.2653. URL: https://www.aeaweb.org/articles?id=10.1257/aer.101.6.2653 (visited on 10/18/2022).

Krishna, Vijay (Sept. 2009). *Auction Theory*. en. Google-Books-ID: qW1128ktG1gC. Academic Press. ISBN: 978-0-08-092293-5.

Lebrun, Bernard (1999). "First Price Auctions in the Asymmetric N Bidder Case". In: *International Economic Review* 40.1. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University], pp. 125–142. ISSN: 0020-6598. URL: https://www.jstor.org/stable/2648842 (visited on 09/08/2023).

Levin, Jonathan and Paul Milgrom (May 2010). "Online Advertising: Heterogeneity and Conflation in Market Design". en. In: *American Economic Review* 100.2, pp. 603–607. ISSN: 0002-8282. DOI: 10.1257/aer.100.2.603. URL: https://www.aeaweb.org/articles?id=10.1257/aer.100.2.603 (visited on 10/09/2022).

Lewis, Randall and David Reiley (2014). "Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!" In: *Quantitative Marketing and Economics (QME)* 12.3. Publisher: Springer, pp. 235–266. ISSN: 1570-7156. URL: https://econpapers.repec.org/article/kapqmktec/v_3a12_3ay_3a2014_3ai_3a3_3ap_3a235-266.htm (visited on 02/14/2023).

Marotta, Veronica, Vibhanshu Abhishek, and Alessandro Acquisti (2019). *Online Tracking and Publishers' Revenues: An Empirical Analysis*. en.

Perrigne, Isabelle and Quang Vuong (2019). "Econometrics of Auctions and Nonlinear Pricing". In: *Annual Review of Economics* 11.1, pp. 27–54. DOI: 10.1146/annurev-economics-080218-025702. URL: https://doi.org/10.1146/annurev-economics-080218-025702.

Peukert, Christian et al. (July 2022). "Regulatory Spillovers and Data Governance: Evidence from the GDPR". In: *Marketing Science* 41.4. Publisher: INFORMS, pp. 318–340. ISSN: 0732-2399. DOI: 10.1287/mksc.2021.1339. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.2021.1339 (visited on 08/18/2022).

Rafieian, Omid and Hema Yoganarasimhan (Mar. 2021). "Targeting and Privacy in Mobile Advertising". In: *Marketing Science* 40.2. Publisher: INFORMS, pp. 193–218. ISSN: 0732-2399. DOI: 10.1287/mksc.2020.1235. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.2020.1235 (visited on 01/10/2022).

Ravichandran, Deepak and Nitish Korula (2019). *Effect of disabling third-party cookies on publisher revenue*. Tech. rep. Google. URL: https://services.google.com/fh/files/misc/disabling_third-party_cookies_publisher_revenue.pdf.

Rutz, Oliver and Randolph Bucklin (2012). "Does banner advertising affect browsing for brands? clickstream choice model says yes, for some". en. In: *Quantitative Marketing and Economics (QME)* 10.2. Publisher: Springer, pp. 231–257. URL: https://ideas.repec.org//a/kap/qmktec/v10y2012i2p231-257.html (visited on 02/14/2023).

Chapter 3

ROBUST ESTIMATION OF RISK PREFERENCES

3.1 Introduction

Economic models estimated from experimental data have the potential to guide decisions and influence policies when they accurately predict real-world behaviors. For instance, financial advisors can use surveys involving risky choices to predict the portfolio composition that best aligns with their clients' risk attitudes. Similarly, experiments on risk preferences can inform policies aimed at regulating high-risk financial behaviors, such as speculative trading or excessive borrowing. However, even within experimental settings, many economic models that successfully rationalize specific behaviors fail to provide reliable out-of-sample predictions (for discussion and examples see, e.g., Agranov et al., 2023; Chapman et al., Chapman et al., 2023a, Chapman et al., 2023b; Dean and Ortoleva, Dean and Ortoleva, 2015, Dean and Ortoleva, 2019). As an alternative predictive approach, machine learning algorithms are gaining attention in the field of economics (Hofman et al., 2021). These methods generate predictions without an explicit economic model, but they have also been criticized for their poor out-of-sample performance.[1]

This paper contributes to the important question of how well we can predict behavior under risk in two ways. First, we conduct an experiment to examine two classical behaviors inconsistent with Expected Utility (EU): the common ratio effect and preferences for randomization. These behaviors have mostly been analyzed in isolation in prior experimental work. Our findings reveal that leading economic models and machine learning algorithms produce inaccurate predictions about one non-EU behavior when trained on choices related to the other. Second, motivated by this negative result, we propose a novel empirical approach to predict behavior under risk. This approach can be used to study any preference that is complete, transitive, and continuous—three common assumptions in decision-making models—without committing to a specific economic model. We then demonstrate the potential of this new approach in producing more accurate out-of-sample predictions, both within our experiment and in high-stakes behaviors outside of the lab.

Section 3.2 introduces our novel experimental design, which allows us to assess

[1] See Athey (2017) and Coveney et al. (2016).

the capability of various approaches to concurrently rationalize and predict risky choices in two important settings. The first setting is the common ratio version of the Allais paradox, where subjects can violate EU by either displaying higher or lower risk aversion when one option is certain, compared to when all options are risky. The former behavior is referred to as the common ratio effect, and its counterpart as the reverse common ratio effect. The second setting examines preferences for lottery mixtures. Subjects can violate EU in this setting by showing a strict preference for mixing lotteries, which we call preferences for randomization, or by displaying a strict aversion to mixing lotteries, which we call aversion to randomization. In our experiment, subjects engaged in incentivized binary choice tasks involving monetary lotteries. The study was conducted on Prolific, with a total of 500 subjects recruited on July 28, 2023. We call choice tasks associated with the common ratio version of the Allais paradox "CR-tasks", and those involving mixture lotteries "R-tasks".

We focus on these two tests of EU because they contributed to the development of many non-EU models under risk—see, for instance, Cerreia-Vioglio et al., 2015; Chew et al., 1991; Gul, 1991; Loomes and Sugden, 1982, Kahneman and Tversky, 1979; Tversky and Kahneman, 1992—and because of two limitations that we identified in prior experimental work. First, previous experiments have studied these EU tests either in isolation or with a predominant focus on one over the other.[2] As a result, little is known about whether a model's rationale for one non-EU behavior leads to accurate predictions for the other. Second, most prior experiments do not allow for the unambiguous elicitation of aversion to randomization.[3] This is important because aversion to randomization often emerges as the prediction of a popular model like Cumulative Prospect Theory (CPT) under standard parametrizations used to explain behavior in the common ratio version of the Allais paradox.

Our experiment employs a within-subjects design to study the common ratio version

[2]To the best of our knowledge, Agranov and Ortoleva (2017) is the only paper that studies both the common ratio version of the Allais paradox and randomization within the same experiment. Their experiment primarily focuses on randomization, with the common ratio version of the Allais paradox implemented through a single pair of binary choice tasks presented to subjects at the end of the experiment.

[3]There are three approaches that are commonly used in the literature to elicit preferences for mixtures. The first approach requires subjects to choose between two lotteries multiple consecutive times (Agranov and Ortoleva, 2017; Agranov et al., 2023; Dwenger et al., 2018). The second approach allows subjects to delegate their choice to an external randomization device (Agranov and Ortoleva, 2017; Cettolin and Riedl, 2019; Sandroni et al., 2013; Dwenger et al., 2018). The third approach presents a choice between two lotteries and some mixtures between them (Agranov and Ortoleva, 2023; Dwenger et al., 2018; Feldman and Rehbeck, 2022; Miao and Zhong, 2018; Sopher and Narramore, 2000). Agranov and Ortoleva (2022) provide an overview of these methods. None of them allows for the direct revelation of an aversion to randomization.

of the Allais paradox in CR-tasks and randomization in R-tasks, using a common set of lotteries. Moreover, to enable ourselves to unambiguously identify both preference for and aversion to randomization, we design R-tasks following a procedure proposed by Camerer and Ho (1994). In particular, we evaluate attitudes towards mixtures between any two lotteries s and r through two distinct R-tasks. In the first task, subjects compare lottery s to a 50-50 mixture of lotteries s and r. In the second task, they compare the 50-50 mixture to lottery r. If a subject chooses the 50-50 mixture in both tasks, this indicates a preference for randomization. On the other hand, if the subject avoids the 50-50 mixture in both tasks, we infer an aversion to randomization.

Section 3.3 summarizes the main findings of the experiment and explores the ability of popular economic models and machine learning algorithms to rationalize them. The common ratio effect and preferences for randomization are the two most frequent non-EU behaviors observed in CR-tasks and R-tasks, respectively. In particular, the common ratio effect accounts for around 63% of all non-EU behaviors in CR-tasks, while preferences for randomization account for around 55% of all non-EU behaviors in R-tasks. Moreover, for fixed values of probabilities and prizes of the lotteries, the percentage of non-EU behavior in CR-tasks attributed to the common ratio effect is strongly positively correlated with the percentage of non-EU behavior in R-tasks attributed to preferences for randomization, with a correlation coefficient of 0.63.

After documenting the emergence of and the correlation between non-EU behaviors in the experiment, we turn to analyzing whether popular economic models and machine learning algorithms can accommodate them. We consider EU and Cumulative Prospect Theory (CPT) as economic models, and gradient boosting trees (GBT) and neural networks (NN) as machine learning algorithms.[4] In particular, we implement two types of out-of-sample exercises. First, we evaluate the ability of various approaches to predict choices in each EU test separately, employing 5-fold cross-validation within CR-tasks and R-tasks. Second, we assess the ability of different approaches to predict the correlation between non-EU behaviors across CR-tasks and R-tasks. To this end, we use choices in CR-tasks as the training set and choices in R-tasks as the test set, and vice versa.

Machine learning algorithms predict non-EU behavior within CR-tasks and R-tasks better than economic models. In particular, GBT correctly classifies more than

[4]Details about the functional form assumptions we made on economic models and about the training procedures for the machine learning algorithms can be found in Appendix C.2.

60% of the observations in cross-validation exercises, while the best-performing economic model, CPT, correctly classifies only around 46% of the observations. However, the ranking between machine learning algorithms and economic models is reversed when attempting to predict one non-EU behavior from another. In these exercises, the predictive accuracy of machine learning models drops by about 30%, raising concerns about their ability to produce generalizable predictions. Moreover, CPT predicts a negative rather than a positive correlation between the common ratio effect and preferences for randomization, performing worse than EU. Overall, EU turns out to be the best model at predicting choices across CR-tasks and R-tasks, despite the fact that non-EU behaviors account for more than 34% of the observations in CR-tasks and more than 47% of the observations in R-tasks.

Our experiment sheds light on a new important instance of a more general problem: economic models and machine learning algorithms often provide poor predictions across settings. Motivated by this observation, Section 3.4 turns to the methodological contribution of the paper, which is the development of a new approach to estimate preferences under risk and make predictions. Intuitively, estimating a model of decision-making under risk requires making a series of assumptions about preferences. Some assumptions are less controversial, as they are embodied in most models. For instance, most models assume that preferences are complete, transitive, and continuous. Other assumptions are more controversial, and this is the case with the independence axiom. The independence axiom is the most empirically challenged assumption of EU, and the common ratio effect and preference for randomization are two examples of behaviors inconsistent with this assumption. Our approach relies on the less controversial assumptions, as it may be more likely that they hold across different settings, while remaining silent about the independence axiom.

Rather than estimating a representation for a preference, which requires making assumptions about the independence axiom, we estimate a representation for its EU-core, which is its largest subrelation that satisfies the independence axiom (Cerreia-Vioglio, 2009). Assuming our preference of interest is complete, transitive, and continuous, its EU-core can be represented by a set of utility functions.[5] This set generates an upper and lower bound for risk aversion, which becomes wider as the preference's inconsistency with the independence axiom increases. We demonstrate that the set of utilities representing the EU-core can be easily estimated using

[5]In particular, Cerreia-Vioglio et al. (2017) clarifies that the EU-core is represented by the set of "local utilities" introduced by Machina (1982).

standard experimental datasets that examine the independence axiom, and we present the estimation results from our experiment. By estimating mixture models, we observe a significant degree of heterogeneity in terms of risk aversion and adherence to EU. In Section 3.5, we assess whether this heterogeneity can be fruitfully exploited to make reliable predictions across different settings.

We perform various types of out-of-sample exercises, distinguished by the differences in the features of training and test data. Initially, we revisit the two out-of-sample exercises within and across CR-tasks and R-tasks introduced in Section 3.3. Because our empirical approach applies to all models that relax EU by violating the independence axiom, it remains silent about how exactly the independence axiom fails. In other words, we can predict the emergence of EU and non-EU behavior, but we cannot differentiate between specific non-EU behaviors. To establish a fair comparison between our approach and other predictive methods, we evaluate the ability of all methods to predict adherence to EU. Consistent with our findings in Section 3.3, machine learning algorithms exhibit better performance when predicting non-EU behaviors in isolation but fall short when predicting one non-EU behavior using choices related to the other as training data. Importantly, our structural model achieves the best performance in these latter exercises, demonstrating potential advantages in making predictions that are not tied to specific economic models.

To provide an additional setting for testing the predictive performance of various approaches, we elicited certainty equivalents for three binary lotteries at the end of our experiment. Using choices from CR-tasks and R-tasks as training data, our method predicts ranges of possible certainty equivalents, which we demonstrate to encompass most of the observed certainty equivalents. Furthermore, we calculate point predictions for certainty equivalents as the midpoints of these predicted ranges.[6] To compare the accuracy of our approach against other methods, we employ mean squared errors and find that machine learning algorithms underperform compared to economic models, while our approach yields the most accurate out-of-sample predictions.

Finally, we explore whether the heterogeneity in risk aversion and adherence to EU, identified through our mixture model using experimental data, correlates with

[6] We abstract away from the question of what theoretical assumptions on preferences justify this aggregation rule. Nevertheless, given that this rule outperforms both the economic models and the machine learning algorithms that we consider, we find its theoretical analysis an interesting direction for future work.

risky behaviors in real-world scenarios. To pursue this, we gathered information on subjects' investment and insurance behaviors. In the context of investments, we assessed whether they had engaged in stock trading and held any cryptocurrencies. Additionally, we verified whether they had insured purchased items, such as mobile phones. This is a classic example of small stakes risk aversion, which can be challenging to rationalize with EU (Rabin, 2000). Our findings reveal that subjects identified as more risk-averse in the experiment are less likely to invest, particularly in cryptocurrencies. Moreover, those identified as less consistent with EU are more likely to insure purchased items. Overall, these correlations hint at our structural model as a potentially useful tool for predicting behavior beyond experimental settings.

Our paper has two broad implications for future research, which we expand upon in our concluding Section 3.6. First, our novel experimental design enables us to document the robust positive correlation that exists between the common ratio effect and preferences for randomization. Popular behavioral models, like CPT, cannot rationalize this correlation, and there is a clear need for new research to develop superior predictive approaches. A natural path forward involves considering alternative theories. However, numerous non-EU models have already been developed over the past decades, and their ability to rationalize and predict various behaviors under risk has proven to be somewhat limited. Our paper proposes a different solution, developing a structural estimation approach that is not tied to specific economic models.

Second, our out-of-sample analysis reveals that the optimal approach for making predictions may depend on at least two factors. The first factor concerns the differences between training and test sets. Machine learning algorithms exhibit superior performance when training and test sets contain choices related to the same non-EU behavior. However, their performance significantly deteriorates in all other out-of-sample exercises we examine. The second factor pertains to the analyst's required level of detail in a prediction. Our model generates predictions that, while potentially more accurate, are less detailed than those produced by conventional predictive approaches. For example, it can only predict a range of possible certainty equivalents for a lottery. If the analyst requires more detailed predictions, our approach can still serve as a valuable complement to conventional tools. For instance, it can be used to generate a range of certainty equivalents, after which economic models or machine learning algorithms can be employed to select a value from this range.

3.2 Experimental Design

The key objective of the experiment is to evaluate the predictive performance of economic models and machine learning algorithms across different settings. To achieve this objective, we have designed an experiment involving binary choice tasks between monetary lotteries. The first two settings that we consider are different tests of the independence axiom: one is the common ratio version of the Allais paradox, and the other assesses attitudes toward mixing lotteries. Moreover, as additional settings for predictions, we elicit certainty equivalents for three binary lotteries and collect information about subjects' financial habits.[7]

Binary Choice Tasks

We elicit choices between lotteries over three monetary prizes $L < M < H$. We represent the three-outcome lottery that gives $\$L$ with probability p_L, $\$M$ with probability p_M, and $\$H$ with probability p_H as $(\$L, p_L; \$M, p_M; \$H, p_H)$. Moreover, we denote by δ_X the degenerate lottery that pays $\$X$ with certainty. The independence axiom imposes consistency requirements on choices across two or more binary choice tasks. We first assess the independence axiom via the common ratio version of the Allais paradox, which involves two types of binary choice tasks that we call CR-tasks:

CR1: $\delta_M = (\$M, 1)$ vs. $r = (\$L, 1 - p_H; \$H, p_H)$.

CR2: $0.3\delta_M + 0.7\delta_L = (\$L, 0.7; \$M, 0.3)$ vs. $0.3r + 0.7\delta_L = (\$L, 1 - 0.3p_H; \$H, 0.3p_H)$.

We use the Marschak-Machina (MM) triangle to describe graphically the lotteries in the experiment (Marschak, 1950; Machina, 1982). The left graph in Figure 3.1 illustrates the CR-tasks in the MM triangle. In the MM triangle, the probability of receiving the highest prize H is on the vertical axis, and the probability of receiving the lowest prize L is on the horizontal axis. Therefore, the generic point (p_L, p_H) in the MM triangle represents the lottery $(\$L, p_L; \$M, 1 - p_L - p_H; \$H, p_H)$. Each dashed segment connecting two lotteries indicate that there is a choice task that involves these lotteries. For instance, the black dashed segments in the left MM

[7] We preregistered the experimental design and the analysis plan at the AEA RCT Registry as AEARCTR-0011749 (Kobayashi and Lucia, 2023).

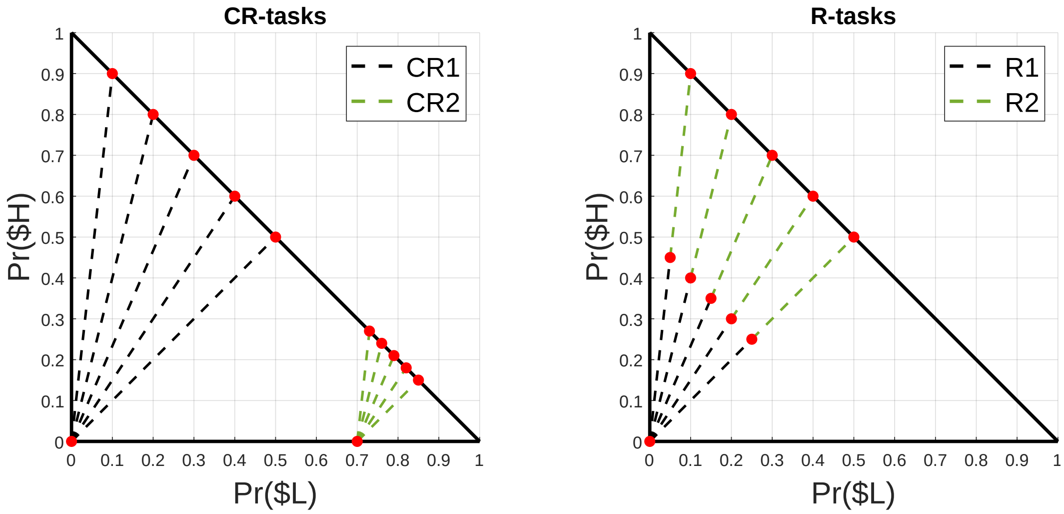


Figure 3.1: Choice Tasks.

triangle of Figure 3.1 represent CR1 choice tasks, while the green dashed segments represent CR2 choice tasks.

There are two possible scenarios in which subjects' choices in CR-tasks are incompatible with the independence axiom. The Common Ratio Effect (CRE) refers to the violation of the independence axiom in which subjects in the experiment choose lottery δ_M in CR1, and lottery $0.3r + 0.7\delta_L$ in CR2. The opposite choices in CR1 and CR2 constitute the other possible violation of the independence axiom, known as the Reverse Common Ratio Effect (RCRE).

As an additional assessment of the independence axiom, we also study subjects' attitudes towards mixtures, i.e., randomization. To this end, we consider the following two types of binary choice tasks that we refer to as R-tasks:

R1: $\delta_M = (\$M, 1)$ vs. $0.5\delta_M + 0.5r = (\$L, 0.5(1 - p_H); \$M, 0.5; \$H, 0.5p_H)$.

R2: $r = (\$L, 1 - p_H; \$H, p_H)$ vs. $0.5\delta_M + 0.5r = (\$L, 0.5(1 - p_H); \$M, 0.5; \$H, 0.5p_H)$.

The right MM triangle in Figure 3.1 represents the R1 choice tasks (depicted by black dashed segments) and the R2 choice tasks (depicted by green dashed segments). In studies exploring preferences for randomization, it is common to combine R1 and R2 into a single choice task in which subjects can select either lottery δ_M, lottery r, or a combination of the two.[8] Choosing a mixture of lotteries δ_M and r is typically interpreted as a preference for randomization. However, this approach has

[8] Agranov and Ortoleva (2022) provide an overview of this literature.

a limitation: it does not allow us to observe whether subjects exhibit aversion to randomization, meaning they prefer either of the lotteries δ_M and r over the mixture. By treating R1 and R2 as separate choice tasks, we can observe both preferences for and aversion to randomization.[9] Specifically, subjects in the experiment display a preference for randomization when they consistently choose the lottery $0.5\delta_M + 0.5r$, and aversion to randomization when they consistently reject the lottery $0.5\delta_M + 0.5r$. Both a preference for randomization and an aversion to it are behaviors that violate the independence axiom.

All subjects engage in the CR1, CR2, R1, and R2 choice tasks involving five different prize triplets (L, M, H): (0, 15, 30), (5, 15, 25), (10, 20, 30), (15, 20, 25), and (0, 10, 20). For each of these triplets, subjects undertake all types of tasks with five different probability values for the high prize p_H: 0.5, 0.6, 0.7, 0.8, and 0.9. Thus, subjects face both CR-tasks and R-tasks at precisely the same values of (L, M, H, p_H). In addition to these 100 choice tasks, the experiment includes two choice tasks in which one lottery stochastically dominates the other (referred to as FOSD choice tasks), and three additional types of choice tasks used to elicit certainty equivalents:[10]

MPL1: $(\$X, 1)$ vs. $(\$0, 0.5; \$20, 0.5)$ for $X \in \{3, \ldots, 13\}$.

MPL2: $(\$X, 1)$ vs. $(\$5, 0.5; \$25, 0.5)$ for $X \in \{8, \ldots, 18\}$.

MPL3: $(\$X, 1)$ vs. $(\$10, 0.5; \$30, 0.5)$ for $X \in \{13, \ldots, 23\}$.

We choose not to incorporate choice tasks between certain amounts and a given lottery into a list, as is typically done using the multiple price list (MPL) method. This design decision is made to minimize the amount of instruction that subjects need to understand, retaining binary choice tasks as the sole method for expressing their preferences.[11] We use the certainty equivalents elicited from MPL1, MPL2, and MPL3 choice tasks to further assess the out-of-sample accuracy of the predictions derived from the empirical analysis of the EU-core. Table 3.1 provides a summary

[9] Camerer and Ho (1994) first uses this approach to distinguish between violations of betweenness, which imposes neutrality over mixing lotteries, and violations of transitivity.

[10] We exclude from the analysis any subjects who violate first-order stochastic dominance more than once. This happens for 3 subjects only.

[11] Different procedures to elicit risk preferences may result in different observed behavior. Freeman et al. (2019) find that embedding a pairwise choice between a certain monetary amount and a risky lottery in a choice list increases the proportion of subjects choosing the risky lottery.

Table 3.1: Summary of the experimental design.

	Block 1					Block 2		
	CR1	CR2	R1	R2	FOSD	MPL1	MPL2	MPL3
# Tasks	25	25	25	25	2	11	11	11
Order Tasks	Randomized					MPL1, MPL2, MPL3		
Order Blocks	Always First					Always Second		

of our experimental design. The choice tasks in the experiment are divided into two blocks: Block 1 and Block 2. Block 1 comprises the choice tasks used to test the independence axiom (CR1, CR2, R1, and R2), along with the FOSD choice tasks. The 102 choice tasks in Block 1 are presented to subjects in a randomized order at the beginning of the experiment.

Upon completing Block 1, subjects then proceed to complete the remaining choice tasks in Block 2 (MPL1, MPL2, and MPL3), specifically designed to elicit certainty equivalents. In Block 2, subjects first encounter MPL1 tasks, followed by MPL2 tasks, and ultimately MPL3 tasks. Within each task type in Block 2, the monetary amounts are presented in ascending order.

Recruitment and Experimental Payments

We recruited 500 subjects from Prolific on July 28, 2022. The experiment was conducted using oTree. Our sample consisted of United States citizens who possessed at least a high school education and maintained a high approval rate on Prolific. For each subject, we collected information about gender, age, income, insurance purchases, and investment behavior.

Each subject received $5.50 upon completing the experiment. Additionally, every subject had a one-in-six chance of being selected to receive an additional bonus payment based on their decisions during the study. Out of the 135 choice tasks, each carried an equal probability of determining the bonus payment amount. Specifically, subjects received the realized amount from the lottery they chose in the randomly selected choice task.[12]

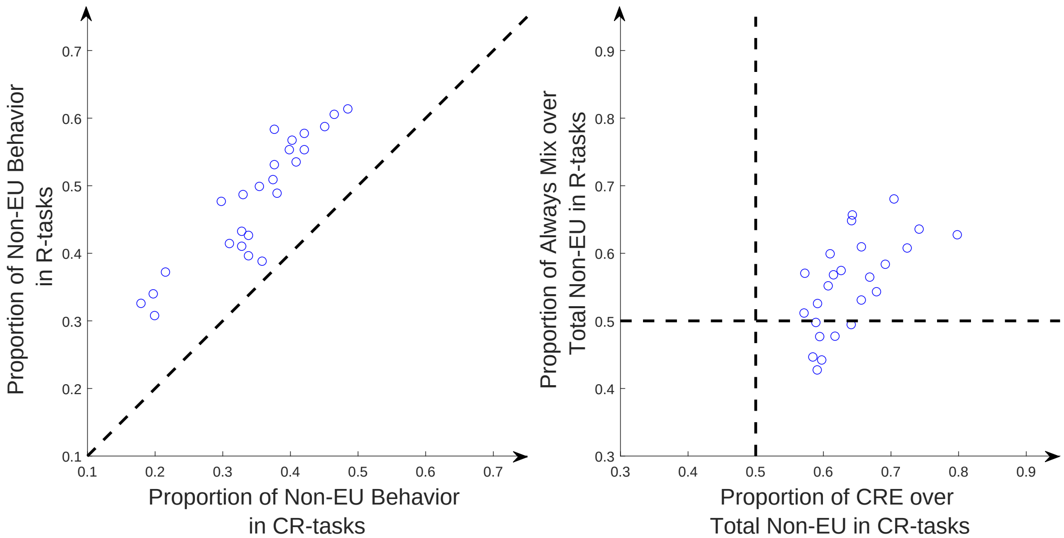


Figure 3.2: Correlation between the overall proportions of non-EU behaviors (left graph) and between the relative proportions of specific non-EU behaviors (right graph) across CR-tasks and R-tasks. Each observation corresponds to a single parameterization (L, M, H, p_H).

3.3 Main Experimental Findings

In this section, we explore the non-EU behaviors observed in CR-tasks and R-tasks, along with the correlation between them. The left graph in Figure 3.2 illustrates the proportion of non-EU choice patterns in CR-tasks (on the x-axis) and R-tasks (on the y-axis) for each pair of lotteries (δ_M, r) in the experiment.[13] Each observation corresponds to a single parameterization (L, M, H, p_H). Given that there are five triplets of prizes in the experiment and five values of p_H for each triplet, there are 25 observations graphed. All observations are situated to the left of the 45-degree dashed line, indicating that non-EU behavior is more prevalent in R-tasks than in CR-tasks for all pairs of lotteries. Moreover, there is a strong positive correlation between non-EU behaviors in CR-tasks and R-tasks, evidenced by a correlation coefficient of 0.9104. This high correlation is noteworthy because it implies that observing EU failures in specific contexts, such as tests of the CRE, may provide insights into other potential EU failures of interest.

Next, we turn to analyzing which of the non-EU behaviors are more likely to be observed in our experiment. For each pair of lotteries (δ_M, r), two potential non-EU choice patterns exist in both CR-tasks and R-tasks. In the context of CR-tasks, these

[12] The complete instructions with screenshots from the experiment are presented in Appendix C.5.

[13] Appendix C.1 provides a comprehensive description of EU and non-EU choice patterns in CR-tasks and R-tasks.

patterns are represented by the CRE and the RCRE. Conversely, within R-tasks, the two non-EU choice patterns manifest as always choosing the mixture (AM) and never choosing the mixture (NM).

The right graph in Figure 3.2 displays the proportion of CRE choice patterns over the total non-EU behaviors in CR-tasks on the x-axis, and the proportion of AM choice patterns over the total non-EU behaviors in R-tasks on the y-axis. Two important observations arise from Figure 3.2. First, CRE and AM are the most frequent non-EU choice patterns for the majority of pairs of lotteries in the experiment. At the same time, the emergence of NM choice patterns is non-negligible for many pairs of lotteries. This suggests that experiments which do not account for the elicitation of aversion to randomization might neglect a critical dimension in the analysis of attitudes towards randomization. The second fact documented in Figure 3.2 is the strong positive correlation between the prevalence of the CRE and of AM choice patterns, with a correlation coefficient of 0.6311.

Result 1 *Non-EU behaviors in CR-tasks and R-tasks are strongly positively correlated, with a correlation coefficient of 0.9104. Among non-EU behaviors, the CRE and AM are the two most frequently observed ones in CR-tasks and R-tasks, respectively. Moreover, these two non-EU behaviors are strongly positively correlated, with a correlation coefficient of 0.6311.*

Out-of-Sample Predictions: Conventional Toolbox

We now investigate whether leading economic models and machine learning algorithms can predict the observed emergence of the two non-EU behaviors and correlation between them. As popular economic models, we analyze EU model and Cumulative Prospect Theory (CPT), which aims to rationalize non-EU behaviors through probability weighting. In addition, we examine the predictive performance of gradient boosting trees (GBT) and neural networks (NN), which are two common machine learning algorithms used for classification tasks. We contrast the performance of EU, CPT, GBT, and NN through two out-of-sample exercises.[14] In the first, we separately analyze behavior in CR-tasks and R-tasks. For each task type, we assess the out-of-sample performance of various methods through cross-validation. In the second exercise, we use choices from either CR-tasks or R-tasks as the training data and aim to predict choices in the alternate tasks.

[14] Details on the procedures we followed to estimate economic models and train machine learning algorithms can be found in Appendix C.2. Instead, Appendix C.3 reports the values of the estimated parameters under EU and CPT.

Table 3.2: Out-of-sample exercises within CR-tasks and R-tasks: predict choice patterns.

Exercise	Pattern	Models			
		EU	CPT	GBT	NN
Train: 80% CR-tasks	Choice	50.62%	46.01%	**63.46%**	46.30%
Test: 20% CR-tasks	Patterns	(0.82)	(1.13)	(1.23)	(0.78)
Train: 80% R-tasks	Choice	40.95%	47.06%	**60.14%**	38.74%
Test: 20% R-tasks	Patterns	(1.73)	(1.22)	(0.56)	(1.06)
Combined	Choice	45.78%	46.54%	**61.80%**	42.39%
	Patterns	(5.25)	(1.24)	(1.97)	(4.22)

Notes: Percentages of correctly classified choice patterns. Standard deviations in parentheses.

For all economic models, we estimate mixture models to account for heterogeneity. We also provide information about each subjects to machine learning algorithms including individual fixed effects in the training data. We use as main metric to evaluate these methods the percentages of choice patterns that they correctly classify. Because these methods provide probabilistic predictions, we identify the predicted choice patterns as the ones with the highest associated predicted probability.[15]

Cross Validation within CR-tasks and R-tasks

We employ a 5-fold cross-validation, treating data from CR-tasks and R-tasks independently. To divide the available data into five equally sized subsets, we use the following randomized procedure: For each subject and every triplet of prizes in the experiment, we randomly allocate all choice tasks associated with one of the five possible probabilities for the high prize p_H in the risky lottery to each of the five subsets. In doing so, we ensure that each iteration of the cross-validation procedure contains sufficient information about all subjects within both training and test sets. This approach enables accounting for heterogeneity in preferences during the training phase and leverages this heterogeneity to formulate predictions.

Table 3.2 presents the average percentages of choice patterns that are accurately classified across the five iterations of the cross-validation exercises within CR-tasks and R-tasks, with standard deviations in parentheses. The first two rows display the

[15] We refer to Appendix C.4 for a more comprehensive probabilistic and deterministic evaluations of all predictive tools. The findings presented in this section are qualitatively similar to the ones obtained in Appendix C.4.

results for the two cross-validation exercises separately, while the last row provides the aggregate results from both exercises. GBT clearly emerges as the winner from this out-of-sample analysis, delivering markedly superior performance compared to both NN and economic models.

Result 2 *GBT surpasses all other approaches in cross-validation exercises within CR-tasks and R-tasks, achieving an average percentage of correctly predicted choice patterns close to 61%. For context, CPT, the runner-up approach, showcases a percentage of correctly predicted choice patterns that is approximately 15% lower.*

Out-of-Sample Predictions across CR-tasks and R-tasks

We conduct two distinct exercises: In the first, CR-tasks serve as the training data and R-tasks as the test data. In the second exercise, conversely, we reverse the roles, employing R-tasks for training and CR-tasks for testing. The percentages of correctly classified choice patterns from these exercises are presented in Table 3.3. The first two rows of Table 3.3 display the results for the two exercises separately, while the last row summarizes the average performance across both exercises.

The descriptive analysis of behaviors in CR-tasks and R-tasks highlight systematic violations of EU. Yet, in these out-of-sample exercises, EU has the best overall out-sample performance. CPT performs worse than EU because it predicts a correlation between non-EU behaviors that is opposite to what we observe in our experiment. In particular, CPT generally rationalizes the CRE through probability weighting. However, probability weighting also implies aversion to randomization for the lotteries in the experiment, while we observe preferences for randomization as the most frequent non-EU behavior.

Machine learning algorithms also demonstrate notably poor performance, with percentages of correctly classified choice patterns around 30% lower than those achieved in cross-validation exercises within CR-tasks and R-tasks. The primary distinction between out-of-sample exercises within a specific type of task and those across tasks is that in the former, lotteries in the training and test sets are identical, while in the latter, they are marginally different. Specifically, choices in R-tasks involve mixture lotteries with three prizes, whereas all lotteries in CR-tasks feature one or two possible prizes. In general, one might expect machine learning algorithms to perform less effectively as the differences between training and test sets increase. However, the strikingly different performance of these approaches in out-of-sample

Table 3.3: Out-of-sample exercises across CR-tasks and R-tasks: predict choice patterns.

Exercise	Prediction	Models			
		EU	CPT	GBT	NN
Train: CR-tasks Test: R-tasks	Choice Patterns	40.72%	**42.70%**	35.71%	35.64%
Train: R-tasks Test: CR-tasks	Choice Patterns	**48.59%**	46.29%	34.16%	21.79%
Combined	Choice Patterns	**44.66%** (5.56)	44.50% (2.54)	34.93% (1.10)	28.72% (9.79)

Notes: Percentages of correctly classified choice patterns. Standard deviations in parentheses.

exercises across choices with very similar lotteries raise concerns about their ability to produce generalizable predictions.

The economic models and the machine learning algorithms that we consider fail in the out-of-sample exercises that are most interesting from an economic point of view. If a predictive approach is not capable of predicting different behaviors within the same experiment, then there is little hope that it can be used as guidance for predicting real world behaviors of interest. The lack of generalizability of conventional predictive approaches calls for the development of new empirical strategies to make predictions.

Result 3 *In cross-validation exercises across CR-tasks and R-tasks, economic models significantly outperform machine learning algorithms. Both EU and CPT allow for approximately 44.5% of choice patterns to be correctly classified across the two out-of-sample exercises. GBT, while achieving the best performance among machine learning algorithms, has a percentage of correctly classified choice patterns that is, on average, around 10% below that of economic models.*

3.4 Empirical Framework

In this section, we begin by reviewing the classic empirical approach used to estimate economic models from choice data. Next, we introduce the notion of the EU-core, and we detail the empirical strategy that we developed for estimating it. Finally, we present our estimation results, which will then be used in Section 3.5 to make out-of-sample predictions.

Classic Estimation Approach

In the classic empirical framework proposed by Hey and Orme (1994), a decision-maker chooses lottery p over lottery q if

$$V(p,q) + \epsilon \geq 0,$$

where $V(p,q)$ is a quantity that is greater or equal to zero if the decision-maker prefers lottery p over lottery q, and ϵ is an error term, assumed to be normally distributed with a mean of zero and a variance of σ. The specific functional form of $V(p,q)$ depends on the assumptions made about the decision-making model. For instance, under EU, $V(p,q)$ represents the difference in expected utilities between lottery p and lottery q.

Within this framework, the analyst must select a decision model and also make parametric assumptions within that chosen model. For example, under EU, a common assumption is that the decision-maker operates with a constant relative risk aversion (CRRA) utility function. With this assumption, the analyst can estimate the free parameter of the CRRA utility function and the variance of the error term using choice data. The estimated model can then be used to undertake counterfactual exercises, such as predicting choices between lotteries in an alternative dataset.

The analysis in Section 3.3 highlights the risks associated with specifying everything about a decision model. For instance, CPT rationalizes the CRE with probability weighting. However, probability weighting also implies aversion to randomization in the experiment, which is inconsistent with what we find. At the same time, the failures of machine learning algorithms in out-of-sample exercises across different types of tasks also emphasize the potential benefit of predictive approaches that rely on an underlying economic structure.

The methodological question we address is: can we generalize the empirical framework of Hey and Orme (1994) to predict behavior under risk without committing to specific decision models? The empirical approach that we propose as an answer to this question builds on the theoretical notion of EU-core.

The EU-core

Given any reflexive, transitive, and continuous preference relation $\succsim$, its EU-core is the subrelation[16] $\succsim^*$ such that for all lotteries p, q, r and for all $\lambda \in (0,1]$,

$$p \succsim^* q \Leftrightarrow \lambda p + (1-\lambda)r \succsim \lambda q + (1-\lambda)r.$$

[16] $\succsim^*$ is a subrelation of $\succsim$ if for all lotteries p and q, $p \succsim^* q$ implies $p \succsim q$.

That is, $p \succsim^* q$ whenever the decision-maker prefers p to q and mixing both lotteries with a third common lottery r does not affect the relative preferences of the decision-maker between p and q.[17] Cerreia-Vioglio (2009) proves that $\succsim^*$ is the greatest subrelation of $\succsim$ that satisfies the independence axiom; that is, if $\succsim^{**}$ is another subrelation of $\succsim$ that satisfies the independence axiom, then $\succsim^{**}$ is a subrelation of $\succsim^*$. If a preference $\succsim$ violates the independence axiom, then its EU-core is an incomplete preference relation and admits a multi-utility representation. In particular, there exists a set of utilities $\mathcal{W}$ such that for all lotteries p and q, we have $p \succsim^* q$ if and only if the difference in expected utilities between p and q is non-negative for all utilities within the set $\mathcal{W}$.

Our empirical approach involves obtaining information about a preference by estimating the set of utility functions that represent its EU-core.[18] In this way, we can obtain estimates and generate predictions that do not rely on specific decision models.

Econometric Specification

We detail our empirical framework in the context of our experimental design. Our study involves lotteries over a finite set of K monetary prizes $X = \{x_1, \dots, x_K\}$, with $x_1 < x_2 < \cdots < x_K$. We consider a set of L utility functions $\mathcal{W} = \{v_1, \dots, v_L\}$, each utility $v_l \colon X \to \mathbb{R}$ is representable as a vector with its k-th component, v_{lk}, being equal to $v_l(x_k)$. We restrict our attention to normalized sets of utilities, setting $v_{11} = \cdots = v_{L1} = 0$ and $v_{1K} = \cdots = v_{LK} = 1$.[19] This means all utilities assign zero to the worst outcome x_1 and one to the best outcome x_K. Moreover, we assume all utilities are weakly increasing; i.e., $v_{l1} \leq v_{l2} \leq \ldots \leq v_{lK}$ for each $v_l \in \mathcal{W}$.

We define $I = \{1, \dots, N\}$ as a set of subjects in our experiment, $\triangle(X)$ as the set of lotteries over X, and by $\mathcal{D} \subseteq \triangle(X)^2$ as a subset of pairs of lotteries where the subjects express their preferences. An empirical analysis of the EU-core requires evaluating whether it holds that $p \succsim_i^* q$ or $q \succsim_i^* p$ for each subject $i \in I$ and each pair of lotteries $(p, q) \in \mathcal{D}$. In the traditional estimation framework, where the objective is to estimate a subject's preferences, the choices made by the subjects

[17]We study the expected utility core by considering only "one-stage" lottery mixtures, rather than two-stage compound lotteries. In other words, we focus on mixture independence, rather than compound independence, as defined in Segal (1990).

[18]This set of utility functions is unique up to the closed convex hull. Our empirical approach aims to estimate the extreme points of the convex set of utility functions that represent the EU-core.

[19]In our estimation procedure, the number of utilities L is an hyperparameter that can be chosen using standard model selection techniques.

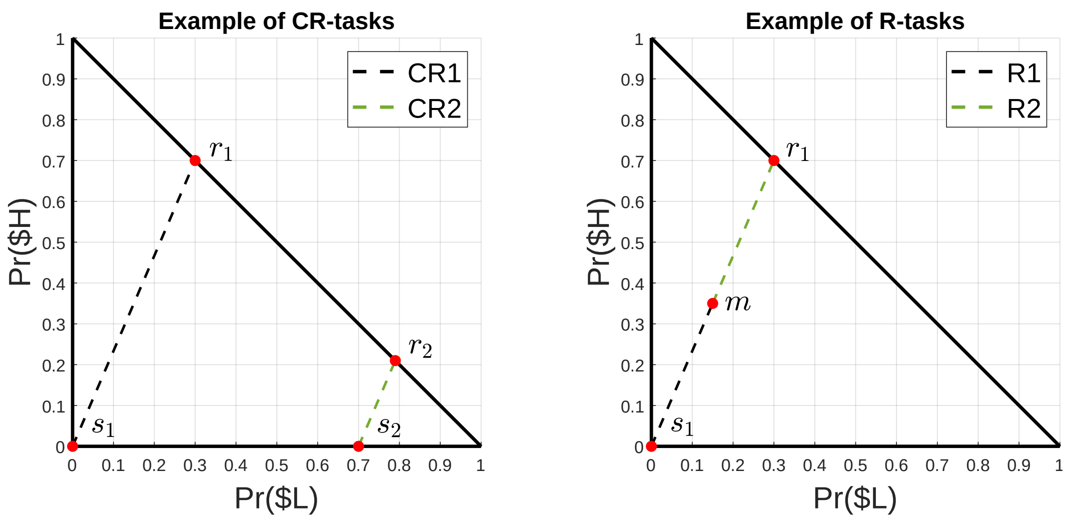


Figure 3.3: Examples of CR-tasks and R-tasks.

can be used directly as inputs for the estimation. However, when the focus of the estimation shifts from a preference relation to its EU-core, additional information becomes necessary. Specifically, we need to assess whether the choices made by the subjects signal a violation of the independence axiom.

By observing subjects' choices in experimental settings that test the independence axiom, we construct an index, $Core_i$, for each subject i as follows: for each pair of lotteries $(p, q) \in \mathcal{D}$,

$$Core_i(p, q) := \begin{cases} 3 & \text{if there is no experimental evidence against } p \succsim_i^* q \\ 2 & \text{if there is no experimental evidence against } q \succsim_i^* p \\ 1 & \text{otherwise.} \end{cases}$$

We construct two distinct versions of the index $Core_i$ within the context of our experiment: $Core_i^{CR}$ and $Core_i^{R}$. Each version builds upon different sets of information. Specifically, $Core_i^{CR}$ uses data solely from CR-tasks, while $Core_i^{R}$ is based on R-tasks only. We explain how we construct these indexes using the examples from Figure 3.3. The left and right graphs in Figure 3.3 respectively depict the CR-tasks and R-tasks associated with a generic pair of lotteries (s_1, r_1) from our experiment.

The absence of experimental evidence contradicting $s_1 \succsim_i^* r_1$ in CR-tasks means that subject i chooses for lottery s_1 in the CR1 task and lottery s_2 in the CR2 task. Observing this choice pattern, we assign the value of 3 to the index $Core_i^{CR}(s_1, r_1)$. In a parallel manner, we find no evidence refuting $r_1 \succsim_i^* s_1$ in the CR-tasks if subject

i chooses lottery r_1 in the CR1 task and lottery r_2 in the CR2 task. In this case, we assign to the index $Core_i^{CR}(s_1, r_1)$ the value of 2. The two remaining choice patterns are the CRE and the RCRE, and are both incompatible with EU. If we observe either the CRE or the RCRE for subject i, we assign to the index $Core_i^{CR}(s_1, r_1)$ the value of 1.

Turning to R-tasks, we find no evidence against $s_1 \succsim_i^* r_1$ if subject i chooses lottery s_1 in the R1 task and lottery m in the R2 task. When this happens, we assign the value of 3 to the index $Core_i^R(s_1, r_1)$. Similarly, if subject i chooses lottery m in the R1 task and lottery r_1 in the R2 task, we have no evidence against $r_1 \succsim_i^* s_1$. In this case, we assign the value of 2 to the index $Core_i^R(s_1, r_1)$. The two remaining non-EU choice patterns are AM and NM. If we observe either AM or NM for subject i, we assign to the index $Core_i^R(s_1, r_1)$ the value of 1.

We define $V(p, q; v_l)$ as the difference in expected utilities between lottery p and lottery q, given a Bernoulli utility function v_l. For each subject $i \in I$, utility $v_l \in \mathcal{W}$, and comparison $(p, q) \in \mathcal{D}$, we associate an error term $\epsilon_{i,l,(p,q)}$. We assume that the vector of error terms $[\epsilon_{i,1,(p,q)}, \ldots, \epsilon_{i,L,(p,q)}]$ across utilities follows a multivariate normal distribution with mean $[0, \ldots, 0] \in \mathbb{R}^L$ and covariance matrix $\Sigma \in \mathbb{R}^{L \times L}$. For any two lotteries p and q, and for any subject i, our empirical framework postulates that

$$Core_i(p, q) = 3 \Leftrightarrow V(p, q, v_l) - \epsilon_{i,l,(p,q)} \geq 0, \text{ for all } l \in \{1, \ldots, L\},$$

and

$$Core_i(p, q) = 2 \Leftrightarrow V(p, q, v_l) - \epsilon_{i,l,(p,q)} < 0, \text{ for all } l \in \{1, \ldots, L\}.$$

In other words, we postulate to find no evidence against $p \succsim_i^* q$ whenever the difference in expected utilities between lotteries p and q, minus an error term, is non-negative for all utilities. Similarly, we expect to find no evidence against $q \succsim_i^* p$ whenever the opposite condition holds.

Our flexible formulation of the error structure extends the normality assumption of the unique error term in Hey and Orme (1994), allowing us to account for potential noise in the $Core_i$ index that might arise from several sources. First, we construct this index by observing the choices of subject i in experimental settings that test the independence axiom. If these choices are noisy, then the resulting $Core_i$ index will also be noisy. Additionally, even in the absence of noise in the choices, the $Core_i$ index might still be noisy due to issues with missing data. For example, we might find no evidence against $p \succsim_i^* q$ simply because we could not observe enough choices involving lotteries p and q.

To account for variation in preferences across subjects, we employ a mixture model and postulate that each subject i belongs to one of C possible different groups (Bruhin et al., 2010). We denote by v_l^c the l-th utility in group c and by Σ_c the covariance matrix in group c, with $c \in \{1, \ldots, C\}$. Within this framework, the probability that we find no experimental evidence against $p \succsim_i^* q$ if subject i belongs to group c is:

$$\Pr(Core_i(p,q) = 3 \mid v_1^c, \ldots, v_L^c, \Sigma_c) = \Phi\left(V(p,q;v_1^c), \ldots, V(p,q;v_L^c); \Sigma_c\right),$$

where Φ represents the cumulative distribution function of the mean-zero multivariate normal distribution. Similarly, the probability that we find no experimental evidence against $q \succsim_i^* p$ if subject i belongs to group c is:

$$\Pr(Core_i(p,q) = 2 \mid v_1^c, \ldots, v_L^c, \Sigma_c) = \Phi\left(-V(p,q;v_1^c), \ldots, -V(p,q;v_L^c); \Sigma_c\right).$$

Given the observed index $Core_i(p,q)$ for all pairs of lotteries $(p,q) \in \mathcal{D}$, we denote by $f(Core_i; v_1^c, \ldots, v_L^c, \Sigma_c)$ the likelihood function for subject i belonging to group c:

$$\begin{aligned}
&\prod_{(p,q)\in\mathcal{D}} \Big(\mathbb{1}\left(Core_i(p,q) = 3\right) \cdot \Pr(Core_i(p,q) = 3 \mid v_1^c, \ldots, v_L^c, \Sigma_c) \\
&+ \mathbb{1}\left(Core_i(p,q) = 2\right) \cdot \Pr(Core_i(p,q) = 2 \mid v_1^c, \ldots, v_L^c, \Sigma_c) \\
&+ \mathbb{1}\left(Core_i(p,q) = 1\right) \cdot (1 - \Pr(Core_i(p,q) = 3) - \Pr(Core_i(p,q) = 2))\Big).
\end{aligned}$$

Let π_c represent the probability of a subject belonging to group type c. The log-likelihood of the finite mixture model is given by:

$$\sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c f(Core_i; v_1^c, \ldots, v_L^c, \Sigma_c),$$

where the first sum is over subjects and the second sum is over groups.

To sum, for each group of subjects c, structural parameters are utility functions $\{v_1^c, \ldots, v_L^c\}$, covariance matrices Σ_c, and probability of group member ship π_c. We estimate these parameters through maximum likelihood estimation by using the log-likelihood given above.

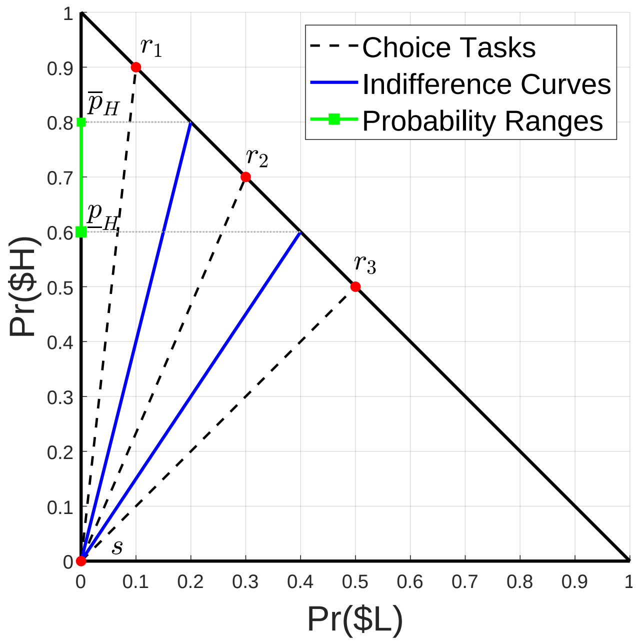


Figure 3.4: Set of utility functions in the MM-triangle.

Model Implications: Risk Aversion and Adherence to EU

By using an example, we explain the behavioral implications of our empirical framework. Figure 3.4 illustrates three generic choice tasks in the MM-triangle: (s, r_1), (s, r_2), and (s, r_3). To obtain predictions from any decision model in these tasks, one only needs to know the shape of the indifference curve passing through lottery s. Lotteries to the left of this curve will be strictly better than s, while those to the right will be strictly worse.

Under our decision model, each subject makes their decision based on not just one but a set of utility functions. This means that there are multiple indifference curves passing through a given lottery, each corresponding to a different utility function. For instance, the blue solid segments in Figure 3.4 represent two indifference curves passing through lottery s. In this new setting, any lottery r that lies to the left of both indifference curves will be unambiguously preferred to s. Specifically, for every such lottery, we can conclude that $r \succsim^* s$, as is the case for lottery r_1 in Figure 3.4. Conversely, any lottery r that lies to the right of both indifference curves will be unambiguously worse than s. In particular, for every such lottery, we can conclude that $s \succsim^* r$, as is the case for lottery r_3 in Figure 3.4. For all other lotteries, without making further assumptions, the only conclusion we can draw is that neither $s \succsim^* r$ nor $r \succsim^* s$ holds.

More generally, given a set of utilities $\mathcal{W}$ and any lottery r on the hypotenuse of the

MM-triangle assigning probability p_H to the high prize H and probability $1 - p_H$ to the low prize L, we can compute the range of probabilities $[\underline{p}_H, \overline{p}_H]$, where

$$\underline{p}_H \coloneqq \max\{p_H \in [0,1] \ : \ v(M) \geq p_H v(H) + (1 - p_H)\, v(L) \ \text{ for all } v \in \mathcal{W}\},$$

and

$$\overline{p}_H \coloneqq \min\{p_H \in [0,1] \ : \ p_H v(H) + (1 - p_H)\, v(L) \geq v(M) \ \text{ for all } v \in \mathcal{W}\}.$$

These two quantities are well-defined because utilities are assumed to be weakly increasing and continuous. In particular, $\underline{p}_H$ is the highest probability of the high prize for which we have $s \succsim^* r$. Similarly, $\overline{p}_H$ is the lowest probability of the high prize for which we have $r \succsim^* s$. The green squares in Figure 3.4 describe the two points $\underline{p}_H$ and $\overline{p}_H$ given the two indifference curves.

The range of probabilities $[\underline{p}_H, \overline{p}_H]$ is a measure for the extent to which a subject adheres to EU. In general, the narrower the range of probabilities, the more consistent the underlying preference is with EU. Under EU, the range would collapse into a fixed probability p_{EU}. Moreover, the range of probabilities $[\underline{p}_H, \overline{p}_H]$ provides information about subjects' risk attitudes. Specifically, in our experiment, the mid-value prize M is always set as the mean of the high prize H and the low prize L. Consequently, given a triplet of prizes, we can classify an EU subject as risk averse if $\underline{p}_H > 0.5$, risk-seeking if $\overline{p}_H < 0.5$, and as neither risk averse nor risk-seeking otherwise. In this way, we generalize the empirical analysis of risk attitude under EU to preferences that may violate the independence axiom.[20] In the special case of EU with a unique probability p_{EU}, we would classify a subject as risk averse if $p_{EU} > 0.5$, risk neutral if $p_{EU} = 0.5$, and risk-seeking if $p_{EU} < 0.5$.

Estimation Results

We present the estimation results from a mixture model with three groups of subjects and two utilities for each group. The estimates are derived using data from both CR-tasks and R-tasks.[21] The three graphs on the left in Figure 3.5 show the estimated utilities in each group, while the three graphs on the right display the ranges of probabilities $[\underline{p}_H, \overline{p}_H]$ that they induce for the five triplets of prizes in our experiment.

The two utilities in Group 1 are very close to each other, indicating that the behavior of subjects in Group 1 can be accurately described by EU. Furthermore, both utilities

[20] Our classification adopts the aversion to mean-preserving spreads as the primitive notion for risk aversion (Rothschild and Stiglitz, 1970).

[21] In particular, we used both the $Core_i^{CR}$ and $Core_i^{R}$ indices as input for the estimation.

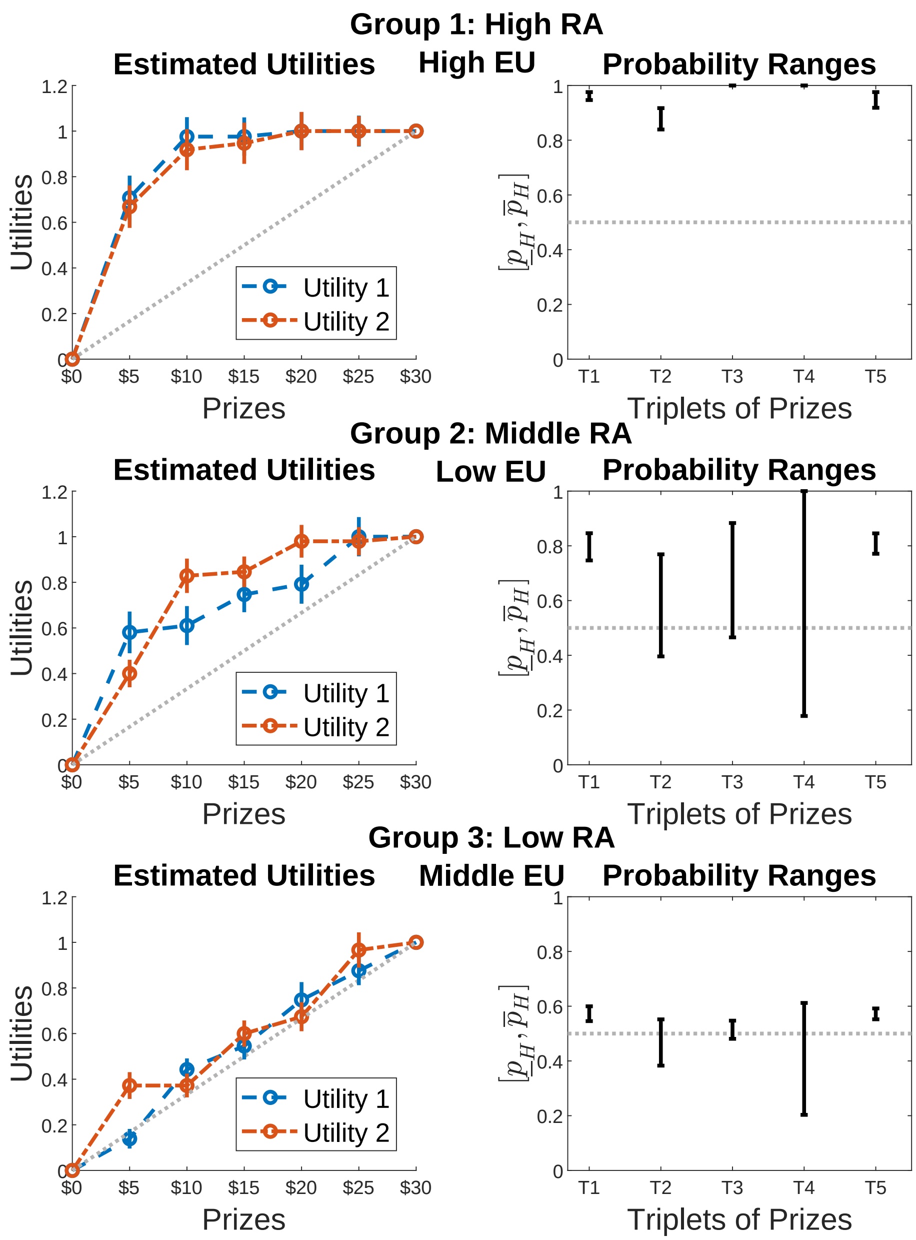


Figure 3.5: Estimated utilities and probability ranges.

Notes: We classify the three groups in terms of risk aversion (RA) and adherence to EU. Estimated group membership probabilities are: 0.2798 for Group 1, 0.4834 for Group 2, and 0.2368 for Group 3. The width of the vertical bars centered around the point estimates in the left graphs indicates the bootstrapped standard errors. We report the values of all estimated parameters in Appendix C.3. In the right graphs, T1 through T5 correspond to prize triplets as follows: T1 = ($0, $15, $30), T2 = ($5, $15, $25), T3 = ($10, $20, $30), T4 = ($15, $20, $25), and T5 = ($0, $10, $20).

in Group 1 are concave and significantly deviate from being linear, pointing to strong levels of risk-aversion. This information is reflected in the range of probabilities induced by the two utilities. Specifically, the ranges across the five triplets of prizes are narrow, indicating a high adherence to EU. Additionally, these ranges encompass very high values for the probability of winning the high prize. For instance, in all the triplets of prizes in our experiment, the two estimated utilities both rank the middle prize above a binary risky lottery that offers the high prize with a probability of approximately 0.8, or the low prize with the complementary probability. Therefore, Group 1 reflects subjects in the experiment who systematically opted for the safer available lottery. The estimated proportion of subjects belonging to Group 1 is 0.2798.

Moving to Group 2 and Group 3, the utilities within these groups are more distinct compared to those in Group 1. The difference in utilities is most pronounced in Group 2, which is evident from the broader range of probabilities that they induce. Interestingly, utilities in both Group 2 and Group 3 are neither strictly concave nor convex. As a result, the range of probabilities in these groups spans values both below and above 0.5, which represents the expected utility threshold for risk-aversion in our experiment. This suggests that subjects' behavior in these two groups doesn't strictly align with either pure risk-aversion or risk-seeking tendencies. However, the probability ranges in Group 2 consistently register higher values than those in Group 3, indicating relatively more risk-averse behavior in comparative terms. The estimated proportion of subjects belonging to Group 2 and 3 are 0.4834 and 0.2368, respectively.

The emergence of non-EU behavior and risk-averse tendencies consistently differs within the groups across the five triplets of prizes in the experiment. Notably, subjects consistently exhibit stronger risk aversion when the lowest prize in the triplet is \$0 compared to when it's a positive amount. On the whole, their behavior aligns more closely with EU. These observations are particularly evident in the probability ranges for the triplets $T1 = (\$0, \$15, \$30)$ and $T5 = (\$0, \$10, \$20)$, which are narrower and encompass higher values than other triplets. Among all the triplets, $T4 = (\$15, \$20, \$25)$ stands out as the one with the most pronounced non-EU behavior across all three groups. This triplet is unique in having prizes that are closely spaced, which could be linked to higher noise in responses.[22]

[22] A potential extension of our model might allow the variance of the error term to be influenced by features of the lotteries, such as the prize amounts.

Result 4 *In the mixture model for the EU-core with 3 groups estimated using both CR-tasks and R-tasks, approximately 28% of subjects (Group 1) are extremely risk-averse and align closely with EU. The remaining subjects (Group 2 and Group 3) deviate markedly from EU and cannot be categorically classified as either risk-averse or risk-seeking.*

3.5 Out-of-Sample Predictions: EU-Core Analysis

In this section, we evaluate whether, and under which conditions, the out-of-sample predictions derived from the analysis of the EU-core are more precise than those obtained from specific economic models or machine learning algorithms. We examine various out-of-sample exercises, each involving different degrees of similarity between training and test sets. Initially, we reexamine the two out-of-sample exercises related to choices in CR-tasks and R-tasks, previously discussed in Section 3.3. In these exercises, training and test sets bear a close relationship. Subsequently, we use all choices from CR-tasks and R-tasks as training data and aim to predict the certainty equivalents derived from choices in Block 2 of the experiment. Lastly, we explore the correlation between estimated levels of risk aversion and adherence to EU with financial habits outside of the experiment.

Out-of-Sample Exercises Within and Across CR-Tasks and R-Tasks

Both economic models and machine learning algorithms can be used to predict the probability of all conceivable choice patterns in CR-tasks and R-tasks. On the other hand, the EU-Core approach only allows us to predict the probability of the possible values of the index $Core$. As a result, we cannot distinguish between non-EU behaviors within CR-tasks and R-tasks. Specifically, we cannot differentiate between CRE and RCRE in CR-tasks, and between AM and NM in R-tasks. To implement a fair comparison that takes into account the different levels of prediction detail attainable by various approaches, we also assess the ability of economic models and machine learning algorithms to predict the index $Core$. Hence, we deem an observed non-EU choice pattern as correctly predicted by these approaches as long as they assign the highest predicted probability to one non-EU behavior, even if the predicted non-EU behavior does not match the observed one.

Table 3.4 shows the percentages of choice patterns with a correctly classified index $Core$ for economic models, machine learning algorithms, and the EU-Core approach. Consistent with the findings in Section 3.3, GBT achieves the best results in out-of-sample exercises within CR-tasks and R-tasks. Our approach performs

Table 3.4: Out-of-sample exercises: adherence to EU.

Exercise	Prediction	Models				
		EU	CPT	GBT	NN	EU-Core
Combined Within Tasks	Index *Core*	45.78% (5.25)	47.73% (1.06)	**64.40%** (1.50)	45.10% (3.93)	60.11% (1.97)
Combined Across Tasks	Index *Core*	44.66% (5.56)	44.94% (2.93)	41.59% (7.09)	42.23% (11.70)	**48.64%** (3.32)

Notes: Percentages of choice patterns with a correctly classified index *Core*. Standard deviations in parentheses. The first row ("Combined Within Tasks") presents the average percentages for the two cross-validation exercises within CR-tasks and R-tasks. For these exercises, we used the same partition of data used in Section 3.3 to assess the predictive accuracy of economic models and machine learning algorithms at predicting specific choice patterns. The second row ("Combined Across Tasks") displays the average percentages for the two out-of-sample exercises across CR-tasks and R-tasks described in Section 3.3.

significantly better than both EU and CPT, while it is around 4% less accurate than GBT. Moreover, in line with the findings in Section 3.3, the ranking between economic models and machine learning algorithms is inverted once we switch from out-of-sample exercises within tasks to across tasks.[23] In this latter type of out-of-sample exercise, the EU-core approach outperforms all other methods, yielding approximately 4% higher accuracy than economic models.

Result 5 *The EU-Core outperforms economic models in out-of-sample exercises within CR-tasks and R-tasks, though it is around 4% less accurate than GBT. Conversely, the EU-core approach attains the most accurate results in out-of-sample exercises across CR-tasks and R-tasks, delivering predictions that are approximately 4% more accurate than those of economic models.*

Certainty Equivalents

We use choices from CR-tasks and R-tasks as training data and evaluate the accuracy of different approaches in predicting the certainty equivalents inferred from choices in Block 2 of our experiment. In Block 2, subjects are asked to compare three risky lotteries and various certain prizes. We focus on the subset of observations where subjects shifted their preference between a fixed lottery and a certain amount at most

[23]Here, CPT and NN do relatively better than EU and GBT, respectively. This is a result of the lenient approach we are using in evaluating these methods. For example, we categorize a choice pattern as correctly classified even when we observe a preference for randomization, while CPT predicts an aversion to randomization.

once.[24] For lotteries where subjects make a single switch, we compute the certainty equivalent as the smallest certain amount preferred over the lottery, reduced by \$0.5. If a lottery is chosen over all the certain prizes, its certainty equivalent is computed as the highest prize in the experiment compared to that lottery. Conversely, if the certain prize is always preferred, the certainty equivalent is computed as the smallest prize in the experiment compared to the lottery.

The EU-core model enables us to predict a range of certainty equivalents. Specifically, given an estimated set of utilities $\hat{W}$, the certainty equivalent for a lottery p is predicted to lie within:

$$[\min_{v\in\hat{W}} c(p,v), \max_{v\in\hat{W}} c(p,v)],$$

where $c(p,v)$ represents the certainty equivalent of lottery p determined using utility function v.

The estimation results presented in Section 3.4 highlight significant heterogeneity in preferences. In particular, we estimated mixture models for three distinct groups of subjects and ranked these groups in terms of risk-aversion and non-EU behavior. The greater the risk aversion, the higher the possible values for the risk premium associated with each lottery.[25] Additionally, increased non-EU behavior implies broader possible ranges of risk premia. Figure 3.6 summarizes with box plots the distribution of risk premia of all lotteries from Block 2 for the three groups of subjects. To construct this graph, we assigned each subject to the group with the highest group membership probability.

Comparing the distribution of risk premia across groups, we observe that the risk premia increase with the predicted level of risk aversion. For instance, the median risk premium in the group with low predicted levels of risk aversion is 0.5, while it is 5.5 in the group with high predicted levels of risk aversion. Furthermore, our model not only accurately predicts the differences in risk premia levels across groups but also the levels within each group. The dashed red lines in Figure 3.6 illustrate the predicted ranges of risk premia in the three groups. In all the groups, the predicted ranges of risk premia are perfectly consistent with the observed ones. In particular, for each of the three lotteries, the predicted range always includes the observed median risk premium.

[24] Appendix C.1 summarizes the distribution of risk premia for the three lotteries presented to subjects in Block 2.

[25] The risk premium of a lottery is the difference between the expected value and the certainty equivalent of a lottery.

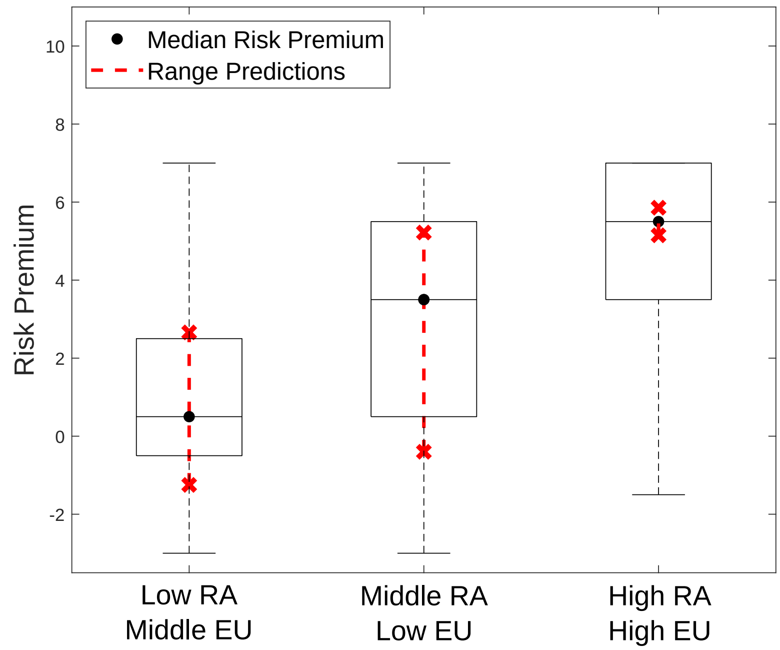


Figure 3.6: Observed risk premia and predicted ranges under the EU-core model.

Notes: Box plots depict the distributions of risk premia in the three groups identified using mixture models with the EU-core. Groups are classified based on risk aversion (RA) and adherence to EU. Observations are considered outliers if they lie more than 1.5 interquartile ranges above the third quartile (75 percent) or below the first quartile (25 percent). We identified 11 outliers in the "High RA High EU" group. All outliers are omitted from the graph for clarity.

To benchmark the performance of our approach, we employ both economic models and machine learning algorithms to predict certainty equivalents. EU and CPT directly yield point predictions for the certainty equivalent of every lottery. Similarly, with GBT and NN, we can predict the specific point in a price list where subjects transition from favoring the certain amount to the risky lottery. We then deduce certainty equivalents using the same approach applied to derive certainty equivalents from the observed choices in Block 2.[26] To establish a fair comparison between the EU-core approach and the other methods, we compute point predictions of the EU-core approach for the certainty equivalents by taking the average values of the predicted ranges.

We use choices from CR-tasks and R-tasks as our training data and compare the mean squared errors of various approaches in predicting certainty equivalents. The results

[26]In 5.25% of observations, GBT predicts either multiple switches between the certain amount and the risky lottery, or a single switch that's directionally incorrect. In this latter scenario, subjects are forecasted to choose the certain prize when its value is low and opt for the risky lottery when the certain prize value is high. We exclude these observations when evaluating the performance of GBT.

Table 3.5: Out-of-sample exercise across tasks in Block 1 and Block 2.

Exercise	Loss	Models				
		EU	CPT	GBT	NN	EU-core
Train: CR-tasks and R-tasks Test: Certainty Equivalents in Block 2	MSE	13.184	16.964	64.148	57.398	**11.444**

Notes: We employ mixture models with three groups to predict risk premia using EU, CPT, and the EU-core approach. For predictions with the EU-core approach, we use the average risk premium within the predicted range.

are presented in Table 3.5, where a lower mean squared error indicates superior model performance. In this out-of-sample evaluation, machine learning algorithms perform considerably worse than other methods. Specifically, GBT predicts that for over 90% of the lotteries, subjects either always prefer the certainty prize, or the risky lottery. Meanwhile, NN yields the analogous prediction for all lotteries. Therefore, both GBT and NN fail to offer reasonable predictions for choices in Block 2. EU and CPT significantly outperform machine learning algorithms in this task, with EU achieving a lower MSE than CPT. Finally, the EU-core approach emerges as the top performer.

Result 6 *In the exercise of predicting risk premia from Block 2 using choices in Block 1 as training data, the EU-core approach outperforms both economic models and machine learning algorithms. Within economic models, EU performs better than CPT, while machine learning algorithms exhibit the poorest performance in this out-of-sample exercise.*

Investment and Insurance Behaviors Outside the Experiment

In this section, we explore whether the heterogeneity identified in the experiments through the EU-core approach, in terms of risk aversion and non-EU behavior, has any correlation with real investment and insurance behaviors. All subjects on Prolific are asked a series of questions about their financial habits when they first enroll on the platform. In the preregistration of the experiment, we chose to evaluate two conjectures.[27] The first conjecture concerns risk aversion, and the second pertains to subjects' adherence to EU.

We posit that individuals who are less risk-averse should be more inclined to invest, especially in volatile assets. To assess this conjecture, we focused on two specific questions. The first inquires whether subjects have made investments (either personal

[27] See page 14 of the analysis plan preregistered at the AEA RCT Registry as AEARCTR-0011749 (Kobayashi and Lucia, 2023).

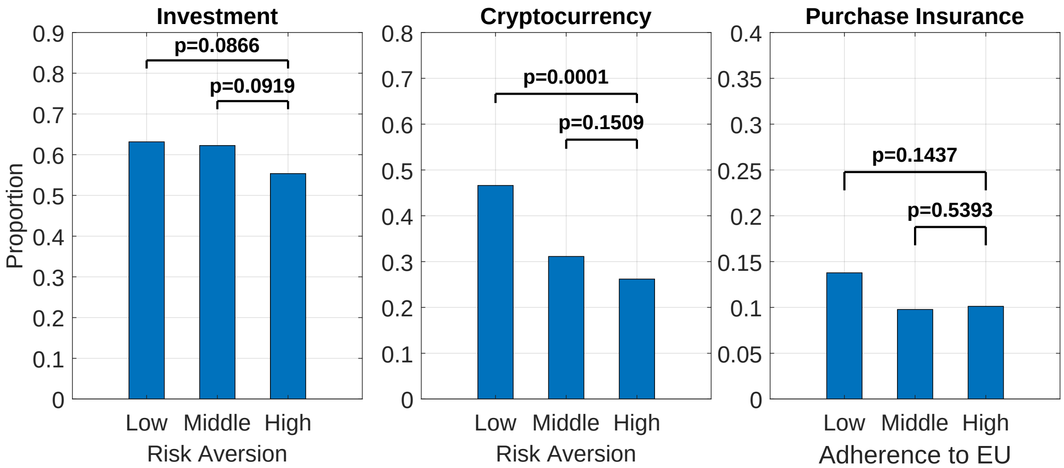


Figure 3.7: Investment and insurance behaviors: proportions.

Notes: We estimated a mixture model for the EU-core with three groups, using CR-tasks and R-tasks. Each subject was assigned to the group with the highest group membership probability. The left graph displays proportions based on affirmative responses to the question, "Have you ever made investments (either personally or through your employment) in the common stock or shares of a company?" The middle graph represents proportions based on affirmative answers to the question, "Do you own/hold any cryptocurrencies?" The right graph, meanwhile, focuses on the question, "Do you actively hold any of the following types of insurance policies?" Specifically, it illustrates the proportion of subjects who selected the option "Purchase Insurance (e.g., Mobile Phone)."

or through their employment) in the common stock or shares of a company. Of course, the act of investing in company shares per se does not necessarily correlate with risk aversion. Much depends on the level of risk associated with the specific stocks considered, information we do not possess. Consequently, we also decided to explore whether subjects declared ownership of cryptocurrencies, serving as a more unambiguous proxy for risky behavior. Moreover, to capture non-EU behavior, we examined whether subjects have purchased insurance for items, such as smartphones. This behavior is a classic example of small stakes risk aversion, which can be challenging to rationalize with EU. Therefore, our second conjecture is that subjects who align more closely with EU should be less inclined to purchase this type of insurance.

Figure 3.7 summarizes the responses to the three questions under consideration, across the different groups of subjects identified with the mixture model for the EU-core estimated using all data from Block 1. For investments in shares of companies (left graph) and cryptocurrencies (middle graph), we order groups from left to right based on their estimated level of risk aversion. For purchasing insurance (right graph), we order groups based on their estimated adherence to EU.

The left and middle graphs of Figure 3.7 provide evidence in line with our first conjecture: groups of subjects characterized by lower estimated levels of risk aversion have higher proportions of individuals investing in shares of companies or cryptocurrencies. Moreover, the negative correlation between risk aversion and investment behavior is more pronounced for cryptocurrencies than for company shares. This observation supports the idea that investments in cryptocurrencies might be a better proxy for risky behavior. Consistent with our second conjecture, the proportion of subjects purchasing insurance is highest among groups with low predicted adherence to EU.

Result 7 *We classified subjects in terms of risk aversion and adherence to EU using the mixture model for the EU-core estimated from choices in CR-tasks and R-tasks. Subjects classified as more risk averse are less likely to invest, particularly in cryptocurrencies. Moreover, subjects classified as less adherent to EU purchase insurance more frequently.*

3.6 Discussion

Our paper offers two main contributions. Empirically, we demonstrate the shortcomings of popular economic models and machine learning algorithms in both rationalizing and predicting two widely documented and influential non-EU behaviors, which were previously mostly analyzed separately in experimental work. Methodologically, we introduce a novel empirical strategy for predicting behavior under risk and showcase its effectiveness through a series of out-of-sample exercises. We conclude by discussing the implications of our results for future research.

A satisfactory model of decision-making under risk should rationalize the strong positive correlation observed between the CRE and preferences for randomization. Conversely, CPT predicts the opposite correlation between these two non-EU behaviors, explaining why this model consistently achieves inferior out-of-sample performance compared to EU in our analysis. A natural direction for future research involves considering alternative models to CPT. Fudenberg et al. (2022) demonstrate that adding to CPT a complexity cost, which increases with the number of prizes in a lottery, yields better out-of-sample predictions. In our experiment, adding a complexity cost to CPT would further strengthen the negative relationship between CRE and preferences for randomization, leading to worse out-of-sample predictions.

Of course, many alternative theories to CPT and its generalizations have been proposed, and some of them may be capable of rationalizing our main experimental findings. One important example is the original Prospect Theory (PT; Kahneman and Tversky, 1979). This model drops the rank dependence assumption that characterizes CPT and can rationalize the observed correlation between CRE and preferences for randomization through simple probability weighting. However, there is also abundant experimental evidence demonstrating failures of PT, with CPT being proposed as a solution to the violations of first-order stochastic dominance implied by PT. The overall absence of an economic model that systematically outperforms the others in terms of predictive accuracy generates interest in exploring alternative approaches for making predictions.

Machine learning algorithms offer an alternative approach for predicting behavior under risk, and a growing body of research compares their predictive capabilities with those of economic models.[28] In our analysis, machine learning algorithms outperformed economic models only when the training and test sets included choices over the exact same lotteries. However, their predictive capabilities dropped significantly in all other out-of-sample exercises. Andrews et al. (2022) obtain a similar result when comparing the out-of-sample performance of economic models and machine learning algorithms in the prediction of certainty equivalents. In particular, they observe that the performance of machine learning algorithms are sensitive to which lotteries are included in the training and test sets. Therefore, the sensitivity of these methods to minor differences between training and test sets raises concerns about their ability to produce generalizable predictions that are at least as substantial as those for economic models.

This paper introduces a novel empirical approach to make predictions that retains an underlying economic structure without being tied to specific models. We show that the predictions of our approach are more accurate but at same time less detailed than those produced by fully specified economic models or machine learning algorithms. For instance, our approach does not allow distinguishing between specific non-EU behaviors, or it does not allow one to directly obtain a point prediction for a certainty equivalent. While this paper focuses on choices under risk, we believe our empirical strategy holds promise for extension to choices under uncertainty, exploiting the notion of "unambiguous preference" introduced by Ghirardato et al. (2004).

Ultimately, determining the "best" method to predict behavior under risk may depend

[28]See Andrews et al. (2022), Camerer et al. (2019), Fudenberg and Liang (2019), Noti et al. (2016), Peterson et al. (2021), Plonsky et al. (2017), Plonsky et al. (2019), and Zhao et al. (2020).

on at least two factors. If the analyst has access to a training set that is sufficiently close to the test set of interest, machine learning algorithms might be the most suitable alternative. However, what constitutes "sufficiently close" can vary based on the specific nature of the problem at hand. Our analysis underscores how minor discrepancies between training and test sets can lead to substantial declines in the performance of machine learning algorithms. Another critical aspect to consider is the level of detail required in the predictions. For instance, if the analyst aims to estimate measures of risk aversion or adherence to EU, our approach offers a promising alternative to traditional predictive tools. Conversely, if our method does not provide the necessary level of detail in predictions, it can still complement other techniques. For example, if a point prediction for a certainty equivalent is needed, our strategy can first offer a range prediction, which economic models or machine learning algorithms can then refine to pinpoint a value within that range.

References

Agranov, Marina, Paul J Healy, and Kirby Nielsen (June 2023). "Stable Randomisation". In: *The Economic Journal* 133.655, pp. 2553–2579. ISSN: 0013-0133. DOI: 10.1093/ej/uead039. eprint: https://academic.oup.com/ej/article-pdf/133/655/2553/51707894/uead039.pdf. URL: https://doi.org/10.1093/ej/uead039.

Agranov, Marina and Pietro Ortoleva (2017). "Stochastic Choice and Preferences for Randomization". In: *Journal of Political Economy* 125.1, pp. 40–68. ISSN: 00223808, 1537534X. URL: https://www.jstor.org/stable/26549925 (visited on 10/16/2023).

– (2022). "Revealed Preferences for Randomization: An Overview". In: *AEA Papers and Proceedings* 112, pp. 426–30. DOI: 10.1257/pandp.20221093. URL: https://www.aeaweb.org/articles?id=10.1257/pandp.20221093.

– (July 2023). "Ranges of Randomization". In: *The Review of Economics and Statistics*, pp. 1–44. ISSN: 0034-6535. DOI: 10.1162/rest_a_01355. eprint: https://direct.mit.edu/rest/article-pdf/doi/10.1162/rest_a_01355/2150363/rest_a_01355.pdf. URL: https://doi.org/10.1162/rest%5C_a%5C_01355.

Andrews, Isaiah et al. (2022). "The Transfer Performance of Economic Models". In: *arXiv preprint arXiv:2202.04796*.

Athey, Susan (2017). "Beyond prediction: Using big data for policy problems". In: *Science* 355.6324, pp. 483–485. DOI: 10.1126/science.aal4321. eprint: https://www.science.org/doi/pdf/10.1126/science.aal4321. URL: https://www.science.org/doi/abs/10.1126/science.aal4321.

Bruhin, Adrian, Helga Fehr-Duda, and Thomas Epper (2010). "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion". In: *Econometrica* 78.4, pp. 1375–1412. ISSN: 1468-0262. DOI: 10.3982/ECTA7139. (Visited on 04/08/2022).

Camerer, Colin F and Teck-Hua Ho (1994). "Violations of the betweenness axiom and nonlinearity in probability". In: *Journal of risk and uncertainty* 8.2, pp. 167–196.

Camerer, Colin F, Gideon Nave, and Alec Smith (2019). "Dynamic unstructured bargaining with private information: theory, experiment, and outcome prediction via machine learning". In: *Management Science* 65.4, pp. 1867–1890.

Cerreia-Vioglio, Simone (2009). *Maxmin expected utility on a subjective state space: Convex preferences under risk*. Tech. rep. Mimeo, Bocconi University.

Cerreia-Vioglio, Simone, David Dillenberger, and Pietro Ortoleva (2015). "Cautious expected utility and the certainty effect". In: *Econometrica* 83.2, pp. 693–728.

Cerreia-Vioglio, Simone, Fabio Maccheroni, and Massimo Marinacci (2017). "Stochastic dominance analysis without the independence axiom". In: *Management Science* 63.4, pp. 1097–1109.

Cettolin, Elena and Arno Riedl (2019). "Revealed preferences under uncertainty: Incomplete preferences and preferences for randomization". In: *Journal of Economic Theory* 181, pp. 547–585.

Chapman, Jonathan et al. (2023a). "Econographics". In: *Journal of Political Economy Microeconomics* 1.1, pp. 115–161. DOI: 10.1086/723044. eprint: https://doi.org/10.1086/723044. URL: https://doi.org/10.1086/723044.

– (2023b). *Willingness to accept, willingness to pay, and loss aversion*. Tech. rep. National Bureau of Economic Research.

Chew, S. H., L. G. Epstein, and U. Segal (1991). "Mixture Symmetry and Quadratic Utility". In: *Econometrica* 59.1, pp. 139–163. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2938244 (visited on 02/20/2024).

Coveney, Peter V., Edward R. Dougherty, and Roger R. Highfield (2016). "Big data need big theory too". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2080, p. 20160153. DOI: 10.1098/rsta.2016.0153. eprint: https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2016.0153. URL: https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2016.0153.

Dean, Mark and Pietro Ortoleva (2015). "Is it All Connected? A Testing Ground for Unified Theories of Behavioral Economics Phenomena". In: *ERN: Uncertainty & Risk Modeling (Topic)*. URL: https://api.semanticscholar.org/CorpusID:43571737.

Dean, Mark and Pietro Ortoleva (2019). "The empirical relationship between nonstandard economic behaviors". In: *Proceedings of the National Academy of Sciences* 116.33, pp. 16262–16267. DOI: 10.1073/pnas.1821353116. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1821353116. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1821353116.

Dwenger, Nadja, Dorothea Kübler, and Georg Weizsäcker (2018). "Flipping a coin: Evidence from university applications". In: *Journal of Public Economics* 167, pp. 240–250. ISSN: 0047-2727. DOI: https://doi.org/10.1016/j.jpubeco.2018.09.014. URL: https://www.sciencedirect.com/science/article/pii/S0047272718301889.

Feldman, Paul and John Rehbeck (2022). "Revealing a preference for mixtures: An experimental study of risk". In: *Quantitative Economics* 13.2, pp. 761–786.

Freeman, David J, Yoram Halevy, and Terri Kneeland (2019). "Eliciting risk preferences using choice lists". In: *Quantitative Economics* 10.1, pp. 217–237.

Fudenberg, Drew and Annie Liang (2019). "Predicting and understanding initial play". In: *American Economic Review* 109.12, pp. 4112–4141.

Fudenberg, Drew et al. (2022). "Measuring the completeness of economic models". In: *Journal of Political Economy* 130.4, pp. 956–990.

Ghirardato, Paolo, Fabio Maccheroni, and Massimo Marinacci (2004). "Differentiating ambiguity and ambiguity attitude". In: *Journal of Economic Theory* 118.2, pp. 133–173. ISSN: 0022-0531. DOI: https://doi.org/10.1016/j.jet.2003.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0022053104000262.

Gul, Faruk (1991). "A theory of disappointment aversion". In: *Econometrica: Journal of the Econometric Society*, pp. 667–686.

Hey, John D and Chris Orme (1994). "Investigating generalizations of expected utility theory using experimental data". In: *Econometrica: Journal of the Econometric Society*, pp. 1291–1326.

Hofman, Jake M et al. (2021). "Integrating explanation and prediction in computational social science". In: *Nature* 595.7866, pp. 181–188.

Kahneman, Daniel and Amos Tversky (1979). "Prospect Theory: An Analysis of Decision under Risk". In: *Econometrica* 47.2, pp. 263–291. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1914185 (visited on 10/15/2023).

Loomes, Graham and Robert Sugden (1982). "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty". In: *The Economic Journal* 92.368, pp. 805–824. ISSN: 00130133, 14680297. URL: http://www.jstor.org/stable/2232669 (visited on 02/20/2024).

Machina, Mark J. (1982). "“Expected Utility” analysis without the independence axiom". In: *Econometrica* 50.2, pp. 277–323. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1912631 (visited on 02/21/2024).

Marschak, Jacob (1950). "Rational behavior, uncertain prospects, and measurable utility". In: *Econometrica: Journal of the Econometric Society*, pp. 111–141.

Miao, Bin and Songfa Zhong (2018). "Probabilistic social preference: how Machina's Mom randomizes her choice". In: *Economic Theory* 65, pp. 1–24.

Noti, Gali et al. (2016). "Behavior-Based Machine-Learning: A Hybrid Approach for Predicting Human Decision Making". In: *ArXiv* abs/1611.10228. URL: https://api.semanticscholar.org/CorpusID:14606035.

Peterson, Joshua C. et al. (2021). "Using large-scale experiments and machine learning to discover theories of human decision-making". In: *Science* 372.6547, pp. 1209–1214. DOI: 10.1126/science.abe2629. eprint: https://www.science.org/doi/pdf/10.1126/science.abe2629. URL: https://www.science.org/doi/abs/10.1126/science.abe2629.

Plonsky, Ori et al. (2017). "Psychological forest: Predicting human behavior". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. 1.

Plonsky, Ori et al. (2019). "Predicting human decisions with behavioral theories and machine learning". In: *arXiv preprint arXiv:1904.06866*.

Rabin, Matthew (2000). "Risk Aversion and Expected-Utility Theory: A Calibration Theorem". In: *Econometrica* 68.5, pp. 1281–1292. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/2999450 (visited on 10/29/2023).

Rothschild, Michael and Joseph E Stiglitz (1970). "Increasing risk: I. A definition". In: *Journal of Economic Theory* 2.3, pp. 225–243. ISSN: 0022-0531. DOI: https://doi.org/10.1016/0022-0531(70)90038-4. URL: https://www.sciencedirect.com/science/article/pii/0022053170900384.

Sandroni, Alvaro, Sandra Ludwig, and Philipp Kircher (2013). "On the difference between social and private goods". In: *The BE Journal of Theoretical Economics* 13.1, pp. 151–177.

Segal, Uzi (1990). "Two-stage lotteries without the reduction axiom". In: *Econometrica: Journal of the Econometric Society*, pp. 349–377.

Sopher and Narramore (2000). "Stochastic Choice and Consistency in Decision Making Under Risk: An Experimental Study". In: *Theory and Decision* 48.4, pp. 323–349. DOI: 10.1023/a:1005289611789.

Tversky, Amos and Daniel Kahneman (1992). "Advances in prospect theory: Cumulative representation of uncertainty". In: *Journal of Risk and uncertainty* 5.4, pp. 297–323.

Zhao, Chen et al. (2020). "Behavioral neural networks". In: *Available at SSRN 3633548*.

APPENDIX TO CHAPTER 1

A.1 Reduced-Form Evidence

Table A.1: Regression discontinuity for price. We use the subsample from the two-hour interval around the budget renewal point.

	log(win_bid)
after renewal	0.4257^{***} (0.0786)
(time − renewal)	-17.60^{*} (10.25)
$(\text{time} - \text{renewal})^2$	-5.057 (4.821)
log(#auctions per 5-min interval)	-0.0927 (0.0802)
computer	-0.2433^{***} (0.0261)
optout	-0.0746^{***} (0.0279)
match_cookie_prop	1.305^{***} (0.0169)
gender = Male	-0.2327 (0.1421)
gender = Female	-0.1599 (0.1441)
age = 25to44	0.1336 (0.1416)
age = 45plus	0.1243 (0.1423)
seg_size	$1.94 \times 10^{-5***}$ (2.04×10^{-6})
num_month_sold	-0.0079^{***} (0.0025)
total_rev	0.0001 (0.0002)
avg_rev	-0.0089^{***} (0.0017)
profile_length	0.0003^{***} (5.47×10^{-5})
after renewal × (time − renewal)	-1.361 (14.32)
after renewal $\times (\text{time} - \text{renewal})^2$	8.965 (7.241)
Site FE	Yes
Browser FE	Yes
City FE	Yes
Observations	30,319
Adjusted R^2	0.32271

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Table A.2: OLS Bid Regression and Logit Entry Regression.

Model:	log(bid) OLS	entry Logit
log(#auctions per 5-min interval)	-0.0701*** (0.0114)	-0.1198*** (0.0246)
computer	-0.0888*** (0.0272)	-0.3689*** (0.1380)
optout	0.0626** (0.0242)	-0.3666 (0.2338)
match_cookie_prop	0.4880*** (0.0571)	2.087*** (0.4137)
gender = Male	-0.0551*** (0.0122)	-0.0130 (0.0323)
gender = Female	-0.0375*** (0.0112)	0.0090 (0.0332)
age = 25to44	0.0314*** (0.0099)	0.0262 (0.0211)
age = 45plus	0.0120 (0.0091)	0.0209 (0.0198)
seg_size	92.10*** (14.58)	642.1*** (109.9)
num_month_sold	-34.46** (16.02)	-272.8* (154.3)
total_rev	-26.59*** (4.851)	-60.62** (25.81)
avg_rev	-35.40*** (4.336)	-317.7*** (47.69)
profile_length	74.67*** (17.68)	586.1*** (210.4)
Site FE	Yes	Yes
Browser FE	Yes	Yes
City FE	Yes	Yes
Day-Hour FE	Yes	Yes
DSP FE	Yes	Yes
Advertiser FE	Yes	No
Observations	8,856,603	45,484,100
Adjusted/Pseudo R^2	0.44974	0.34448

Double-clustered (DSP & 5-min interval) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

A.2 First-step estimates

	$\hat{r}_t$	$\hat{p}_t$
$t = 1$	617.9 (440.6)	0.159 (0.095)
$t = 2$	2225.1 (2350.6)	0.498 (0.264)
$t = 3$	11055.2 (18673.2)	0.875 (0.185)
$t = 4$	1038.7 (910.5)	0.472 (0.218)
$t = 5$	238551.6 (288525.0)	0.995 (0.006)
$t = 6$	3933.6 (4070.4)	0.701 (0.217)
$t = 7$	822.3 (563.5)	0.202 (0.110)
$t = 8$	1079.0 (681.5)	0.151 (0.081)
$t = 9$	458.7 (298.2)	0.044 (0.027)
$t = 10$	421.5 (274.4)	0.028 (0.018)
$t = 11$	322.0 (197.4)	0.020 (0.012)
$t = 12$	1475.1 (1662.2)	0.080 (0.083)
$t = 13$	360.4 (227.6)	0.022 (0.013)
$t = 14$	232.0 (142.1)	0.015 (0.009)
$t = 15$	195.5 (125.0)	0.013 (0.008)
$t = 16$	199.2 (127.3)	0.013 (0.008)
$t = 17$	125.0 (78.1)	0.009 (0.006)
$t = 18$	154.5 (99.0)	0.014 (0.009)
$t = 19$	203.6 (131.6)	0.022 (0.014)
$t = 20$	108.1 (67.7)	0.013 (0.008)
$t = 21$	83.3 (52.9)	0.011 (0.007)
$t = 22$	109.6 (70.2)	0.016 (0.010)
$t = 23$	109.6 (70.5)	0.019 (0.012)
$t = 24$	232.6 (155.2)	0.053 (0.034)

Table A.3: Estimated parameters for $F_K^{(t)} = \text{NegativeBinomial}(r_t, p_t)$, the distribution of the number of auctions for each period t.

A.3 Bidders' objectives under the second-price auction

Let $\widetilde{\Psi}_t(\cdot \mid K)$ be the belief on rivals' highest bid P conditional on the current number of auction K under the SPA format. This belief represents the probability of winning an auction and also the distribution of the price in each auction. Given the ex-ante value function $\widetilde{EV}_{t+1}(w) = E[\widetilde{V}_{t+1}(K_{t+1}, w)]$ with $\widetilde{EV}_{T+1}(w) = \eta Q(w)$, the Bellman formulation of the objective is

$$\max_{\gamma,\tau} K_t F_C(\tau) \left[E\left[\int_0^{b^\gamma(X)} (X - p) d\widetilde{\Psi}_t(p \mid K_t) \right] - E[C \mid C \leq \tau] \right] + E\left[\widetilde{EV}_{t+1}(w_{it+1}) \mid \gamma, \tau \right]. \tag{A.1}$$

Now, we provide the first-order necessary conditions for the bidding problem under the second-price format while assuming differentiablity. The one with respect to the bid function parameter γ is given by

$$E\Big[\underbrace{\Psi_t'(b^\gamma(X)\mid K_t)(X-b^\gamma(X))}_{\text{Static FOC}}\nabla_\gamma b^\gamma(X)\Big]+\underbrace{\frac{1}{K_tF_C(\tau)}\nabla_\gamma E\left[\widetilde{EV}_{t+1}(w_{it+1})\mid\gamma,\tau\right]}_{\text{Dynamic Tradeoff}}=0.$$

Meanwhile, the first-order condition with respect to the entry threshold τ is

$$\tau=\underbrace{E\left[\int_0^{b^\gamma(X)}(X-p)d\widetilde{\Psi}_t(p\mid K_t)\right]}_{\text{Static Threshold}}+\underbrace{\frac{1}{Kf_C(t)}\frac{\partial}{\partial\tau}E\left[\widetilde{EV}_{t+1}(w_{it+1})\mid\gamma,\tau\right]}_{\text{Dynamic Tradeoff}}.$$

Again, similarly to the first-order conditions under the first-price format for (1.1), we have both static and dynamic components in the optimality conditions here. For instance, in the condition for the optimal bid strategy, the static component encourages truthful bidding, but the dynamic component provides a counteracting force.

A.4 Algorithm for solving for an equilibrium

We setup the algorithm by first making a grid over the state space for (K_t, w_{it}). For the remaining budget, we make a grid over $[-M, M]$ where M is the upper bound of the initial budgets. For the number of auctions, we take Monte Carlo draws from $\widehat{F}_K^{(t)}$ for each period t. We set $\{(\gamma_t^{(0)}(K,w),\tau_t^{(0)}(K,w)\}_{t=1}^T$ as the initial strategies.

Then, we execute the following loop: For each $m = 1,\ldots,$

- Forward simulate numerous paths of $\{(K_t, w_{1t},\ldots,w_{Nt})\}_{t=1}^T$ using $\{(\gamma_t^{(m-1)}(K,w),\tau_t^{(m-1)}(K,w)\}_{t=1}^T$ to numerically obtain the belief:

  $$\begin{aligned}&\Psi_t^{(m)}(b\mid K)\\&=E\left[\prod_{j\neq i}\Big(1-F_C(\tau_t^{(m-1)}(K,w_{jt}))+F_C(\tau_t^{(m-1)}(K,w_{jt}))F_X(b^{-1}(b\mid\gamma_t^{(m-1)}(K,w_{jt})))\Big)\right]\end{aligned}\tag{A.2}$$

  over the grid of K_t for each t.

- Obtain $\{(\gamma_t^{(m)}(K,w),\tau_t^{(m)}(K,w)\}_{t=1}^T$ over the grid of states that best respond given $\Psi_t^{(m)}(b\mid K)$ by solving (1.1) via backward induction.

- Break if $\|\{(\gamma_t^{(m)}(K,w),\tau_t^{(m)}(K,w)\}_{t=1}^T-\{(\gamma_t^{(m-1)}(K,w),\tau_t^{(m-1)}(K,w)\}_{t=1}^T\|$ is below some tolerance.

Appendix B

APPENDIX TO CHAPTER 2

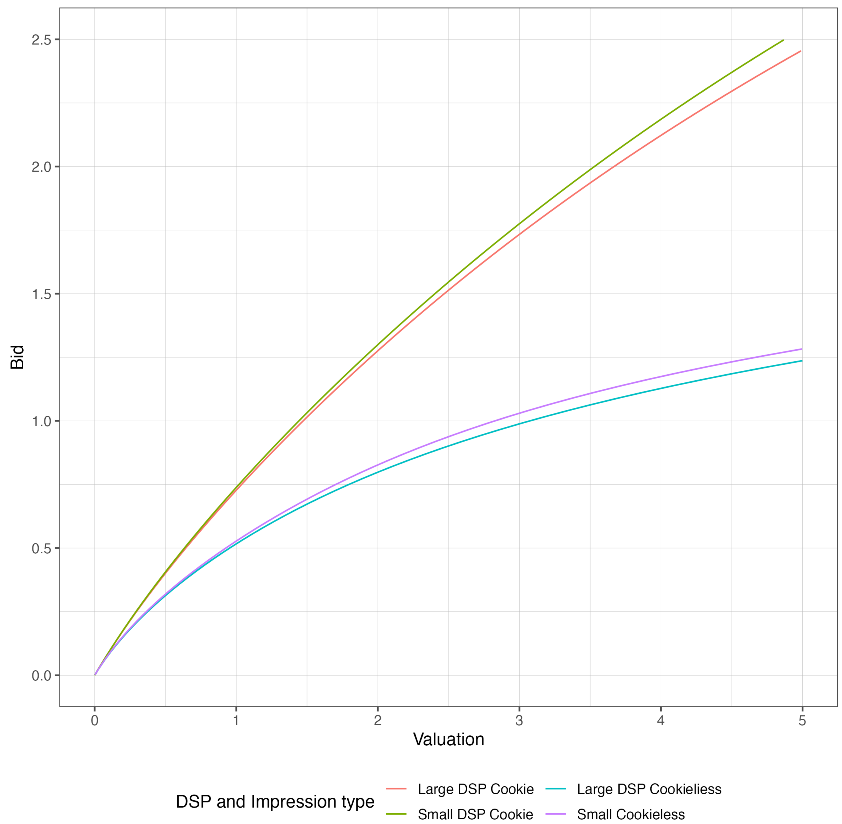
B.1 Additional Tables and Figures

Table B.1: Regression results of logit model of entry decision.

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

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

Figure B.1: Bidding functions of large and small general-purpose DSPs.



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

Appendix C

APPENDIX TO CHAPTER 3

C.1 Descriptive Analysis

We provide a descriptive analysis of behavior in Block 1 and Block 2 of the experiment.

Block 1

Figure C.1 displays the percentages of various choice patterns observed in CR-tasks (on the left graph) and R-tasks (on the right graph).

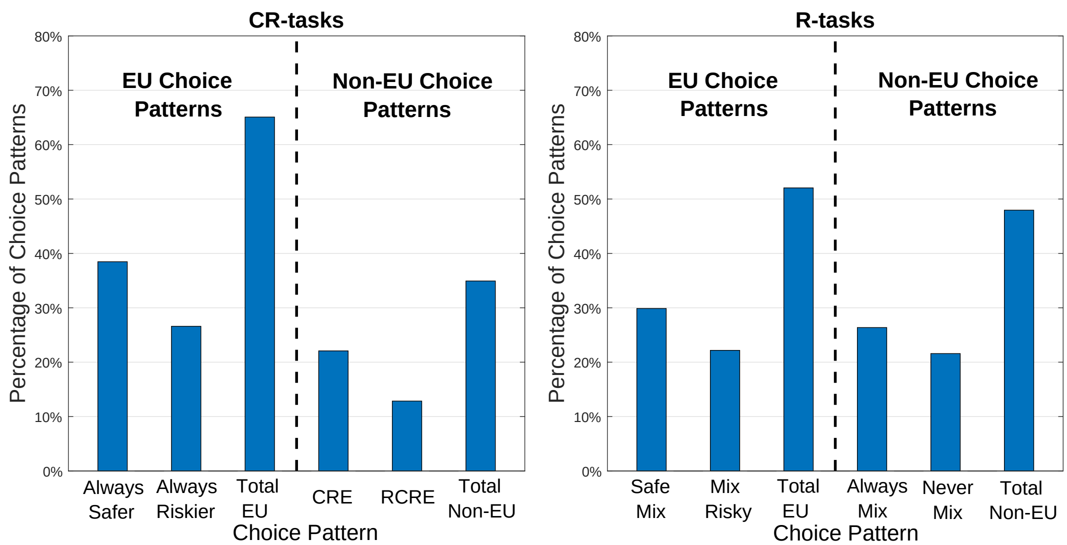


Figure C.1: Percentages of the different choice patterns in CR-tasks and R-tasks.

Notes: There are four possible choice patterns for each pair of lotteries (δ_M, r) in CR-tasks (left graph) and R-tasks (right graph). In CR-tasks, "Always Safer" means consistently selecting the safer lottery, while "Always Riskier" denotes the opposite choice. The CRE and RCRE are the two possible non-EU choice patterns in CR-tasks. Within R-tasks, "Safe Mix" indicates choosing the safe lottery over the mixture, and the mixture over the risky lottery. Vice versa, "Mix Risky" indicates the opposite behavior. For non-EU choice patterns, "Always Mix" indicates always choosing the mixture, while "Never Mix" indicates the opposite behavior. Finally, "Total EU" and "Total Non-EU" indicates respectively the aggregate percentages of EU and non-EU choice patterns within CR-tasks (left graph) and R-tasks (right graph).

Within the CR-tasks, two choice patterns are consistent with EU: always choosing the safer lottery ("Always Safer") and always choosing the riskier lottery ("Always

Riskier"), which together make up 65.07% of all choice patterns. Among the non-EU choice patterns in CR-tasks, the CRE is the most frequent, occurring in roughly 9% more cases than the RCRE. In the context of R-tasks, the EU-consistent patterns emerge when subjects either choose the safe lottery over the mixture and then the mixture over the risky one ("Safe Mix"), or vice versa ("Mix Risky"). EU choice patterns in R-tasks account for 52.05% of all choice patterns—a drop of roughly 13% compared to CR-tasks. Turning to non-EU patterns in R-tasks, always choosing the mixture ("Always Mix") constitutes the 26.37% of choice patterns, while never choosing the mixture ("Never Mix") constitutes the 21.58%.

Block 2

Figure C.2 uses box plots to summarize the distribution of risk premia for the three lotteries presented to subjects in Block 2. The distributions of risk premia

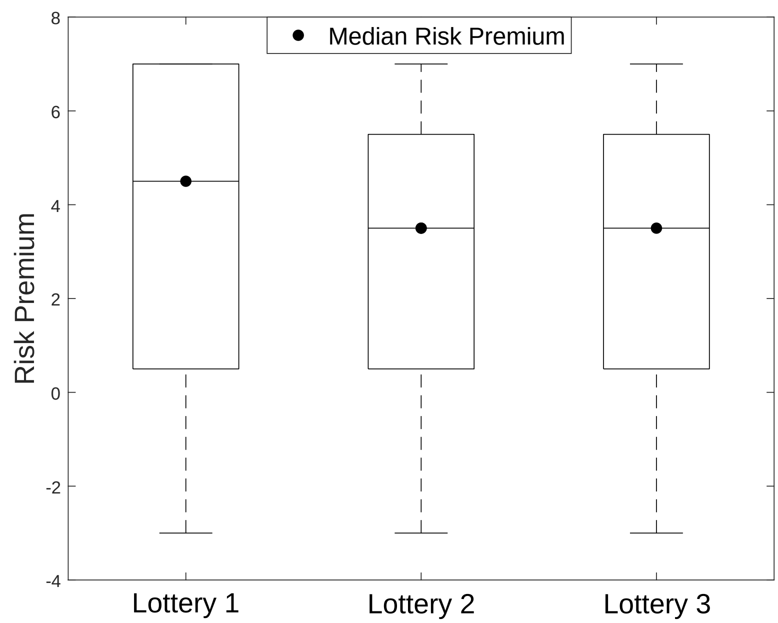


Figure C.2: Distributions of certainty equivalents for the three lotteries considered in Block 2.

Notes: We focus on the subset of observations from Block 2 where subjects shifted their preference between a fixed lottery and a certain amount at most once. The box plots summarize the distributions of risk premia for the three lotteries from Block 2. Lottery 1 pays \$0 or \$20 with equal chance. Lottery 2 pays \$5 or \$25 with equal chance. Lottery 3 pays \$10 or \$30 with equal chance.

are comparable across the three lotteries. For all three lotteries, the median risk premium is around \$4, and more than 75% of subjects have a positive risk premium. Moreover, for each lottery, we observe a significant level of heterogeneity in the risk premia.

C.2 Econometric Procedures

We provide details about the procedures that we follow to estimate economic models and train machine learning algorithms.

Economic Models

Cumulative Prospect Theory (CPT).

The value of lottery $p = (\$L, p_L; \$M, p_M; \$H, p_H)$ under CPT is

$$U_{CPT}(p) = \pi(p_H)u(H) + [\pi(p_H + p_M) - \pi(p_M)]\, u(M) + [1 - \pi(p_H + p_M)]\, u(L),$$

where $v(\cdot)$ is a utility function and $\pi(\cdot)$ is a probability weighting function. For estimation purposes, we consider the functional forms for utility and probability weighting functions proposed by Tversky and Kahneman, 1992:

$$u(x) = x^{\alpha};$$
$$\pi(p) = \frac{p^{\gamma}}{[p^{\gamma} + (1-p)^{\gamma}]^{\frac{1}{\gamma}}}.$$

Within the empirical framework proposed by Hey and Orme, 1994, a CPT decision-maker chooses lottery p over lottery q if

$$U_{CPT}(p) - U_{CPT}(q) \geq \epsilon,$$

where ϵ is an error term normally distributed with a mean of zero and a variance of $\sigma > 0$.

We define $I = \{1, \ldots, N\}$ as a set of subjects in our experiment, $\triangle(X)$ as the set of lotteries over X, and by $\mathcal{D} \subseteq \triangle(X)^2$ as a subset of pairs of lotteries where the subjects express their preferences. We construct an index, $Choice_i$, for each subject i as follows: for each pair of lotteries $(p, q) \in \mathcal{D}$,

$$Choice_i(p,q) \coloneqq \begin{cases} 2 & \text{if subject } i \text{chooses lottery } p \text{ over lottery } q \\ 1 & \text{otherwise.} \end{cases}$$

We estimate mixture models with three groups. Within each group c, we estimate the risk-aversion coefficient α_c, the probability weighting function coefficient γ_c, and the variance of the error term σ_c. We denote by $f(Choice_i; \alpha_c, \gamma_c, \sigma_c)$ the likelihood function for subject i belonging to group c:

$$\prod_{(p,q)\in\mathcal{D}} \Big(\mathbb{1}\,(Choice_i(p,q) = 2) \cdot \Pr(U_{CPT}(p) - U_{CPT}(q) \geq \epsilon \mid \alpha_c, \gamma_c, \sigma_c)$$
$$+ \mathbb{1}\,(Choice_i(p,q) = 1) \cdot \Pr(U_{CPT}(p) - U_{CPT}(q) < \epsilon \mid \alpha_c, \gamma_c, \sigma_c) \Big).$$

Let π_c represent the probability of a subject belonging to group type c. The log-likelihood of the finite mixture model is given by:

$$\sum_{i=1}^{N} \ln \sum_{c=1}^{C} \pi_c f(Core_i; \alpha_c, \gamma_c, \sigma_c),$$

where the first sum is over subjects and the second sum is over groups.

We estimate the utility functions, the parameters of the covariance matrices, and the probabilities of group membership through maximum likelihood estimation. We employ the Global Search algorithm in Matlab to maximize the log-likelihood. To ensure that the algorithms converges to a global maximum, we employ a multi-start approach initiating multiple searches with 200 different starting points. We also evaluate the robustness of our estimates by estimating the model using an expectation-maximization algorithm (Dempster et al., 1977).

Expected Utility (EU). To estimate EU, we repeat the same procedure fixing the value of the probability weighting function to one for all the three groups.

Machine Learning Algorithms

Gradient Boosting Trees (GBT). We employ the LogitBoost algorithm using MATLAB's "fitcensemble" function, a specialized gradient boosting methodology tailored for binary classification. This method uses an ensemble of weak decision tree learners, optimizing the logistic loss to enhance classification accuracy. We allow the algorithm to use the following set of features: probabilities of the lotteries (p, q) and an indicator for each subject. The output of the algorithm is the probability of choosing p over q.

The LogitBoost operates in a stage-wise fashion. For each iteration, the algorithm focuses on the residuals or errors made by the present ensemble, these errors being a product of the logistic loss. Instead of directly approximating the class labels, LogitBoost models the posterior probabilities of the classes. At every step, a new decision tree is trained to fit the current residuals. This tree, once trained, is amalgamated into the ensemble. To optimize performance, we've implemented hyperparameter optimization, adjusting the number of learning cycles, learning rate, and minimum leaf size for the decision trees. The optimization uses a 10-fold cross-validation, parallel computation, and is bounded to a maximum of 150 evaluations with a 4-hour time constraint.

Neural Networks (NN). We employ a neural network classifier using MATLAB's "fitcnet" function. We allow the NN to use the following sets of input variables: probabilities of the lotteries, prizes of the lotteries, observed choices and an indicator for each subject. The outputs of the algorithm are choice probabilities.

The NN follows a structured methodology. Initially, the input data is processed with standardization, ensuring all features have a mean of zero and a variance of one. This pre-processing step aids in stabilizing and speeding up the network's convergence during training. Next, we implement a 10-fold cross-validation strategy to optimize the following hyperparameters of the NN: activation functions, regularization strength and the size of the hidden layers. To accelerate the training, parallel computation is leveraged, and the optimization is constrained to a maximum of 150 evaluations with a 4-hour time limit.

C.3 Estimation Results

We report the estimation results arising from the EU-core analysis, as well as those arising from EU and CPT.

EU-core Analysis

Table C.1 presents the estimation results from the EU-core analysis, which were obtained using a mixture model with three groups. The estimates are presented with bootstrapped standard errors in parentheses. The row labeled with π presents the estimated sizes of each group. The row labeled with σ shows the standard deviation of the error terms for each utility. The row labeled with ρ indicates the correlation coefficients among the error terms.

EU and CPT

Table C.2 reports the estimation results for EU and CPT derived from mixture models with three groups. For EU, we rank groups of subjects based on risk aversion, which depends solely on the parameter that shapes the curvature of the utility functions (α). For CPT, we rank groups of subjects in terms of risk aversion and adherence to

Table C.1: EU-core mixture model: Estimation results.

Groups	Dataset: CR-tasks and R-tasks					
	High RA High EU		Middle RA Low EU		Low RA Middle EU	
π	0.2798 (0.1355)		0.4834 (0.1128)		0.2368 (0.0973)	
	Utility 1	**Utility 2**	**Utility 1**	**Utility 2**	**Utility 1**	**Utility 2**
\$5	0.7074 (0.1944)	0.6686 (0.1863)	0.5805 (0.1827)	0.4002 (0.1206)	0.1389 (0.0860)	0.3722 (0.1174)
\$10	0.9758 (0.1705)	0.9184 (0.1801)	0.6105 (0.1708)	0.8286 (0.1507)	0.4424 (0.0975)	0.3722 (0.1039)
\$15	0.9758 (0.1684)	0.9467 (0.1807)	0.7466 (0.1553)	0.8461 (0.1343)	0.5455 (0.1181)	0.5997 (0.1158)
\$20	1.0000 (0.1661)	1.0000 (0.1681)	0.7918 (0.1710)	0.9800 (0.1435)	0.7475 (0.1564)	0.6741 (0.1276)
\$25	1.0000 (0.1355)	1.0000 (0.1314)	1.0000 (0.1722)	0.9800 (0.1248)	0.8756 (0.1268)	0.9663 (0.1552)
σ	0.2759 (0.3551)	0.0294 (0.4093)	0.3052 (0.3016)	0.1632 (0.5821)	0.1106 (0.1978)	0.2502 (0.5068)
ρ	0.0034 (0.1364)		0.0000 (0.1110)		0.0090 (0.1261)	

EU. We use the predicted proportion of choices in CR-tasks and R-tasks where the safest available lottery was selected as a proxy for risk aversion.[1] Adherence to EU can be directly inferred from the probability weighting function parameter, γ.

C.4 Out-of-Sample Analysis: Additional Results

We develop a more comprehensive probabilistic and deterministic evaluation of all the predictive approaches considered in this paper. We first focus on the ability of economic models and machine learning algorithms to predict choice patterns, thus extending the analysis in Section 3.3. Next, we extend the EU-core analysis in Section 3.5.

Predict Choice Patterns. There are four possible choice patterns in CR-tasks and R-tasks. We can thus define a model f as a function that links each lottery pair (δ_M, r)

[1]Risk aversion under CPT is influenced by both utility and probability weighting functions. Hence, it is determined by the interaction between α and γ.

Table C.2: EU and CPT mixture models: Estimation results.

	Dataset: CR-tasks and R-tasks					
Model	**Expected Utility (EU)**			**Cumulative Prospect Theory (CPT)**		
Groups	High RA	Middle RA	Low RA	High RA Low EU	Middle RA Middle EU	Low RA High EU
α	0.0269 (0.2231)	0.4450 (0.3355)	0.9364 (0.2595)	1.0051 (0.1474)	0.2004 (0.2113)	0.7461 (0.1647)
γ				0.1216 (0.2678)	0.9800 (0.2092)	0.9463 (0.1975)
σ	0.1419 (0.1096)	0.5396 (0.5513)	2.0863 (1.1591)	6.0000 (0.7305)	0.2010 (0.0582)	1.3625 (0.3348)
π	0.4935 (0.2209)	0.3463 (0.2183)	0.1602 (0.1410)	0.2401 (0.0544)	0.4166 (0.1377)	0.3433 (0.1085)

Notes: Estimates are presented with bootstrapped standard errors in parentheses. In this table, α denotes the risk aversion coefficient, γ denotes the probability weighting function coefficient, σ denotes the standard deviation of the error term, and π denotes the estimated size of each group.

and subject i, with a vector of characteristics X_i, to a vector $f(\delta_M, r; X_i) \in \mathbb{R}^4$. Each component of this vector indicates the estimated probability under model f of one of the four choice patterns. Denoting by $P_i(\delta_M, r) \in \mathbb{R}^4$ the degenerate probability distribution that indicates which choice pattern, associated with the pair of lotteries (δ_M, r) and subject i, was actually observed, we define the loss of a model as follows:

$$L_p(f) := \sum_{(\delta_M, r)\in\mathcal{D}} \sum_{i=1}^{500} l\left(P_i(\delta_M, r), f(\delta_M, r; X_i)\right),$$

where $\mathcal{D}$ is the set of the 25 pair of lotteries (δ_M, r) in our experiment, and $l\colon \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}$ is a loss function, which we assume to be the Euclidean distance.

Moreover, we analyze the ability of different models to provide accurate deterministic predictions. In particular, we denote by $\tilde{f}(\delta_M, r) \in \mathbb{R}^4$ the vector whose all components are equal to zero, except from the one associated with the choice pattern having the highest predicted probability. To measure deterministic accuracy, we define the deterministic loss of model f as the fraction of choice patterns that are misclassified:

$$L_d(f) := \frac{1}{|\mathcal{D}| \times 500} \sum_{(\delta_M, r)\in\mathcal{D}} \sum_{i=1}^{500} \mathbb{1}\left(P_i(\delta_M, r) \neq f(\delta_M, r)\right),$$

where $|\mathcal{D}|$ denotes the cardinality of the set $\mathcal{D}$.

Table C.3: Choice patterns analysis: deterministic and probabilistic evaluations.

Exercise	Prediction	Loss	Models			
			EU	CPT	GBT	NN
Combined Within Task	Choice Pattern	Det.	6.1869 (1.1196)	6.1136 (1.1023)	**4.3690** **(0.7961)**	6.5877 (1.2214)
		Prob.	3.3573 (0.4031)	3.3824 (0.4277)	**2.5535** **(0.3010)**	3.6912 (0.4520)
Combined Across Task	Choice Pattern	Det.	**1.9933** **(0.0160)**	2.0041 (0.0939)	2.3563 (0.2575)	2.5958 (0.5927)
		Prob.	**1.6767** **(0.1415)**	1.7309 (0.1576)	1.9381 (0.2201)	2.1545 (0.3754)

Notes: Normalized loss functions with standard deviation in parentheses. "Det." stands for "Deterministic", while "Prob." stands for "Probabilistic". The smallest in-sample loss was obtained with the GBT algorithm in all out-of-sample exercises.

In general, lower values for both probabilistic and deterministic losses indicate higher probabilistic and deterministic accuracy of a model. However, interpreting the absolute magnitude of these losses can be challenging. To facilitate interpretation, we introduce a normalized loss measure. For a given loss L, its normalized version is defined as:[2]

$$\hat{L}(f) \coloneqq \frac{L(f)}{L^*},$$

where f is any generic model, and L^* is the lowest possible loss that can be achieved by training machine learning algorithms directly on the test data. Therefore, the normalized loss quantifies how many times greater the loss of a model trained on the training data is compared to the lowest possible loss that can be achieved by training a model directly on the test data.

Table C.3 summarizes the average normalized losses of the various methods for the out-of-sample exercises within and across tasks conducted in Section 3.3 and Section 3.3. The deterministic and probabilistic assessments of all predictive approaches yield the same result. The GBT outperforms EU and CPT in cross-validation exercises within CR-tasks and R-tasks. At the same time, EU achieves the smallest deterministic and probabilistic normalized losses in out-of-sample exercises across tasks. Furthermore, the values of the normalized losses inform us about the magnitude of a model's loss compared to the smallest achievable loss. Overall, all the normalized losses are significantly greater than one, indicating a substantial cost in

[2]See Fudenberg et al., 2022.

Table C.4: Adherence to EU: deterministic and probabilistic evaluations.

Exercise	Prediction	Loss	Models				
			EU	CPT	GBT	NN	EU-core
Combined Within Task	Index *Core*	Det.	6.3195 (1.1297)	6.1190 (1.1742)	**4.1648** **(0.7809)**	6.4227 (1.2513)	4.6769 (0.9830)
		Prob.	3.0751 (0.4194)	3.1845 (0.4557)	**2.4295** **(0.2879)**	3.4511 (0.4736)	2.9069 (0.3836)
Combined Across Task	Index *Core*	Det.	2.0843 (0.0316)	2.0778 (0.0671)	2.2207 (0.4566)	2.2038 (0.6289)	**1.9374** **(0.0401)**
		Prob.	**1.5927** **(0.1874)**	1.6656 (0.1966)	1.8322 (0.3495)	1.8215 (0.4960)	1.6006 (0.1505)

Notes: Normalized loss functions with standard deviation in parentheses. The smallest in-sample loss was obtained with the GBT algorithm in all out-of-sample exercises.

predictive accuracy when transitioning from in-sample to out-of-sample predictions.

EU-core Analysis. When the objective shifts from predicting choice patterns to forecasting the index $Core$, we can define a model as a function that associates each lottery pair (δ_M, r) and subject i, with a vector of characteristics X_i, to a vector $f(\delta_M, r; X_i) \in \mathbb{R}^3$. The j-th component of this vector represents the estimated probability under model f that the index $Core(\delta_M, r)$ assumes the value of $j \in 1, 2, 3$. For each model, we derive deterministic and probabilistic normalized loss functions, replicating the analysis previously described for choice patterns.

Table C.4 summarizes the average normalized losses of the various methods for the out-of-sample exercises within and across tasks conducted in Section 3.5. The GBT outperforms all other predictive approaches in cross-validation exercises within CR-tasks and R-tasks. The probabilistic performances of both EU and the EU-core approach are almost identical, with EU achieving a slightly smaller normalized loss. All other predictive approaches have significantly worse probabilistic performance.

C.5 Instructions

General Instructions. You will receive \$4 if you complete the entire study. We anticipate that the study will take about 20 minutes, on average. In addition to this payment, 1 out of every 5 participants will be randomly selected to receive a bonus payment. The smallest possible bonus payment is \$0 and the largest possible bonus payment is \$30. You will be informed of how your decisions will influence your bonus payment if you were to be randomly selected.

Description of the Experiment. In this study, we will ask you questions about *lotteries*. A lottery specifies different payments you may receive with different chances.

For example, one lottery might be the following:

Lottery

15% chance of **$17**

25% chance of **$9**

60% chance of **$2**

You can think of this lottery in the following way:

- In 15 out of 100 chances (15% chance) the lottery pays $17.
- In 25 out of 100 chances (25% chance) the lottery pays $9.
- In 60 out of 100 chances (60% chance) the lottery pays $2.

We may allow you to play a lottery and receive the outcome of the lottery as bonus payment. The outcome of the lottery will be determined by the computer using the chances specified. You will learn more about your bonus payment in the following instruction pages.

Check for Understanding. Before we proceed, here is a question to test your understanding. Consider the lottery below:

In how many out of 100 chances does this lottery pay $9?

o 15

o 25

o 60

Lottery

15% chance of **$17**

25% chance of **$9**

60% chance of **$2**

o 90

o None of the above

[Subjects are required to provide the correct answer in order to proceed. When they select a wrong answer, we show them the following error message: "The lottery pays $9 with a 25% chance. The correct answer is 25. Please revise your answer."]

Blocks of the Experiment. We will ask you to make choices over lotteries in two different blocks. If you are selected to receive a bonus payment, then we will randomly pick one of these blocks. We will describe within each block how your bonus would be determined if that block were randomly selected. Please click to learn about Block 1.

Block 1. In Block 1, we will show you two lotteries and will ask you to choose between the following two answer choices:

1. I prefer Lottery A
2. I prefer Lottery B

Block 1: Bonus Payment. If Block 1 is selected to determine your bonus payment, how would we pay you? We will randomly select one task from Block 1, and we will let you play the lottery you preferred. The lottery's outcome will be your bonus payment. We ask you to complete a brief training session to check your understanding of the tasks in Block 1.

Please proceed to start the training session!

Training task 1 of 2 - Block 1

We ask you to complete the following training task **assuming that**:

- You prefer **Lottery A over Lottery B**.

Lottery A	Lottery B
30% chance of **$40**	**40%** chance of **$30**
70% chance of **$5**	**60%** chance of **$10**

Which lottery do you **prefer**?

I prefer Lottery A ○ I prefer Lottery B ○

Training task 2 of 2 - Block 1

Suppose that this task is selected for bonus payment and your answer is:

Lottery A	Lottery B
30% chance of **$40**	**40%** chance of **$30**
70% chance of **$5**	**60%** chance of **$10**

Which lottery do you **prefer**?

I prefer Lottery A ◉ I prefer Lottery B ○

Please answer the following comprehension question:

How do we determine your bonus payment in this example?

○ We let you play Lottery A to determine your bonus payment.

○ We let you play Lottery B to determine your bonus payment.

○ We let you play Lottery A to determine your bonus payment with a 50 in 100 chance, or Lottery B with a 50 in 100 chance.

[Subjects are required to provide the correct answers in both training tasks in order to proceed. When they select a wrong answer in the first training task, we show them the following error message: "We are asking you to answer the question assuming that you prefer Lottery A. Please revise your answer." When they select a wrong answer in the second training task, we show them the following error message: "Given that you preferred Lottery A in this example, we would let you play Lottery A to determine your payment. Please revise your answer."]

Begin Block 1. Thank you for completing the training session. There will be 102 tasks in Block 1. Please answer all the questions thoughtfully to the best of your ability. Remember that there are no right or wrong answers. We are only interested in studying your preferences. Please proceed to start Block 1!

[Subjects complete CR-tasks, R-tasks and FOSD-tasks presented to them in a randomized order.]

Block 2. In Block 2, we will show you two lotteries and will ask you to choose between the following two answer choices:

1. I prefer Lottery A
2. I prefer Lottery B

Block 2: Bonus Payment. If Block 2 is selected to determine your bonus payment, how would we pay you? We will randomly select one task from Block 2, and we will let you play the lottery you preferred. The lottery's outcome will be your bonus payment.

Begin Block 2. There will be 33 tasks in Block 2. Please answer all the questions thoughtfully to the best of your ability. Remember that *there are no right or wrong answers*. We are only interested in studying your preferences.

Please proceed to start Block 2!

Block 2 - Part 1. In the following 11 tasks of Block 2:

- **Lottery A** pays a monetary amount with a **100%** chance that is different in every task. In task 1, the monetary amount paid for sure by Lottery A is **$3**. As you move from one task to the next one, the monetary amount paid for sure by Lottery A increases by $1. For instance, in task 2 is **$4**, in task 3 is **$5**, etc.
- **Lottery B** always pays **$20** with a **50%** chance, or **$0** with a **50%** chance.

[Subjects complete the 11 MPL1 tasks.]

Block 2 - Part 1. In the following 11 tasks of Block 2:

- **Lottery A** pays a monetary amount with a **100%** chance that is different in every task. In task 12, the monetary amount paid for sure by Lottery A is **$8**. As you move from one task to the next one, the monetary amount paid for sure by Lottery A increases by $1. For instance, in task 13 is **$9**, in task 14 is **$10**, etc.
- **Lottery B** always pays **$25** with a **50%** chance, or **$5** with a **50%** chance.

[Subjects complete the 11 MPL2 tasks.]

Block 2 - Part 1. In the following 11 tasks of Block 2:

- **Lottery A** pays a monetary amount with a **100%** chance that is different in every task. In task 23, the monetary amount paid for sure by Lottery A is **$13**. As you move from one task to the next one, the monetary amount paid for sure by Lottery A increases by $1. For instance, in task 24 is **$14**, in task 25 is **$15**, etc.
- **Lottery B** always pays **$30** with a **50%** chance, or **$10** with a **50%** chance.

[Subjects complete the 11 MPL3 tasks.]