

Online Causal Inference for Advertising in Real-Time Bidding Auctions*

Caio Waisman Harikesh S. Nair Carlos Carrion Nan Xu

This draft: March 8, 2021

Abstract

Real-time bidding (RTB) systems, which leverage auctions to programmatically allocate user impressions to multiple competing advertisers, continue to enjoy widespread success in digital advertising. Assessing the effectiveness of such advertising remains a lingering challenge in research and practice. This paper presents a new experimental design to perform causal inference on advertising bought through such mechanisms. Our method leverages the economic structure of first- and second-price auctions, which are ubiquitous in RTB systems, embedded within a multi-armed bandit (MAB) setup for online adaptive experimentation. We implement it via a modified Thompson sampling (TS) algorithm that estimates causal effects of advertising while minimizing the costs of experimentation to the advertiser by simultaneously learning the optimal bidding policy that maximizes her expected payoffs from auction participation. Simulations show that not only the proposed method successfully accomplishes the advertiser’s goals, but also does so at a much lower cost than more conventional experimentation policies aimed at performing causal inference.

Keywords: Causal inference, multi-armed bandits, advertising auctions

*This paper was previously circulated under the title “Online inference for advertising auctions.” The first three authors are part of JD Intelligent Ads Lab. We thank Tong Geng, Jun Hao, Xiliang Lin, Lei Wu, Paul Yan, Bo Zhang, Liang Zhang and Lizhou Zheng from JD.com for their collaboration; seminar participants at Cornell Johnson, Berkeley EECS, Stanford GSB: OIT/Marketing, UCSD Rady, Yale SOM; at the 2019 Conference on Structural Dynamic Models (Chicago Booth), the 2019 Midwest IO Fest, the 2020 Conference on AI/ML and Business Analytics (Temple Fox), and the 2020 Marketing Science Conference; and Mohsen Bayati, Rob Bray, Isa Chaves, Yoni Gur, Yanjun Han, Günter Hitsch, Kanishka Misra, Sanjog Misra, Rob Porter and Stefan Wager in particular for helpful comments. Please contact the authors at caio.waisman@kellogg.northwestern.edu (Waisman); harikesh.nair@stanford.edu (Nair); carlos.carrion@jd.com (Carrion); or nanxu@umn.edu (Xu) for correspondence.

1 Introduction

The dominant way of selling ad-impressions on ad-exchanges (AdXs) is now through real-time bidding (RTB) systems. These systems leverage auctions at scale to allocate user impressions arriving at various digital publishers to competing bidding advertisers or intermediaries such as demand side platforms (DSPs). Most RTB auctions on AdXs are single-unit auctions implemented typically via second-price auctions (SPAs) and, more recently, via first-price auctions (FPAs).¹ The speed, scale and complexity of RTB transactions is astounding. There are billions of auctions for impressions on an AdX on any given day, each consummated on average in less than 100 milliseconds, and each transaction corresponds to different impression characteristics and possibly requires a unique bid and clearing price. The complexity and scale of available ad-inventory, the speed of the transaction that allows little time for deliberation and the complex nature of competition imply that advertisers participating in RTB auctions have significant uncertainty about the value of the impressions they are bidding for as well as the nature of competition they are bidding against. Developing accurate and reliable ways of measuring the value of advertising in this environment is therefore essential for the advertiser to profitably trade on the exchange and to ensure that acquired ad-impressions generate sufficient value. Measurement needs to deliver incremental effects of ads for different types of ad and impression characteristics and needs to be automated. Experimentation thus becomes attractive as a device to obtain credible estimates of such causal effects.

Motivated by this, the current paper presents an experimental design to perform causal inference on RTB advertising in both SPA and FPA settings. Our design enables learning heterogeneity in the inferred average causal effects across ad and impression characteristics. The novelty of the proposed experimental design is in addressing two principal challenges that confront developing a scalable experimental design for RTB advertising.

The first challenge is that measuring the average treatment effect (*ATE*) of advertising requires comparing outcomes of users who are exposed to ads with outcomes of users who are not. The complication of the RTB setting is that ad-exposure is not under complete control of the experimenter because it is determined by an auction. This precludes simple designs in which ad-exposure is randomized directly to users. Instead, we need a design in which the experimenter controls only an input to the auction (bids), but

¹See Muthukrishnan (2009) and Choi et al. (2020) for an overview of the economics of AdXs and Despotakis et al. (2019) for the reasons behind the recent shift to FPAs.

wishes to perform inference on the effect of a stochastic outcome induced by this input (ad-exposure).

The second challenge pertains to the cost of experimentation on the AdX. Obtaining experimental impressions is costly: one has to win the auction to see the outcome with ad-exposure, and one has to lose the auction to see the outcome without. When bidding is not optimized, the economic costs of both can be substantial. These economic costs arise when one bids too much to obtain an impression with ad-exposure (overbidding) or decides deliberately not to induce ad-exposure on what would have been a very valuable impression for the advertiser (underbidding/opportunity cost). With millions of auctions potentially in the experiment, suboptimal bidding can make the experiment unviable. Therefore, to be successful an effective experimental design has to deliver inference on the causal effect of ads while also judiciously managing the cost of experimentation by implementing a bidding policy that is optimized to the value of the impressions encountered.

It is not obvious how an experiment can be designed prior to actual implementation to address both challenges simultaneously. Optimal bidding requires knowing the value of each impression, which was the goal of experimentation in the first place. *Online* methods, which introduce randomization to induce exploration of the value of advertising and combine it with concurrent exploitation of the information learned to optimize bidding, thus become highly attractive in such a setting.

At the core of online methods is the need to balance the goal of learning the expected effect of ad-exposure (henceforth called the advertiser’s “inference goal”) with the goal of learning the optimal bidding policy (henceforth called the advertiser’s “economic goal”). The tension is that finding the best bidding policy does not guarantee proper estimation of ad-effectiveness and vice versa. At one extreme, with a bidding policy that delivers on the economic goal, the advertiser could win most of the time, making it difficult to measure ad-effectiveness since outcomes with no ad-exposures would be scarcely observed. At the other extreme, with pure bid randomization the advertiser could estimate unconditional ad-effectiveness (the *ATE*) or how ad-effectiveness varies with observed heterogeneity (the *CATEs*) and deliver on the inference goal, but may end up incurring large economic losses in the process.

The contribution of this paper is to present a multi-armed bandit design (MAB) and statistical learning framework that address both these considerations. In our design, observed heterogeneity is summarized by a context, x , bids form arms, and the advertiser’s

monetary payoffs form the rewards, so that the best arm, or optimal bid, maximizes the advertiser’s expected payoff from auction participation given x . Exploiting the economic structure of SPAs and FPAs, we derive, under conditions we outline, the relationship between the optimal bid at a given x and the *CATE* of the ad at the value x , or $CATE(x)$. For SPAs, we show that these two objects are equal, so that, in our experimental design, the twin tasks of learning the optimal bid and estimating ad-effectiveness not only can build off each other, but are perfectly aligned. For FPAs, we demonstrate that the two goals are closely related, though only imperfectly aligned. Leveraging this relationship improves the efficiency of learning and generates an online experimental design that delivers on both goals for the advertiser at minimal cost under both auction formats.

To implement the design, we present a Thompson Sampling (TS) algorithm customized to our auction environment trained online via a Markov Chain Monte Carlo (MCMC) method. The algorithm adaptively chooses bids across rounds based on current estimates of which arm is the optimal one. These estimates are updated on each round via MCMC through Gibbs sampling, which leverages data augmentation to impute the missing potential outcomes and the censored or missing competing bids in each round. Simulations show that the algorithm is able to recover treatment effect heterogeneity as represented by the *CATE*s of advertising and considerably reduces the expected costs of experimentation compared to non-adaptive “A/B/n” tests, explore-then-commit (ETC) strategies and a canonical off-the-shelf TS algorithm.

The rest of the paper discusses the relationship between the approach presented here with the existing literature and explains our contribution relative to existing work. The following section defines the design problem. Sections 3 and 4 show how we leverage auction theory to balance the experimenting advertiser’s objectives. Section 5 presents the modified TS algorithm we use to implement the experimental design. Section 6 shows extensive simulations documenting the performance of the proposed algorithm and shows its advantages over competing designs. The last section concludes.

Related literature

Our paper lies at the intersection of three broad fields of study: the literature on online learning in ad-auctions, the literature on experimentation in digital advertising, and the literature on causal inference with bandits. Given that each of these streams of literature is mature, we discuss only the most related papers for brevity, and the reader is referred

to some of the cited papers for further reading.

Literature on online learning in ad-auctions

While the primary focus of this paper is on causal inference and experimentation, to the extent that we solve an online learning-to-bid problem during experimentation, this paper is also closely related to the literature on online learning in ad-auctions, in which learning strategies are suggested for use by auction participants to make good auction-related decisions. One feature of this literature is that the majority of studies adopt the seller’s perspective, focusing on the problem of designing mechanisms that maximize the seller’s expected profit (e.g., finding an optimal reserve price in SPAs when the distribution of valuations of the bidders is unknown *a priori* to the auctioneer). Examples include [Cesa-Bianchi et al. \(2014\)](#), [Mohri and Medina \(2016\)](#), [Roughgarden and Wang \(2019\)](#), [Haoyu and Wei \(2020\)](#) and [Kanoria and Nazerzadeh \(2021\)](#).

Our study, which addresses the problem of ad-experimentation from the advertiser’s perspective, is more closely related to a smaller subliterature within this stream at the intersection of online learning and auction theory that has studied the problem of bidding from the perspective of the bidder. The key issue is that RTB bidders have significant uncertainty about the value of advertising they buy on AdXs. In turn, MAB policies are appealing devices for AdXs to learn which advertisers to pick as suggested in [McAfee \(2011\)](#), although neither a specific algorithm for advertisers to learn-to-bid nor a strategy for them to conduct online experimentation on AdXs is outlined.

[Balseiro et al. \(2019\)](#) present an algorithm for contextual bandits with cross-learning across contexts and discuss an application to advertisers learning-to-bid in FPAs. Their insight is that if we consider the value of the ad to the advertiser, v , as the context, observing the reward from bidding at context v yields information about the reward at another context v' for the same bid, so there is cross-learning across contexts within each arm. They show that leveraging this information improves the efficiency of learning-to-bid algorithms.² However, the informational assumption in [Balseiro et al. \(2019\)](#) – that the bidder’s value for the item (i.e., the context) is known to her prior to bidding – implies that there is no scope for causal inference on this value to the bidder, unlike the situation we consider here. [Han et al. \(2020b\)](#) extend this analysis and present stochastic bandit

²This feature is shared by the bandit problem presented in this paper as well, with an added advantage that the existence of a shared payoff structure across arms implies that the problem displays *cross-arm* learning in addition to the within-arm, cross-context learning pointed out by [Balseiro et al. \(2019\)](#).

algorithms for learning-to-bid in FPAs. Han et al. (2020a) present an analogous algorithm for the adversarial case. They discuss ways to overcome a key impediment to resolving the uncertainty a learning advertiser has over competitors' bids, which is that the highest competing bid is not observed when she wins the auction, leading to censored feedback and a form of a winner's curse in learning. Han et al. (2020b) and Han et al. (2020a) maintain the same assumption as Balseiro et al. (2019): bidders have uncertainty about the distribution of competing bids, but know their own valuation exactly prior to bidding. Therefore, again, there is no role for causal inference on the intrinsic value of advertising from the advertiser's perspective.³

Weed et al. (2016) and Feng et al. (2018) relax the assumption that advertisers know their private value and present learning-to-bid algorithms for SPAs. Like the above papers, their emphasis is on addressing the challenge associated with the censored "win-only" or "outcome-based" feedback that arises in SPAs. While these papers relax the assumption of exact knowledge of the ad's value by the advertiser, the way the informational assumptions are relaxed is different from ours, and implies a much more limited scope for causal inference on the value of the ad to the advertiser. Weed et al. (2016)'s assumption is that the bidder is uncertain about her valuation prior to bidding, but the value of the good is fully observed if the auction is won. In turn, Feng et al. (2018)'s assumption is that the bidder is uncertain about her valuation prior to bidding, but the value of the ad is fully observed if the auction is won and the ad is clicked on. This implies an implicit assumption that the impression generates no value to the advertiser when the auction is lost and no ad-exposure or ad-click occurs. In canonical auction-theoretic settings, it is assumed that the bidder gets no utility if she loses the auction. However, in the context of ads a user may have a non-zero propensity to buy the advertiser's products even in the absence of ad-exposure or ad-click, and winning the auction increments this propensity, so this assumption may not be as well suited.

Relaxing this assumption changes the implied inference problem substantively. In causal inference terms, if we call $Y(1)$ and $Y(0)$ respectively the potential outcomes to the advertiser with and without ad-exposure to the user, the value of the ad is $Y(1) - Y(0)$. The assumption above implies that $Y(0) \equiv 0$. This assumes away a principal challenge in causally inferring the effect of ads addressed in this paper, that $Y(1)$ and $Y(0)$ are not observed together for the same impression because the auction outcome censors the po-

³Han et al. (2020b)'s informational assumptions regarding competing bids are also different from ours: we assume the advertiser does not observe the highest competing bid in an FPA when winning or losing (full censoring); they assume the advertiser does not observe the highest competing bid if she wins, but observes it if she loses (partial censoring).

tential outcomes (this is [Holland \(1986\)](#)’s “fundamental problem” of causal inference). Performing inference on the effect of ads while addressing this censoring problem is a major focus of the experimental design in this paper. Another difference is that when winning helps learn the valuation fully, the bandit’s exploration motive prioritizes higher bids, because that increases the chance of winning. In our setting, learning the value of advertising necessarily requires losing the auction sometimes, because that allows observing $Y(0)$. Hence, this force for higher bidding in exploration is less pronounced, which in turn affects the nature of bidding and the induced cost of experimentation.

This paper is related to a series of recent studies that analyze market equilibrium when learning bidders interact in repeated RTB auctions. [Dikkala and Tardos \(2013\)](#) characterize an equilibrium credit provision strategy for an ad-publisher who faces bidders that are uncertain about their values from participating in SPAs for ads. The credit provided by the publisher to bidders incentivizes them to experiment with their bidding to resolve their uncertainty, and, when set optimally, improves the publisher’s revenue. [Iyer et al. \(2014\)](#) adopt a mean-field approximation for large markets, in which bidders track only an aggregate and stationary representation of the competitors’ bids to study repeated SPAs where bidders learn about their own private values over time. They apply it to the problem of selecting optimal reserve prices for the publisher. [Balseiro et al. \(2015\)](#) characterize equilibrium strategic interactions between budget-constrained advertisers who face no uncertainty about their private valuations, but have uncertainty about the number, bids and budgets of their competitors. They also develop a fluid mean-field approximate equilibrium, and use their characterization to recommend optimal budget pacing strategies for advertisers and optimal reserve prices for the publisher. Finally, [Balseiro and Gur \(2019\)](#) and [Tunuguntla and Hoban \(2021\)](#) provide pacing algorithms and characterize equilibria in bidding when budget-constrained advertisers who observe their current private values before bidding (but are uncertain about their and competitors’ future values and budgets) interact over time in repeated SPAs. [Tunuguntla and Hoban \(2021\)](#) also discuss augmenting their algorithm with bandit exploration when the advertiser’s valuation has to be learned. Overall, the goals of these papers – to characterize equilibria and to suggest equilibrium budget pacing strategies, credits or reserve prices – are different from ours, which is to develop an experimental design from the advertiser’s perspective for causal inference on ads bought via SPAs and FPAs. We note the methods in these studies could form the basis for extending the analysis in this paper to develop experimental designs for simultaneous experimentation by multiple advertisers on an AdX.

There is a smaller subliteration that uses more general reinforcement learning (RL)

approaches beyond bandits to optimize RTB advertising policies (Cai et al., 2017; Wu et al., 2018; Jin et al., 2018). These papers are not concerned with estimating ad-effects, and we depart from these approaches in that our goal is to perform causal inference. Further, we achieve this goal by leveraging key properties of the auction format, thus contributing to a nascent literature, to our knowledge, on direct applications of auction theory to enable causal inference. While several studies combined experimental designs with auction theory, their goals were to identify optimal policies such as bids as in the aforementioned studies, reserve prices (Austin et al., 2016; Ostrovsky and Schwarz, 2016; Pouget-Abadie et al., 2018; Rhuggenaath et al., 2019) or auction formats (Chawla et al., 2016), not to estimate the causal effect of an action such as advertising determined by the outcome of an auction.

Literature on experimentation in digital advertising

This paper is related to the literature on pure experimental approaches to measure the effect of digital advertising.⁴ One feature that distinguishes this paper from several studies in this stream is its focus on developing an experimental design from the advertiser’s perspective; in contrast, many of the proposed experimental designs, such as “ghost-ads” for display advertising (Johnson et al., 2017) or search ads (e.g., Simonov et al., 2018), are designed from the publisher’s perspective and require observation of the ad-auction logs or cooperation with the publisher for implementation. In RTB settings, the advertiser bidding on AdXs does not control the auction and does not have access to these logs, precluding such designs (the next section provides a more detailed discussion).

Existing experimental designs that have been proposed from the advertiser’s perspective include geo-level randomization (e.g., Blake et al., 2015) or inducing randomization of ad-intensities by manipulating ad campaign frequency caps on DSPs (e.g., Sahni et al., 2019). Unlike these papers, our design leverages bid randomization, is tailored to the RTB setting, is an online, rather than offline, inferential procedure, and leverages auction theory for inference.

Lewis and Wong (2018) suggest using bid randomization as a device to infer the causal effects of RTB ads. Their method uses bids as an instrumental variable for ad-exposure and delivers estimates of the local average treatment effect of ads, unlike the

⁴For a more detailed review we refer the reader to Gordon et al. (2021), who provide in their Section 1 a recent overview and critical discussion.

experimental design proposed here, which leverages the link to auction theory to deliver *ATEs* for SPAs and FPAs. Also, unlike the experimental design outlined here, their method is not adaptive and involves only pure exploration. Therefore, it does not have the feature that the bid randomization policy also minimizes the costs from experimentation by concurrently exploiting the information learned during the test to optimize advertiser payoffs. Finally, adaptive experimental designs for picking the best creative for ad-campaigns are presented in Scott (2015), Schwartz et al. (2017), Ju et al. (2019) and Geng et al. (2020). While related, the problem addressed in these papers of selecting a best performing variant from a set of candidates is conceptually distinct from the problem addressed in this paper of measuring the causal effect of an RTB ad-campaign.

Finally, while there are differences in implementation, our philosophy towards experimental design – which aligns the goal of the design with the payoff-maximization objective of the advertiser – is aligned with that of Feit and Berman (2019), who advocate ad-experimental designs that are profit maximizing. The advantage of the bandit-based allocation here is that traffic is adaptively assigned to bid-arms in a way that respects the advertiser’s profits, analogous to Feit and Berman (2019)’s setup. Some salient differences in implementation are that we adopt a many-period bandit-allocation, while Feit and Berman (2019) use a two-period setup, and that our design shows how to implement profit maximizing tests for outcomes over which the advertiser only has imperfect control.

Literature on causal inference with bandits

There is an emerging literature that discusses online causal inference with bandits in various general settings.⁵ Performing causal inference with bandits is complicated by the adaptive nature of data collection, wherein future data collection depends on the data already collected. Although bandits possess attractive properties in finding the best arm, estimates of arm-specific expected rewards typically exhibit a bias often referred to as “adaptive bias” (Xu et al., 2013; Villar et al., 2015). In particular, Nie et al. (2018) show that archetypal bandit algorithms such as Upper Confidence Bound (UCB) and TS compute

⁵Note that existing approaches to inference with bandits differ based on whether they pertain to the *offline* setting, where pre-collected data is available to the analyst, or the *online* setting, where data arrive sequentially, with the online literature being relatively more recent. Unlike online methods, offline methods are meant to be implemented *ex-post*, which implies that the data collection, though done sequentially, is typically not made explicitly to facilitate inference upon its completion. Also, the methods are meant for retrospective application on logged data, which means that data collection does not explicitly reflect in real-time the progress made towards the inferential goal. This paper relates more closely to the online stream, so we discuss only papers related to online inference.

estimates of arm-specific expected rewards that are biased downwards. Due to this problem, leveraging bandits for causal inference for RTB ads is complicated even in a simple case where assignment of users to ads is fully under the control of the advertiser. If we set up ad-exposure and no-ad-exposure as bandit arms, so that the difference in rewards between the two arms represents the causal effect of the ad, adaptive bias in estimating the respective arm-rewards contaminates this difference.

Online methods to find the best arm while correcting for such adaptive bias include Goldenshluger and Zeevi (2013), Nie et al. (2018), and Bastani and Bayati (2020), who suggest data-splitting by forced-sampling of arms at prescribed times, and Dimakopoulou et al. (2018) and Hadad et al. (2021), who correct for the bias via balancing and inverse probability weighting. Online methods to perform frequentist statistical inference that is valid for bandits, but which avoid issues of explicitly bias-correcting estimates of arm-specific expected rewards, are presented in Yang et al. (2017), Jamieson and Jain (2018) and Ju et al. (2019).

This paper has a different focus on inference compared to the above studies. Broadly speaking, the above methods aim to either find the best arm or learn without bias the expected reward associated with the best arm. In contrast, our goal is to obtain an unbiased estimate of the effect of an action (ad-exposure) that is *imperfectly obtained by pulling arms* (placing bids). Therefore, in our setup the target treatment whose effect is to be learned is not an arm, but a shared stochastic outcome that arises from pulling arms. Hence, arms are more appropriately thought of as *encouragements* for treatments, which makes our setup the online analogue of an offline encouragement design from the program evaluation literature (e.g., Imbens and Rubin, 1997). In addition to this difference, our bandit design, which treats bids and their associated payoffs as arms/rewards (rather than ad-exposure and the payoff of ad-exposure as arms/rewards), presents a different approach to avoiding the aforementioned adaptive bias. In our approach, we obtain the object we would like to estimate and perform inference on via the *identity* of the best arm and the theoretical relationship between the two rather than the expected value of its reward. Since typical MAB algorithms recover the identity of the best arm without bias, we are able to leverage them for inference on ad-effects without bias in an online environment. Again, this is achieved by maintaining a close link to auction theory, which makes our approach different in spirit from the above, more theory-agnostic approaches.

Our setup also shares similarities with the instrument-armed bandit setup of Kallus (2018), in which there is a difference between the treatment-arm pulled and the treatment-

arm applied due to the possibility of user non-compliance. However, the difference between the pulled and applied treatments, which is important to the design here, is not a feature of the design considered by Kallus (2018), because the pulled and applied treatments belong to the same set in his design. Also, unlike the setup in Kallus (2018), where exposure to a treatment is the outcome of a choice by the user to comply with the pulled arm, exposure here is obtained via a multi-agent game that is not directly affected by the user, thus characterizing a different exposure mechanism.

Bandits have been explicitly embedded within the *structural causal framework* of Pearl (2009) in a series of papers by Bareinboim et al. (2015), Lattimore et al. (2016) and Forney et al. (2017). Our paper is related to this stream as our application is a specific instance of a structural causal model tailored to the auction setting: we assume the existence of a probabilistic, microfounded generative process that is the common shared causal structure behind the distributions of the rewards for each arm. As this stream has emphasized, the link to the model in our application is helpful to making progress on the inference problem. This approach has also been followed by other papers in economics (see, for example, the references in Bergemann and Välimäki, 2008) and marketing (Misra et al., 2019) that study pricing problems where firms aim to learn the optimal price from a grid of prices, corresponding to arms, which share the same underlying demand function.

2 Problem formulation

Our goal is to develop an experimental design to measure the causal effect of the ads an advertiser buys programmatically on AdXs. To buy the ad, the advertiser (she) needs to participate in an auction ran by the AdX for each impression opportunity. Winning the auction allows the advertiser to display her ad to the user. The AdX's auction format can either be a SPA or a FPA.

Recall that we define the advertiser' goal of estimating the expected effect of displaying the ad to a user (he) as her *inference goal*. To state this goal more precisely, let $Y(1)$ denote the revenue the advertiser receives when her ad is shown to the user and let $Y(0)$ denote the revenue she receives when her ad is not shown. $Y(1)$ and $Y(0)$ are potential outcomes to the advertiser expressed in monetary units and the causal effect of the ad is $Y(1) - Y(0)$. All the information the advertiser has about the user and impression opportunity is captured by a variable x that can take P different values, so that

$x \in \mathbb{X} \equiv \{x_1, \dots, x_P\}$.⁶

The advertiser’s inference goal specifically is to estimate a set of *conditional average treatment effects* (CATEs) in which exposure to the ad is the treatment, where $CATE(x) = \mathbb{E}_{1,0}[Y(1) - Y(0)|x]$. The CATEs represent heterogeneity in the average treatment effect of the ad across subsets of users spanned by x . The subscripts “1,0” on the expectation operator in the CATE explicitly indicate that the expectation is taken with respect to the conditional distribution of $Y(1)$ and $Y(0)$ on x , about which the advertiser has uncertainty. The advertiser needs to estimate this object because she does not have complete knowledge of the distribution of potential outcomes $Y(1)$ and $Y(0)$ conditional on x . Therefore, achieving the inference goal requires the collection of data informative of this distribution.

An experimental design that delivers on the advertiser’s inference goal needs to address four issues, which we discuss in sequence below. All four are generated by the distinguishing feature of the AdX environment that the treatment – ad-exposure – can only be obtained by winning an RTB auction.

Issue 1: The advertiser cannot randomize treatment directly

The first issue is that the existence of the mediating auction precludes typical experimental approaches to measuring treatment effects that involve collecting data while randomizing treatment and then using these data to estimate CATEs. The outcome of the RTB auction is not under the advertiser’s complete control because she does not determine the highest competing bid on the AdX. This lack of control disallows her from randomizing the treatment, ad-exposure.

Although ad-exposure cannot be perfectly controlled, participation in auctions is fully under the advertiser’s control. Therefore, a viable alternative design involves randomizing her participation across auctions. Without loss of generality, we can think of auction participation randomization as analogous to *bid randomization*, with “no participation” corresponding to a bid of 0, and “participation” corresponding to a positive bid. Consequently, we could consider bid randomization as an alternative identification strategy to recover CATEs in this environment.

⁶In our MAB setup, x is the *context* of the auction. It can be obtained from a vector Z of observable display opportunity variables that can include, for example, user and publisher characteristics, with P being the total number of different combinations of values across all elements of Z . Consumer segmentation of user characteristics in this manner is common in the literature; see, for instance, [Misra et al. \(2019\)](#).

Issue 2: Bid randomization alone is insufficient for identification

This leads to the second issue: the auction environment constrains what can be learned from participation/bid randomization experimental designs. In particular, bid randomization is generally *not* sufficient to yield identification of *CATEs* if the relationship between $Y(1)$, $Y(0)$ and competing bids remains completely unrestricted.

To see this, let the highest bid against which the experimenting advertiser competes in the auction be denoted by B_{CP} . For simplicity, assume that $P = 1$ so that x can only take one value and hence can be ignored. Let $D \equiv \mathbb{1}\{B_{CP} \leq b\}$ denote winning the auction (i.e., ad-exposure), where b is the experimenting advertiser's bid. Let $Y \equiv D \times Y(1) + (1 - D) \times Y(0)$, $Y(1) = \lambda_1 + \eta_1$ and $Y(0) = \lambda_0 + \eta_0$, where λ_1 and λ_0 are constants and $\mathbb{E}[\eta_1] = \mathbb{E}[\eta_0] = 0$, so that $ATE = \lambda_1 - \lambda_0$.

Consider measuring the *ATE* via a regression. A regression of Y on D using data collected with bid randomization corresponds to:

$$Y_i = \lambda_0 + ATE \times D_i + D_i \eta_{1i} + (1 - D_i) \eta_{0i}, \quad (1)$$

where i indexes an observation. For the OLS estimator of *ATE* to be consistent the indicator D_i has to be uncorrelated with the errors η_{1i} and η_{0i} , and there are two potential sources of such correlation, b_i and $B_{CP,i}$. Thus, even if b_i is picked at random, a correlation can still exist through $B_{CP,i}$.

This consideration motivates Assumption 1 below, which we maintain for the rest of the paper.

Assumption 1. *“Private values” / Conditional independence*

$\{Y(1), Y(0)\} \perp\!\!\!\perp B_{CP} | x$.

One way to interpret Assumption 1 is from the perspective of auction theory. Noting that the value of the ad-exposure to the advertiser depends on $Y(1)$ and $Y(0)$, Assumption 1 implies that, conditional on x , knowledge of B_{CP} has no effect on the experimenting bidder's assessment of $Y(1)$ and $Y(0)$, and, consequently, on her assessment of her own willingness-to-pay, or valuation. Therefore, we can think of Assumption 1 as analogous to a typical private values condition.⁷ This does not mean that there are no correlations

⁷We say that it is “analogous to” but not exactly a private values condition because the specific model of ad-auctions that we will present below is slightly different from the canonical model for auctions presented

between the values that competing advertisers assign to an auctioned impression. The maintained assumption is that these common components of bidder valuations are accommodated in the observed vector x . Part of the motivation arises from the programmatic bidding environment. When advertisers bid in AdXs, they match the id (cookie/mobile-identifier) of the impression with their own private data. If x is large and encompasses what can be commonly known about the impression, each advertiser’s private value after conditioning on x would only be weakly correlated.

Another way to think of Assumption 1 is in causal inference terms as an unconfoundedness assumption. Statistically, it simply is a conditional independence assumption, which is more likely to hold when x is large and spans the common information set that auction participants have about the user impression. This is most likely to happen when the experimenter is a large advertiser or an intermediary such as a large DSP, which has access to large amounts of user data that can be matched to auctioned user impressions in real-time.

Along with bid randomization, Assumption 1 yields identification of $CATE(x)$. For instance, from equation (1) we see that OLS identifies ATE consistently under this condition, because both b and $B_{CP,i}$ are independent of η_{1i}, η_{0i} .⁸

Issue 3: High costs of experimentation

While bid randomization combined with Assumption 1 is sufficient for identification, a third issue to consider is the cost of experimentation. As mentioned in the introduction, inducing ad-exposure involves paying the winning price in the auction, and inducing no-

in the theory literature (e.g., Milgrom and Weber, 1982). In the canonical model, a bidder obtains a signal about the item she is bidding on prior to the auction, and uses that to form her expectation of its value. Under a private values condition, it is then without loss of generality to normalize the bidder’s valuation to be the signal itself (e.g., Athey and Haile, 2002, pp. 2110–2112). If we apply that formulation here, this would imply that the signal would correspond to the treatment effect $Y(1) - Y(0)$ itself. This formulation simply does not fit our empirical setting in which the treatment effect is not known to the advertiser prior to bidding (this is the motivation for running the experiment in the first place). Reflecting this, the model we present does not have signals. As a consequence, it does not map exactly to the canonical dichotomy between private and interdependent values, which is framed in terms of bidders’ signals.

⁸Another advantage is that, under this assumption, the target estimand, $CATE(x)$, becomes a common component of the expected reward associated with all the bid-arms for a given x , thereby facilitating cross-arm learning of both this estimand and optimal bids. If potential outcomes were instead correlated with others’ bids given x , the resulting expected reward associated with each bid-arm would be not be a function of the $CATE(x)$, but rather of a more complex expectation of the potential outcomes that varies across arms, precluding such cross-arm learning efficiencies. See also Balseiro et al. (2019) who make a similar independence assumption and point out its usefulness for cross-context learning.

exposure involves foregoing the value generated from ad-display. Therefore, collecting experimental units on the AdX involves costs.

These costs can be high under suboptimal bidding. Bidding higher than what is required to win the auction involves wastage of resources from overbidding, and bidding 0 involves possible opportunity costs from underbidding, especially on users to whom ad-exposure would have been beneficial. In a high frequency RTB environment, a typical experiment can involve millions of auctions, so that if bidding is not properly managed, the resulting economic losses can be substantial.⁹ These costs can form an impediment to implementation of the experiment in practice, precluding its use.

The key to controlling costs is to optimize the bidding, specifically by finding the optimal bid, b^* , to submit to the auction for each value of x . Henceforth, we refer to the advertiser's goal of obtaining the optimal bidding policy in the experiment, $b^*(x)$, as her *economic goal*. As we discuss in more detail below, the advertiser's inference and economic goals are directly related, though not necessarily perfectly aligned.

To characterize the optimal bid more formally and relate it to the advertiser's inference goal, we first turn to the advertiser's optimization problem from auction participation. The advertiser's payoff as a function of her bid, b , is denoted $\pi(b, Y(1), Y(0), B_{CP})$.

In an SPA, $\pi(\cdot)$ is

$$\begin{aligned}\pi(b, Y(1), Y(0), B_{CP}) &= \mathbb{1}\{B_{CP} \leq b\} \times [Y(1) - B_{CP}] + \mathbb{1}\{B_{CP} > b\} \times Y(0) \\ &= \mathbb{1}\{B_{CP} \leq b\} \times \{[Y(1) - Y(0)] - B_{CP}\} + Y(0),\end{aligned}\quad (2)$$

while in an FPA, $\pi(\cdot)$ is

$$\begin{aligned}\pi(b, Y(1), Y(0), B_{CP}) &= \mathbb{1}\{B_{CP} \leq b\} \times [Y(1) - b] + \mathbb{1}\{B_{CP} > b\} \times Y(0) \\ &= \mathbb{1}\{B_{CP} \leq b\} \times \{[Y(1) - Y(0)] - b\} + Y(0).\end{aligned}\quad (3)$$

As an aside, notice the formulation of auction payoffs in equations (2) and (3) is different from typical set-ups. As mentioned previously, in most auction models the term $Y(0)$ is set to zero because it is assumed that a bidder only accrues utility when she wins the auction. However, this convention is not suitable to our setting given the interpreta-

⁹With a fixed budget, poor bidding also affects the quality of inference when wastage of experimental resources leads to reduced collection of experimental data, leading to smaller samples and reduced statistical precision.

tion of the terms $Y(1)$ and $Y(0)$. In particular, a consumer might have a baseline propensity to purchase the advertiser’s product even if he is not exposed to her ad, which is associated with the term $Y(0)$. Exposure to the ad might affect this propensity, further implying that $Y(1) \neq Y(0)$.

Optimal bidding

The advertiser is assumed to be risk-neutral, and to look for a bid that maximizes the expected payoff from participating in the auction,

$$\bar{\pi}(b|x) = \mathbb{E}_{1,0,CP}[\pi(b, Y(1), Y(0), B_{CP})|x], \quad (4)$$

where the subscripts “1, 0, CP” on the expectation operator indicate that the expectation is taken with respect to the conditional distribution of $Y(1)$, $Y(0)$ and B_{CP} on x . Analogous to the assumption that the advertiser does not fully know the distribution of $Y(1)$ and $Y(0)$, we assume that she also does not know the distribution of B_{CP} . This implies that the advertiser faces uncertainty over the joint conditional distribution of $Y(1)$, $Y(0)$ and B_{CP} on x , which we denote by $F(\cdot, \cdot, \cdot|x)$. It is important to note that the fact that we postulate that there exists a distribution for B_{CP} does not imply that competitors are randomizing bids or following mixed strategies, although it does allow for it. As typical in game-theoretic approaches to auctions, a given bidder, in this case the advertiser, treats the actions taken by her competitors as random variables, which is why we treat B_{CP} as being drawn from a probability distribution.¹⁰

Solving for $b^*(x)$ in the presence of uncertainty over $F(\cdot, \cdot, \cdot|x)$ is a non-standard and highly non-trivial auction problem. The problem is non-standard because under the outlined circumstances, the advertiser faces two levels of uncertainty. The first, “lower-

¹⁰This modeling approach for competition, summarizing it by B_{CP} , is reduced-form. Typically B_{CP} is more precisely defined because more information about the environment is available. For example, if the advertiser knew she was competing against M competitors, then B_{CP} would correspond to the highest order statistic out of the M competing bids. If a reserve price was in place, then B_{CP} would further be the maximum between this reserve price and the highest bid. This order statistic could be further characterized depending on what assumptions are made by the signals or information that competitors use to pick their bids, such as symmetry. If M is unknown but the advertiser knows the probability distribution governing M , then B_{CP} is the highest order statistic integrated against such distribution. The reason why we follow this reduced-form approach is twofold. The first is due to practical constraints: in settings such as ours, advertisers rarely have information about the number and identities of competitors they face, so conditioning on or incorporating it would be infeasible. The second is that we are focusing on the optimization problem faced by a single advertiser, who takes the actions of other agents as given. Since B_{CP} can incorporate both a varying number of competitors and a reserve price, it is a convenient modeling device to solve this problem.

order” uncertainty is similar to the one faced by bidders in typical auction models: the advertiser is uncertain about what her competitors bid, which is encapsulated by B_{CP} . She is also uncertain about her own valuation. As mentioned above, under this formulation the advertiser’s valuation corresponds to the treatment effect, $Y(1) - Y(0)$, which is never observed in practice.

The second, “higher-order” uncertainty is not present in standard auction models and is also the source of the inference goal. While in the majority of auction models a given bidder does not have complete knowledge of her valuation or of her competing bids, she does know the distributions from which these objects are drawn, which would correspond to the conditional joint distribution of $Y(1)$, $Y(0)$ and B_{CP} on x in our model. This is not the case here where the advertiser faces uncertainty regarding $F(\cdot, \cdot, \cdot | x)$.

To see why this bidding problem is non-trivial, notice that without access to data informative of this distribution, the advertiser would have to integrate over $F(\cdot, \cdot, \cdot | x)$ to construct the expected payoff from auction participation. The optimization problem she would then need to solve is:

$$\max_b \mathbb{E}_F \{ \mathbb{E}_{1,0,CP} [\pi(b, Y(1), Y(0), B_{CP}) | F, x] | x \}. \quad (5)$$

In equation (5), the inner expectation is taken with respect to $Y(1)$, $Y(0)$ and B_{CP} while keeping their joint distribution fixed, where $\pi(b, Y(1), Y(0), B_{CP})$ is given in equations (2) and (3). Thus, it reflects the aforementioned lower-order uncertainty. In turn, the outer expectation is taken with respect to $F(\cdot, \cdot, \cdot | x)$, which is reflected in the subscript “ F ” and on the conditioning on F in the inner expectation. At the most general level, the advertiser would consider all trivariate probability distribution functions whose support is the three-dimensional positive real line. As such, the optimization problem given in (5) is neither standard nor tractable, and the solution to this problem is in all likelihood highly sensitive to the beliefs over distributions the advertiser can have.

Without access to data, the advertiser would have no choice but to try to solve the optimization problem in (5). However, as we noted above, achieving the inference goal requires the collection of data, which can also be used to address her economic goal. Therefore, instead of tackling the optimization problem in (5), one strategy for optimizing bidding would be to use the data collected in the experiment to construct updated estimates of $\mathbb{E}_{1,0,CP} [\pi(b, Y(1), Y(0), B_{CP}) | F, x]$, and to perform bid optimization with respect to this estimate as the experiment progresses. This way, the data generated from the experiment

are used to address both the advertiser’s inference goal and to optimize expected profits in order to address her economic goal, so that both goals are pursued simultaneously.

Issue 4: Aligning the inference and economic goals

This leads to the final issue: how to balance the simultaneous pursuit of both the inference and economic goals in the experiment? This issue arises because typical strategies aimed at tackling one of the goals can possibly have negative impacts on accomplishing the other, suggesting an apparent tension between the two.

To see this, consider what would happen if the experiment focused only on the advertiser’s inference goal by randomizing bids without any concurrent bid optimization. We already alluded to the consequences of this for the advertiser’s economic goal in our previous discussion: pure bid randomization can hurt the advertiser’s economic goal by inducing costs from suboptimal bidding.

Consider now what would happen if the experiment focused only on the advertiser’s economic goal: learning $b^*(\cdot)$. Notice from equations (2), (3) and (4) that the payoffs from bidding b are stochastic from the advertiser’s perspective, and that the optimal bid, $b^*(\cdot)$, is the maximizer of the expected payoff from auction participation, $\bar{\pi}(b|x)$, which is an unknown objective function to the advertiser. Therefore, in this setup, pursuing the economic goal involves finding the best bid to play in an environment where payoffs are stochastic, and maximizing expected payoffs against a distribution which has to be learned by exploration. A MAB or RL approach is thus attractive in this situation because it can recover $b^*(\cdot)$ while minimizing the costs from suboptimal bidding, which pure randomization does not assess. By following this strategy the advertiser would adaptively collect data to learn a good bidding policy by continuously re-estimating and re-optimizing $\bar{\pi}(b|x)$.

In principle, these data could also be used to estimate $CATE(x)$ by running the regression in (1) for each x , for example. However, the adaptive nature of the data collection procedure induces autocorrelation in the data, which can impact asymptotic statistical and econometric properties of estimators. Moreover, even if all desired properties hold, underlying properties of the data and the algorithm used can affect the estimator adversely. To see this, consider the following example. For simplicity, assume once again that $P = 1$ so that x can be ignored and further assume that $\Pr(B_{CP} \leq b^*) \approx 1$. A good al-

gorithm would converge quickly, eventually yielding relatively few observations of $Y(0)$ compared to those of $Y(1)$, which would hinder the inference goal since typical estimators for *CATEs* based on such imbalanced data tend to be noisy.

This framing demonstrates that the inference and economic goals can possibly be in conflict. The advertiser’s goals are clearly related since they both depend on knowledge about the distribution $F(\cdot, \cdot, \cdot | \cdot)$. The challenge faced by the advertiser in accomplishing her goals is that this distribution is unknown to her. While there are known approaches to gather data and tackle each of the advertiser’s goals in isolation, it is unclear whether they can perform satisfactorily in achieving both goals concurrently. Addressing this is the core remaining piece for experimental design, which is discussed next.

3 Balancing the advertiser’s objectives

The strategy we adopt to balance the two goals leverages the microfoundations of the experiment by recognizing that the goals are linked to each other by the economic logic of optimal bidding. Because the bidding logic depends on the auction format, the extent to which the two goals are balanced will also differ by auction format. In particular, we will show that, in SPAs, the inference and economic goals can be perfectly aligned, while in FPAs, leveraging this linkage is still helpful, but the goals can only be imperfectly aligned.

To characterize our approach, we consider the limiting outcome of maximizing the true expected profit function with respect to bids when the joint distribution $F(\cdot, \cdot, \cdot | \cdot)$ is known to the advertiser. In what follows, we will use the expressions in equations (2) and (3) ignoring the second term, $Y(0)$, because it does not depend on the advertiser’s bid, b . This expression also has the benefit of directly connecting the potential outcomes to this auction-theoretic setting, with the treatment effect $Y(1) - Y(0)$ taking the role of the advertiser’s valuation.

We combine equations (2) and (4) to write the object she aims to learn in an SPA as the maximizer with respect to b of:

$$\begin{aligned} \bar{\pi}(b|x) &\equiv \mathbb{E} [\pi(b, Y(1), Y(0), B_{CP})|x] \\ &= \Pr(B_{CP} \leq b|x) \times \mathbb{E} \{ [Y(1) - Y(0)] - B_{CP} | B_{CP} \leq b; x \}, \end{aligned} \quad (6)$$

In an FPA, we combine equations (3) and (4) to write the object she aims to learn as the

maximizer with respect to b of:

$$\begin{aligned}\bar{\pi}(b|x) &\equiv \mathbb{E} [\pi(b, Y(1), Y(0), B_{CP})|x] \\ &= \Pr (B_{CP} \leq b|x) \times \mathbb{E} \{[Y(1) - Y(0)] - b|B_{CP} \leq b; x\}.\end{aligned}\quad (7)$$

Once again, the expectation in equations (6) and (7) is taken with respect to $Y(1)$, $Y(0)$ and B_{CP} , but the subscript “1,0,CP” is omitted to ease notation. The conditioning on the distribution F is also omitted since $\bar{\pi}(\cdot|x)$ is the true expected profit function, which therefore utilizes the true conditional distribution $F(\cdot, \cdot, \cdot|x)$ to compute the relevant expectations and probabilities. We denote the maximizers of these respective expressions by $b^*(x)$.

To ensure that $b^*(x)$ is well-defined, we maintain the following weak technical assumption on $F(\cdot, \cdot, \cdot|x)$.

Assumption 2. *Well behaved $F(\cdot, \cdot, \cdot|x)$*

- (i) *The joint distribution $F(\cdot, \cdot, \cdot|x)$ admits a continuous density, $f(\cdot, \cdot, \cdot|x)$, over \mathbb{R}_+^3 for all x .*
- (ii) *$\mathbb{E}[Y(1)|x] < \infty$, $\mathbb{E}[Y(0)|x] < \infty$, and $\mathbb{E}[B_{CP}|x] < \infty$ for all x .*
- (iii) *The density of B_{CP} given x , $f_{CP}(\cdot|x)$, is strictly positive in the interior of \mathbb{R}_+ for all x .*
- (iv) *$\frac{f_{CP}(b_{CP}|x)}{F_{CP}(b_{CP}|x)}$ is decreasing in b_{CP} for all x .*

Assumption 2 not only is mild but also relatively common in auction models and is made solely for tractability. Assumptions 2(i) and 2(ii) are minimal requirements. In causal inference terms, Assumption 2(iii) is equivalent to an overlap assumption that $0 < P(D = 1|x) < 1$, where $P(D = 1|x)$ is the propensity score. Letting $D \equiv \mathbb{1}\{B_{CP} \leq b\}$ as before, so that $P(D = 1|x) \equiv P(B_{CP} \leq b|x)$, Assumption 2(iii) implies $0 < P(B_{CP} \leq b|x) < 1$. In addition, Assumption 2(iii) could in principle be relaxed as we will mention below. Finally, Assumption 2(iv) is only required to determine $b^*(x)$ for FPAs. It states that the conditional distribution of B_{CP} on x has a decreasing reversed hazard rate. As argued by Block et al. (1998), this property holds for several distributions, including all decreasing hazard rate distributions and increasing hazard rate Weibull, gamma and lognormal distributions.

We now investigate the relationship between $b^*(\cdot)$ and $CATE(\cdot)$ under Assumptions 1 and 2. If the distribution $F(\cdot, \cdot, \cdot|x)$ was known, computing $CATE(\cdot)$ would be straightforward, as would be solving for $b^*(\cdot)$ by maximizing expression (6) or expression (7). The results below characterize this relationship first for SPAs and then for FPAs.

Proposition 1. *Optimal bid in SPAs*

If Assumptions 1 and 2 hold and the auction is an SPA, $b^(x) = \max\{0, CATE(x)\}$.*

Proof. To prove Proposition 1, we first rewrite equation (6):

$$\begin{aligned}
\bar{\pi}(b|x) &= \Pr \{B_{CP} \leq b|x\} \{ \mathbb{E} [Y(1) - Y(0)|B_{CP} \leq b; x] - \mathbb{E} [B_{CP}|B_{CP} \leq b; x] \} \\
&= \Pr \{B_{CP} \leq b|x\} \{ \mathbb{E} [Y(1) - Y(0)|x] - \mathbb{E} [B_{CP}|B_{CP} \leq b; x] \} \\
&= \Pr \{B_{CP} \leq b|x\} \{ CATE(x) - \mathbb{E} [B_{CP}|B_{CP} \leq b; x] \} \\
&= \int_0^b [CATE(x) - b_{CP}] f_{CP}(b_{CP}|x) db_{CP},
\end{aligned}$$

where the second equality follows from Assumption 1. Notice that $\bar{\pi}(b|x)$ becomes a bidder's expected payoff from a second-price sealed-bid auction in which the bidder's *private value* equals $CATE(x)$. Because the advertiser cannot submit negative bids, when $CATE(x) \leq 0$ the optimal bid is $b^*(x) = 0$ since the integrand is negative and $f_{CP}(\cdot|x) > 0$ in the interior of \mathbb{R}_+ due to Assumption 2(iii). Otherwise, notice that the integrand is non-negative as long as $b(x) \leq CATE(x)$, which implies the optimal bid cannot be greater than $CATE(x)$. Once again, due to Assumption 2(iii), $f_{CP}(\cdot|x) > 0$, so that $b^*(x) = CATE(x)$.¹¹ \square

We now characterize the analogous relationship for FPAs.

Proposition 2. *Optimal bid in FPAs*

If Assumptions 1 and 2 hold and the auction is an FPA, $b^*(x) = \max \{0, \chi^{-1} [CATE(x)]\}$, where $\chi(b) = b + \frac{F_{CP}(b|x)}{f_{CP}(b|x)}$.

Proof. To prove Proposition 2, we proceed as above, by rewriting equation (7):

$$\begin{aligned}
\bar{\pi}(b|x) &= \Pr \{B_{CP} \leq b|x\} \{ \mathbb{E} [Y(1) - Y(0) - b|B_{CP} \leq b; x] \} \\
&= \Pr \{B_{CP} \leq b|x\} \{ \mathbb{E} [Y(1) - Y(0)|x] - b \} \\
&= F_{CP}(b|x) \{ CATE(x) - b \},
\end{aligned}$$

where, once again, the second equality follows from Assumption 1. If $CATE(x) \leq 0$, it is straightforward to verify that $b^*(x) = 0$ since the advertiser cannot submit negative bids. Consider now the case where $CATE(x) > 0$. The first-order condition with respect to b is given by

$$CATE(x) = b + \frac{F_{CP}(b|x)}{f_{CP}(b|x)} \equiv \chi(b). \quad (8)$$

¹¹Notice that Assumption 2(iii) could be relaxed to assuming that $f_{CP}(\cdot|x)$ is strictly positive on neighborhoods around 0 and $CATE(x)$.

Assumption 2(iii) ensures that $\chi(0) = 0$. In addition, Assumption 2(iv) implies that the right-hand side of equation (8) is monotonically increasing in b . Hence, $\chi(\cdot)$ is invertible, which yields a unique solution. Denoting its inverse by $\chi^{-1}(\cdot)$, the optimal bid is therefore given by $b^*(x) = \max \{0, \chi^{-1} [CATE(x)]\}$. \square

Implications

Propositions 1 and 2 have three implications. First, they show how the optimal bids, $b^*(x)$, are related to $CATE(x)$. In an SPA, Proposition 1 shows that whenever displaying the ad is beneficial, that is, when $CATE(x) \geq 0$, one should bid the $CATE(x)$. In an FPA, Proposition 2 shows that whenever displaying the ad is beneficial, one should bid less than the $CATE(x)$.¹² In turn, when ad-exposure is detrimental, that is, when $CATE(x) < 0$, the propositions convert this qualitative fact into a clear economic policy as the advertiser would have no interest in displaying the ad in the first place, which can be guaranteed by a bid of zero. A consequence of these relationships is that the advertiser can follow a MAB or RL strategy in the experiment to learn $b^*(\cdot)$, and once they are learned, can obtain $CATE$ s by leveraging these relationships. This will form the basis of the algorithm we propose in the next section.

Second, one sees from the propositions that, for both SPAs and FPAs, the object of inference, $CATE(x)$, is a common component of the expected payoff associated with the bids one could consider for a given x . Thus, if one thinks of bids as “arms” in the sense of a bandit, leveraging the economic structure of the problem induces cross-arm learning within each context (i.e., pulling each arm contributes to learning $CATE$ s). This cross-arm learning, which is purely a consequence of maintaining a link to economic theory, helps the bandit-learner proposed in the next section infer the $CATE(x)$ more efficiently, which, in turn, also allows a more efficient learning of the optimal bidding policy for each context.

Third, the propositions show precisely how the inference and economic goals can be aligned in the experiment. For SPAs, Proposition 1 is powerful because it implies that whenever displaying the ad is beneficial, the advertiser’s economic and inference goals are *perfectly* aligned, as learning $b^*(x)$ and estimating $CATE(x)$ consist of the *same* task. Our proposed experimental design for SPAs will be to find the best bid for each x , $b^*(x)$, and set that to be the $CATE(x)$. This design will have the feature that achieving good

¹²Equation (8) shows that $b^*(x) \neq CATE(x)$ whenever $CATE(x) \neq 0$ because the right-hand side consists of a sum of two non-negative terms. In particular, if $CATE(x) > 0$ it follows that $b^*(x) < CATE(x)$ due to the typical bid shading in FPAs.

performance in learning the best bid does not come at the cost of reduced performance on measuring *CATE*s or vice versa.

For FPAs, Proposition 2 is also helpful because it implies that whenever displaying the ad is beneficial, the advertiser’s economic and inference goals reinforce each other, as learning $b^*(x)$ helps learning $CATE(x)$ in the experiment. Our proposed experimental design for FPAs will be to find the best bid for each x , $b^*(x)$, and set $\chi[b^*(x)]$ to be $CATE(x)$. However, because the inference and economic goals are not perfectly aligned, the design will have the feature that good performance in learning optimal bids may come at the cost of reduced performance in learning the *CATE*s or vice versa.

4 Accomplishing the advertiser’s objectives

We leverage Propositions 1 and 2 to develop an experimental design that concurrently accomplishes the advertiser’s goals. Our proposal comprises an adaptive design that learns $b^*(\cdot)$ over a sequence of display opportunities. We begin by stating the following assumption, which we maintain throughout the analysis.

Assumption 3. *Independent and identically distributed (i.i.d.) data*
 $\{Y_i(1), Y_i(0), B_{CP,i}\} \stackrel{iid}{\sim} F(\cdot, \cdot, \cdot | x_i)$ and $x_i \stackrel{iid}{\sim} F_x(\cdot)$.

Assumption 3 is a typical assumption made in stochastic bandit problems, that the randomness in payoffs is i.i.d. across occurrences of play. It imposes restrictions on the data generating process (DGP). For instance, if the same user appeared more than once and if his potential outcomes $Y(1)$ and $Y(0)$ were serially correlated, this condition would not hold. In turn, if competing bidders solved a dynamic problem because of longer-term dependencies, budget or impression constraints, B_{CP} could become serially correlated as a result, in which case this condition would also not hold.

Assumption 3 justifies casting the advertiser’s problem as a MAB. In particular, when Assumptions 2 and 3 hold, $b^*(x)$ is well-defined and common across auctions for every x for SPAs and FPAs. Under these assumptions, it is natural to represent the advertiser’s economic goal as a contextual MAB problem. In such setting, the advertiser considers a potentially context-specific set of $r_x = 1, \dots, R_x$ arms, each of which associated with a bid, b_{r_x} . The advertiser’s goal can be expressed as minimizing cumulative regret from potentially bidding suboptimally over a sequence of auctions while learning $b^*(\cdot)$.

Hence, as customary, we implicitly assume that for each x the grid contains the optimal bid, $b^*(x)$.

Algorithms used to solve MAB problems typically base the decision of which bid to play in round t , b_t , on a tradeoff between randomly picking a bid to obtain more information about its associated payoff (exploration) and the information gathered until then on the optimality of each bid (exploitation). The existing information at the beginning of round t is a function of all data collected until then, which we denote by W_{t-1} . Each observation i in these data, whose structure is displayed in Table 1 for SPAs and FPAs, is an ad display auction.

Table 1: Snapshot of data structure

i	b_i	x_i	D_i	Y_i	$Y_i(1)$	$Y_i(0)$	$B_{CP,i}$	
							SPA	FPA
1	b_1	x_1	1	y_1	y_1	—	$b_{CP,1}$	$\leq b_1$
2	b_2	x_2	0	y_2	—	y_2	$\geq b_2$	$\geq b_2$
\vdots	\vdots							

For the analysis of SPAs it will be useful to define the variable $\bar{B}_{CP} \equiv \min\{B_{CP}, b\}$. Stacking the data presented in Table 1 across auctions for each round τ , it follows that we can write $W_t = \{b_\tau, x_\tau, D_\tau, Y_\tau, \bar{B}_{CP,\tau}, \omega_\tau\}_{\tau=1}^t$ for SPAs and $W_t = \{b_\tau, x_\tau, D_\tau, Y_\tau, \omega_\tau\}_{\tau=1}^t$ for FPAs. In both cases, the ω s are seeds, independent from all other variables, required for randomization depending on which algorithm is used.

Notice that these data suffer from two issues. The first, common to both auction formats, is the fundamental missing data problem in causal inference referenced before (Holland, 1986): that $Y(1)$ and $Y(0)$ are not observed together at the same time. The second regards what we observe regarding B_{CP} and differs across the two auction formats. For SPAs, we have a censoring problem related to the competitive environment: for SPAs, B_{CP} is only observed when the advertiser wins the auction; otherwise, all she knows is that it was larger than the bid she submitted. Hence, the observed data have a similar structure to the one in the model defined by Amemyia (1984) as the Type 4 Tobit model. However, for FPAs this restriction is stronger: we never observe B_{CP} and only have either a lower or upper bound on it depending on whether the advertiser wins the auction, so

that the observed data has a Type 5 Tobit model structure.

There are two points of departure between this setup and a standard MAB problem. First, in the latter, each arm is associated with a different DGP, so it is commonly assumed that the reward draws are independent across arms. This is not true in our setting: given the economic structure of the problem, conditional on x the values of $\{Y(1), Y(0), B_{CP}\}$ are the same regardless of which arm is pulled, which creates correlation between rewards across arms. In particular, this is a nonlinear stochastic bandit problem as defined by [Bubeck and Cesa-Bianchi \(2012\)](#). Second, on pulling an arm the advertiser observes three different forms of feedback: an indicator for whether she wins the auction and obtains treatment (ad-exposure), the highest competing bid conditional on winning for SPAs or a bound on it conditional on losing, or a bound on the highest competing bid for FPAs, and the reward. This contrasts with the canonical case in which the reward forms the only source of feedback, and fits into the class of “complex online problems” studied by [Gopalan et al. \(2014\)](#).

5 Bidding Thompson Sampling (BITS) algorithm

We now propose a specific procedure to achieve the advertiser’s goals, which is a version of the TS algorithm. Since it aims to learn the advertiser’s optimal bid, we refer to it as Bidding Thompson Sampling (BITS).

5.1 General procedure

It is not our goal to solve for or implement the optimal learning policy that minimizes cumulative regret over a finite number of rounds of play. In fact, a general solution for MAB problems with correlated rewards across contexts and arms such as the one we consider is not yet known. What we require is an algorithm that performs “well” in terms of minimizing cumulative regret and that can easily accommodate and account for information shared across arms. Hence, we make use of the TS algorithm ([Thompson, 1933](#)), which is a Bayesian heuristic to solve MAB problems.¹³ TS typically starts by parametrizing the distribution of rewards associated with each arm. Since our problem departs from standard MAB problems in that the DGP behind each of the arms, that is, the distribution $F(\cdot, \cdot, \cdot | \cdot)$,

¹³See [Scott \(2015\)](#) for an application to computational advertising and [Russo et al. \(2018\)](#) for an overview.

is the same, we choose to parametrize it instead and denote our vector of parameters of interest by θ . Expected rewards depend on θ , so we will often write $\bar{\pi}(\cdot|\cdot, \theta)$. This is the same approach followed by [Gopalan et al. \(2014\)](#), who showed that, in a setting that has similarities to ours, the TS algorithm exhibits a logarithmic regret bound. This is the sense in which we consider it to perform “well.”

The algorithm runs while a criterion, c_t , is below a threshold, T . After round t , the prior over θ is updated by the likelihood of all data gathered by the end of round t , W_t . We denote the number of observations gathered on round t by n_t and the total number of observations gathered by the end of round t by $N_t = \sum_{\tau=1}^t n_\tau$. If $n_t = 1$ for all t the algorithm proceeds auction by auction. We present it in this way to accommodate batch updates. Given the posterior distribution of θ given W_t , we calculate

$$\psi_t(b_{r_x}|x) \equiv \Pr(\text{arm } r_x \text{ is the best arm} | W_t; x) \tag{9}$$

and update the criterion c_t . In round $t + 1$, arm r_x is pulled for each observation with context x with probability $\psi_t(b_{r_x}|x)$. The generic structure of the TS algorithm is outlined below.

Algorithm 1: Thompson Sampling
<ol style="list-style-type: none"> 1 Set priors, $\psi_0(\cdot \cdot)$, c_0 and T. <li style="padding-left: 2em;">while ($c_t < T$) do <li style="padding-left: 2em;">2 Pull arms according to $\psi_{t-1}(\cdot \cdot)$. <li style="padding-left: 2em;">3 Combine new data with previously obtained data in W_t. <li style="padding-left: 2em;">4 Update the posterior distribution of θ with W_t. <li style="padding-left: 2em;">5 Compute $\psi_t(\cdot \cdot)$, c_t and $b_t^*(\cdot)$. <li style="padding-left: 2em;">end

5.2 Parametrizing distribution of rewards

We now present the specific parametrization we use in our problem. Because of the structure of the data we described in Section 4 and because of the algorithm we use, our procedure requires reimplementing a Bayesian estimator to a Type 4 or Type 5 Tobit model on each round. Hence, the specific parametric structure we impose is chosen to make this estimator as simple as possible and, consequently, to speed up the implementation of the algorithm.

Let X_i be the following P -dimensional vector of mutually exclusive dummies:

$$X_i \equiv [\mathbb{1}\{x_i = x_1\}, \mathbb{1}\{x_i = x_2\}, \dots, \mathbb{1}\{x_i = x_P\}]'. \quad (10)$$

Notice that there is a one-to-one correspondence between the vector X_i and the variable x_i . Hence, we will use them interchangeably. We assume that

$$\begin{bmatrix} \log Y_i(1) \\ \log Y_i(0) \\ \log B_{CP,i} \end{bmatrix} \Big|_{X_i} \overset{i.i.d.}{\sim} N \left(\begin{bmatrix} X_i' \delta_1 \\ X_i' \delta_0 \\ X_i' \delta_{CP} \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_0 & 0 \\ \rho \sigma_1 \sigma_0 & \sigma_0^2 & 0 \\ 0 & 0 & \sigma_{CP}^2 \end{bmatrix} \right) \equiv N \left(\begin{bmatrix} \Delta' X_i \\ X_i' \delta_{CP} \end{bmatrix}, \begin{bmatrix} \Sigma & 0 \\ 0' & \sigma_{CP}^2 \end{bmatrix} \right), \quad (11)$$

where $\Delta \equiv [\delta_1, \delta_0]$. We collect the parameters in $\theta \equiv [\delta_1', \delta_0', \delta_{CP}', \sigma_1^2, \sigma_0^2, \sigma_{CP}^2, \rho]'$.

Notice that this parametrization directly imposes Assumption 1 and that it implies that $CATE(X_i) = \exp\{X_i' \delta_1 + 0.5\sigma_1^2\} - \exp\{X_i' \delta_0 + 0.5\sigma_0^2\}$. In addition, since the potential outcomes are never observed simultaneously, ρ is not point identified without further restrictions.¹⁴ Hence, since our interest is in $CATE(\cdot)$ and since it does not depend on ρ , we follow Chib and Hamilton (2000) and explicitly assume that $\rho = 0$. This assumption has the benefit of simplifying the algorithm we present. A more general version that allows for $\rho \neq 0$ is given in Appendix C. Finally, notice that (11) also implies that for SPAs the expected payoff is:

$$\begin{aligned} \bar{\pi}(b|X_i, \theta) &= \Phi \left(\frac{\log b - X_i' \delta_{CP}}{\sigma_{CP}} \right) \times CATE(X_i) \\ &\quad - \Phi \left(\frac{\log b - X_i' \delta_{CP}}{\sigma_{CP}} - \sigma_{CP} \right) \times \exp \left\{ X_i' \delta_{CP} + 0.5\sigma_{CP}^2 \right\}, \end{aligned} \quad (12)$$

and for FPAs the expected payoff is:

$$\bar{\pi}(b|X_i, \theta) = \Phi \left(\frac{\log b - X_i' \delta_{CP}}{\sigma_{CP}} \right) \times [CATE(X_i) - b], \quad (13)$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and where we omit the terms that do not depend on b for brevity.

¹⁴However, it is possible to exploit the positive semidefiniteness of Σ to partially identify ρ . See, for example, Vijverberg (1993) and Koop and Poirier (1997).

5.3 Choice of priors

We choose independent normal-gamma priors, which are conjugate to the normal likelihood induced by the DGP in (11). We choose these priors solely for convenience since they speed up the algorithm. For $k \in \{1, 0, CP\}$, we set:

$$\begin{aligned}\sigma_k^{-2} &\sim \Gamma(\alpha_k, \beta_k) \\ \delta_k | \sigma_k^2 &\sim N\left(\mu_{\delta_k}, \sigma_k^2 A_k^{-1}\right),\end{aligned}\tag{14}$$

where $\{\alpha_k, \beta_k\}_{k \in \{1, 0, CP\}}$ are non-negative scalars, $\{\mu_{\delta_k}\}_{k \in \{1, 0, CP\}}$ are P -dimensional vectors and $\{A_k\}_{k \in \{1, 0, CP\}}$ are P -by- P matrices. For the gamma distribution, the parametrization is such that if $G \sim \Gamma(\alpha, \beta)$, then $\mathbb{E}[G] = \alpha/\beta$. We discuss how to use historical data to choose the parameters of the prior distributions below.

5.4 Drawing from posterior: Gibbs sampling

Implementing the algorithm requires computing updated probabilities, $\psi_t(\cdot|\cdot)$, which cannot be done analytically because of the missingness and censoring in the feedback data. Nevertheless, it is still possible to exploit *conditional* conjugacy via data augmentation and use Gibbs sampling to obtain draws from the posterior distribution of θ given W_t . Using these draws we can then estimate $\psi_t(\cdot|\cdot)$ via Monte Carlo integration. We first describe the steps of this estimation procedure for SPAs, which combines the methods introduced by Chib (1992) and Koop and Poirier (1997) in a single Gibbs sampling algorithm with data augmentation, and then describe how it can be modified to accommodate FPAs. Bayesian data augmentation forms an elegant way to solve the censoring and missingness problems induced by the ad-auction. In each draw, we augment the Markov chain with the missing potential outcomes and competing bids, and then perform Bayesian inference on the required treatment effects conditioning on these augmented variables.

5.4.1 Data augmentation

The first step in our procedure is to draw the missing values from our data conditional on (W_t, θ) . We begin by drawing the missing values $\{\log B_{CP,i}\}_{i:D_i=0}$. Given (W_t, θ) and

under (11), it follows that:

$$\begin{aligned} \log B_{CP,i}^{miss} \Big| D_i = 0, \log Y_i, \log \bar{B}_{CP,i}, \log b_i, X_i, \theta &\stackrel{d}{=} \\ \log B_{CP,i}^{miss} \Big| D_i = 0, \log b_i, X_i, \delta_{CP}, \sigma_{CP}^2 &\sim TN \left(X_i' \delta_{CP}, \sigma_{CP}^2, \log b_i, +\infty \right), \end{aligned} \quad (15)$$

where $\stackrel{d}{=}$ means equality in distribution and $TN(\delta_*, \sigma_*^2, l, u)$ denotes the truncated normal distribution with mean δ_* , variance σ_*^2 , lower truncation at l and upper truncation at u .

Now we proceed to draw the missing values $\{\log Y_i(1)\}_{i:D_i=0}$ and $\{\log Y_i(0)\}_{i:D_i=1}$. Given (W_t, θ) and under (11), it follows that:

$$\begin{aligned} \log Y_i^{miss}(1) \Big| D_i = 0, \log Y_i, \log \bar{B}_{CP,i}, \log b_i, X_i, \theta &\stackrel{d}{=} \\ \log Y_i^{miss}(1) \Big| D_i = 0, X_i, \delta_1, \sigma_1^2 &\sim N \left(X_i' \delta_1, \sigma_1^2 \right) \end{aligned} \quad (16)$$

and,

$$\begin{aligned} \log Y_i^{miss}(0) \Big| D_i = 1, \log Y_i, \log \bar{B}_{CP,i}, \log b_i, X_i, \theta &\stackrel{d}{=} \\ \log Y_i^{miss}(0) \Big| D_i = 1, X_i, \delta_0, \sigma_0^2 &\sim N \left(X_i' \delta_0, \sigma_0^2 \right). \end{aligned} \quad (17)$$

Now, defining,

$$\delta_i^{miss} = D_i \times X_i' \delta_0 + (1 - D_i) \times X_i' \delta_1 \quad (18)$$

$$\sigma_i^{2,miss} = D_i \times \sigma_0^2 + (1 - D_i) \times \sigma_1^2, \quad (19)$$

we can combine (16) and (17) into:

$$\begin{aligned} \log Y_i^{miss} \Big| \log Y_i, D_i, \log \bar{B}_{CP,i}, \log b_i, X_i, \theta &\stackrel{d}{=} \\ \log Y_i^{miss} \Big| D_i, X_i, \delta_1, \delta_0, \sigma_1^2, \sigma_0^2 &\sim N \left(\delta_i^{miss}, \sigma_i^{2,miss} \right). \end{aligned} \quad (20)$$

5.4.2 Creating the “complete” data

Given a draw from the distributions given above, $\{\log Y_i^{miss}, \log B_{CP,i}^{miss}\}$, we can construct the “complete” data implied by that draw of the Markov chain:

$$\begin{aligned}\log \tilde{Y}_i(1) &= D_i \log Y_i + (1 - D_i) \log Y_i^{miss} \\ \log \tilde{Y}_i(0) &= D_i \log Y_i^{miss} + (1 - D_i) \log Y_i \\ \log \tilde{B}_{CP,i} &= D_i \log \tilde{B}_{CP,i} + (1 - D_i) \log B_{CP,i}^{miss}.\end{aligned}\tag{21}$$

5.4.3 Drawing from posterior distribution

The last step is to draw new parameters from their full conditional distributions. Collect the parameters of the priors in $\theta_{\text{prior}} \equiv \{\mu_{\delta_k}, A_k, \alpha_k, \beta_k\}_{k \in \{1,0,CP\}}$. For ease of notation, we stack all the “complete” data by the end of round t in the following N_t -by-1 vectors: $\log \tilde{Y}_t(1)$, $\log \tilde{Y}_t(0)$, $\log \tilde{B}_{CP,t}$, D_t and $\log b_t$. We also use the N_t -by- P matrix X_t , whose i^{th} row is the vector X'_i , and collect them all in the complete data set $\tilde{W}_t \equiv [\log \tilde{Y}_t(1), \log \tilde{Y}_t(0), \log \tilde{B}_{CP,t}, \log b_t, D_t, X_t]$. Finally, let the $(q-1)^{\text{th}}$ draw of the parameters be $\theta^{(q-1)} = [\delta_1^{(q-1)}, \delta_0^{(q-1)}, \delta_{CP}^{(q-1)}, \sigma_1^{2,(q-1)}, \sigma_0^{2,(q-1)}, \sigma_{CP}^{2,(q-1)}]'$. Given the structure of the model, it then follows that the full conditional distributions of the parameters simplify in the following way:

$$\begin{aligned}\sigma_{CP}^{2,(q)} \Big| \theta^{(q-1)}, \theta_{\text{prior}}, \tilde{W}_t &\stackrel{d}{=} \sigma_{CP}^{2,(q)} \Big| \log \tilde{B}_{CP,t}, X_t, \mu_{\delta_{CP}}, A_{CP}, \alpha_{CP}, \beta_{CP} \\ \sigma_1^{2,(q)} \Big| \theta^{(q-1)}, \theta_{\text{prior}}, \tilde{W}_t &\stackrel{d}{=} \sigma_1^{2,(q)} \Big| \log \tilde{Y}_t(1), X_t, \mu_{\delta_1}, A_1, \alpha_1, \beta_1 \\ \sigma_0^{2,(q)} \Big| \theta^{(q-1)}, \theta_{\text{prior}}, \tilde{W}_t &\stackrel{d}{=} \sigma_0^{2,(q)} \Big| \log \tilde{Y}_t(0), X_t, \mu_{\delta_0}, A_0, \alpha_0, \beta_0\end{aligned}\tag{22}$$

and, letting $\sigma^{2,(q)} \equiv [\sigma_1^{2,(q)}, \sigma_0^{2,(q)}, \sigma_{CP}^{2,(q)}]'$,

$$\begin{aligned}\delta_{CP}^{(q)} \Big| \sigma^{2,(q)}, \delta_1^{(q-1)}, \delta_0^{(q-1)}, \delta_{CP}^{(q-1)}, \theta_{\text{prior}}, \tilde{W}_t &\stackrel{d}{=} \delta_{CP}^{(q)} \Big| \sigma_{CP}^{2,(q)}, \log \tilde{B}_{CP,t}, X_t, \mu_{\delta_{CP}}, A_{CP} \\ \delta_1^{(q)} \Big| \sigma^{2,(q)}, \delta_1^{(q-1)}, \delta_0^{(q-1)}, \delta_{CP}^{(q-1)}, \theta_{\text{prior}}, \tilde{W}_t &\stackrel{d}{=} \delta_1^{(q)} \Big| \sigma_1^{2,(q)}, \log \tilde{Y}_t(1), X_t, \mu_{\delta_1}, A_1 \\ \delta_0^{(q)} \Big| \sigma^{2,(q)}, \delta_1^{(q-1)}, \delta_0^{(q-1)}, \delta_{CP}^{(q-1)}, \theta_{\text{prior}}, \tilde{W}_t &\stackrel{d}{=} \delta_0^{(q)} \Big| \sigma_0^{2,(q)}, \log \tilde{Y}_t(0), X_t, \mu_{\delta_0}, A_0.\end{aligned}\tag{23}$$

The specific forms of these full conditional distributions are presented in Appendix A.

5.4.4 Summary

The full Gibbs sampling procedure is summarized below. If one wishes to allow for $\rho \neq 0$ the procedure has to be adjusted. We present this more general algorithm in Appendix C.

Algorithm 2: Gibbs sampling	
1	Set $\theta^{(0)}$ and θ_{prior} .
for $(q = 1, \dots, Q)$ do	
2	Draw $\left\{ \log Y_i^{\text{miss},(q)}(1), \log Y_i^{\text{miss},(q)}(0), \log B_{CP,i}^{\text{miss},(q)} \right\}_{i=1}^{N_t}$ according to equations (15)–(20).
3	Construct $\left\{ \log \tilde{Y}_i^{(q)}(1), \log \tilde{Y}_i^{(q)}(0), \log \tilde{B}_{CP,i}^{(q)} \right\}_{i=1}^{N_t}$ according to equation (21).
4	Draw $\theta^{(q)}$ according to equations (22)–(23).
end	

5.5 Adaptation to FPAs

The procedure described above for SPAs can be used with minor adjustments to handle FPAs. Accommodating FPAs involves two substantive changes. The first is that we use the expected profit function for FPAs in (13) rather than the one for SPAs in equation (12). The second change is a function of the different data the advertiser would have access to in an FPA. As noted before, under an FPA, B_{CP} is always missing, which necessitates the normalization $\sigma_{CP}^2 = 1$. However, under Assumption 1 B_{CP} is conditionally independent from $Y(1)$ and $Y(0)$ given x , and therefore the augmentation step for B_{CP} would not depend on observed outcomes given x . Therefore, instead of embedding the Bayesian approach to the Tobit model by Chib (1992) in the BITS algorithm, we use the Bayesian approach to a Probit model introduced by Albert and Chib (1993).

More concretely, in addition to drawing the missing values $\{\log B_{CP,i}\}_{i:D_i=0}$ according to (15) with $\sigma_{CP}^2 = 1$, we also draw the missing values $\{\log B_{CP,i}\}_{i:D_i=1}$ according to,

$$\begin{aligned} \log B_{CP,i}^{\text{miss}} \Big| D_i = 1, \log Y_i, \log \bar{B}_{CP,i}, \log b_i, X_i, \theta &\stackrel{d}{=} \\ \log B_{CP,i}^{\text{miss}} \Big| D_i = 1, \log b_i, X_i, \delta_{CP} &\sim TN(X_i' \delta_{CP}, 1, -\infty, \log b_i). \end{aligned} \quad (24)$$

Hence, instead of creating the variable $\log \tilde{B}_{CP,i}$ according to equation (21), we equal it to $\log B_{CP,i}^{\text{miss}}$ for all i . Other than these two modifications, the rest of the procedure remains

unchanged.

5.6 Estimating optimality probability of each arm and implied *CATEs*

Once the draws from the posterior $\theta^{(q)}$ are obtained as above in round t , for each draw $\theta^{(q)}$, context x and arm b_{r_x} , we can compute $\bar{\pi}(b_{r_x}|x, \theta^{(q)})$ via equation (12) for SPAs or equation (13) for FPAs. The probability that an arm b_{r_x} is best for context x as of round t is estimated by averaging over the Q draws:

$$\hat{\psi}_t(b_{r_x}|x) = \frac{1}{Q} \sum_{q=1}^Q \mathbb{1} \left\{ \bar{\pi}(b_{r_x}|x, \theta^{(q)}) > \bar{\pi}(b_{r'_x}|x, \theta^{(q)}) \text{ for all } r'_x \neq r_x \right\}, \quad (25)$$

which is used then for allocation of traffic as outlined in Algorithm 1.

Given $\hat{\psi}_t(b_{r_x}|x)$, and leveraging Proposition 1, the procedure implies that, for an SPA, $CATE(x)$ is

$$b_{r_x} \text{ with probability } \hat{\psi}_t(b_{r_x}|x), \quad (26)$$

that is, we read off $CATE(x)$ as the label associated with what the procedure implies is the best bid-arm for that x . This leads to the following estimator of $CATE(x)$ in an SPA,

$$\widehat{CATE}_t(x) = \sum_{r_x=1}^{R_x} \hat{\psi}_t(b_{r_x}|x) \times b_{r_x}. \quad (27)$$

For an FPA, we can estimate the bid-adjustment in equation (8) for each bid-arm b_{r_x} by averaging the inverse of the reversed hazard rate of B_{CP} at that x over the $q = 1, \dots, Q$ draws of $\theta^{(q)}$. Leveraging Proposition 2, the procedure implies that in an FPA,

$$\hat{\chi}_t(b_{r_x}) \equiv b_{r_x} + \frac{1}{Q} \sum_{q=1}^Q \frac{F_{CP}(b_{r_x}|x; \theta^{(q)})}{f_{CP}(b_{r_x}|x; \theta^{(q)})} \text{ with probability } \hat{\psi}_t(b_{r_x}|x), \quad (28)$$

that is, we read off $CATE(x)$ as the label associated with what the procedure implies is the best bid-arm for that x plus the bid-adjustment term for that bid-arm. This leads to

the following estimator of $CATE(x)$ for FPAs

$$\widehat{CATE}_t(x) = \sum_{r_x=1}^{R_x} \hat{\psi}_t(b_{r_x}|x) \times \hat{\chi}_t(b_{r_x}). \quad (29)$$

In equations (28) and (29), $F_{CP}(\cdot|\cdot)$ and $f_{CP}(\cdot|\cdot)$ are, respectively, the conditional *CDF* and *PDF* of B_{CP} on x evaluated at the draws $\theta^{(q)}$, so that the bid-adjustment is the inverse of the reversed hazard rate of B_{CP} averaged over the Q draws from the updated posterior distribution. By assumption 2(iv), the inverse of the reversed hazard rate is monotonically increasing in the bid for each draw. The average over draws is therefore a weighted sum of monotonically increasing functions, which is also monotonically increasing. Therefore, we are guaranteed that the estimates $\hat{\chi}_t(b_{r_x})$ in equation (28) are monotonically increasing in b_{r_x} , so a well-defined mapping between the bid-value and $\widehat{CATE}_t(x)$ exists. For the specific parametrization chosen in equation (11), $F_{CP}(\cdot|\cdot)$ and $f_{CP}(\cdot|\cdot)$ correspond to a lognormal distribution with parameters $X'_i\delta_{CP}$ and σ_{CP}^2 , which makes it easy to compute the right-hand side of equation (29) at the end of each round.

5.7 Stopping the experiment

The last piece required to complete the discussion of the experiment is a decision criterion for when to stop the experiment. Before outlining our suggested criterion, we note at the outset that an advantage of the adaptive cost control in the experiment is that the adverse profit consequences of allowing it to run without stopping are low: as the experiment proceeds, more traffic is assigned to the bid-arm that provides highest expected payoff from auction participation to the advertiser, so continuing to let the experiment run protects the advertiser's interests from continued auction participation. In this situation, the proposed experimental design can be viewed simply as an explore-exploit scheme for optimal bidding that also has an auxiliary benefit of delivering estimates of *CATEs* to the advertiser. In other use-cases, the proposed experimental design forms the basis of an explicit test the advertiser runs in order to measure *CATEs*, for which developing a principled approach to stoppage is useful (see Scott, 2015 and Geng et al., 2020 for examples). With this latter perspective in mind, this section discusses stopping rules that terminate the experiment when the inference goal is achieved with reasonable precision.

The simplest stopping rule is to specify the total number of rounds the algorithm has

to run through, in which case $c_t = t$ and T is some exogenous threshold, which could map to a notion of time or budget available to run the experiment. A more nuanced stopping rule uses the data collected through the algorithm to inform the decision of when to stop the experiment. The stopping rule discussed below follows this approach. We motivate it first in a non-contextual setting to provide intuition, and then generalize it to the more complex contextual case.

5.7.1 Non-contextual case

Consider first a non-contextual MAB problem, that is, x can take only one value. Thus, we omit x for the remaining of this section to ease notation. The algorithm aims to identify the best bid-arm while minimizing the costs of experimentation. Therefore, we can leverage a stopping rule based on the confidence with which the optimal arm was found. More precisely, suppose we set $T = 0.95$ and:

$$c_t = \max_r \hat{\psi}_t(b_r). \quad (30)$$

We can interpret this as a decision to stop when the posterior distribution of θ given W_t leads us to believe that the arm with current highest probability of being the optimal arm is the true best arm with at least 95% probability. By virtue of equation (26) for SPAs, $\hat{\psi}_t(b_r)$ also represents the probability, based on the current posterior, with which we believe that bid-arm value b_r is the true *ATE*. Analogously, by virtue of equation (28) for FPAs, $\hat{\psi}_t(b_r)$ represents the probability, based on the current posterior, with which we believe the adjusted bid-arm value $\hat{\chi}_t(b_r)$ is the true *ATE*. Thus, we can also interpret the stopping rule in equation (30) as a decision to stop when the posterior distribution of θ given W_t leads us to believe that the *ATE* value associated with the arm with current highest probability of being the true best arm is the true *ATE* with at least 95% probability.

This stopping rule has an attractive feature in that it has a well-defined interpretation in terms of Bayes factors, which are often used for Bayesian hypothesis testing. Let $\zeta_t(b_r)$

be the posterior odds ratio of arm r being the optimal arm by the end of round t . Then,

$$\begin{aligned}\zeta_t(b_r) &= \frac{\Pr_t(b_r \text{ is the optimal bid})}{\Pr_t(b_r \text{ is not the optimal bid})} \\ &= \frac{\Pr_t(b_r \text{ is the optimal bid})}{1 - \Pr_t(b_r \text{ is the optimal bid})} \\ &= \frac{\psi_t(b_r)}{1 - \psi_t(b_r)}.\end{aligned}\tag{31}$$

Thus, c_t can alternatively be constructed as $\max_r \hat{\zeta}_t(b_r)$, with corresponding threshold $T = 19$, so that stopping is based off a threshold on the implied Bayes factor.¹⁵

5.7.2 Contextual case

The contextual case is more complex because now there is not a single best arm, but P best arms. Thus, a natural but conservative approach would be to require 95% posterior probability over a list of P arms as being the optimal ones. In this case, the threshold rule can be expressed by:

$$c_t = \min_{x \in \mathbb{X}} \max_{r_x} \hat{\psi}_t(b_{r_x}|x),\tag{32}$$

while maintaining the requirement that $c_t > T = 0.95$. Consequently, upon stoppage there would be at least 95% posterior probability on the $CATE(x)$ value associated with the bid-arms for each x .

In some scenarios, the advertiser's inference objective may be to estimate and perform inference on the unconditional ATE . Under these circumstances, the stopping rule above is likely to be too stringent. The ATE is the weighted average of $CATE(x)$ over the distribution of x . To achieve the goal of learning the weighted average of $CATEs$ with a given level of precision, it is not necessary to learn every $CATE(x)$ with the same level of precision. We now present a slightly less demanding stoppage criterion that reflects this.

Recall that the context x takes P values, indexed by $p = 1, \dots, P$, and that for each value x_p we consider R_{x_p} different bid-arms. Consequently, considering only the val-

¹⁵It is important to emphasize that following this stopping rule is not equivalent to conducting a sequential Bayesian hypothesis test. Such procedure would require us to establish a null hypothesis that one specific arm was the best and base the decision to stop solely on this arm's Bayes factor or posterior odds ratio. Instead, here we remain agnostic as to which arm is the best, and base our decision to stop on which arm has strongest evidence in its favor.

ues from the grids the ATE can take at most $R \equiv \prod_{p=1}^P R_{x_p}$ values because $ATE = \sum_{p=1}^P F_x(x_p) \times CATE(x_p)$ and because each $CATE(x_p)$ can take R_{x_p} values. Consider the R values that ATE can take and select a grid composed of the Y unique values among these R , which we denote o_v for $v = 1, \dots, Y$. An alternative criterion is to stop the experiment when the posterior at the end of a round implies with at least 95% probability that the ATE is equal to one of the o_v values in this grid.

To make this criterion precise, consider, for each o_v , a sequence s over the P contexts such that the implied estimate of ATE from these values equals o_v . In other words, s is a sequence of values $\{b_{r_{x_p}}^s\}_{p=1, \dots, P}$, each taken from one of the R_{x_p} values in the grid of each context x_p , such that $o_v = \sum_{p=1}^P F_x(x_p) \times b_{r_{x_p}}^s$ for SPAs and $o_v = \sum_{p=1}^P F_x(x_p) \times \hat{\chi}_t(b_{r_{x_p}}^s)$ for FPAs. Let S_v be the total number of such sequences. Then the alternative stopping criterion is given by:

$$c_t = \max_{v \in \{1, \dots, Y\}} \left\{ \sum_{s=1}^{S_v} \sum_{p=1}^P F_x(x_p) \times \hat{\psi}_t(b_{r_{x_p}}^s | x_p) \right\} \quad (33)$$

and we stop when $c_t > T = 0.95$. Notice that while this decision rule depends on the confidence with which we believe to have found the true ATE as implied by the posterior distribution of θ given W_t , traffic is still allocated to each arm according to (25). Hence, the decision to stop the experiment is aligned with the advertiser's inference goal, while the way it performs randomization is aligned with her economic goal.

Notice that this stopping criterion presupposes that the distribution from which contexts are drawn, $F_x(\cdot)$, is known to the researcher. When this is the case, the grid of values $\{o_v\}_{v=1, \dots, Y}$ is fixed for SPAs, but it changes for FPAs because the values $\hat{\chi}_t(b_{r_x})$ change over the rounds. When this is not the case, one could replace it with empirical frequencies estimated using data collected via the algorithm, in which case the grids will vary across rounds for both auction formats. While we expect the stopping rule given in (33) to shorten the duration of the experiment when compared to the one given in (32), we found in simulations that the difference between these two rules is minimal.¹⁶

¹⁶It is important to mention that the statistical implications of data-driven stoppage in sequential experiments is still being debated in the literature. Even though data-based stopping rules are known to interfere with frequentist inference, which motivated the development of new methods to explicitly account for this interference both for non-adaptive (Johari et al., 2019) and adaptive (Yang et al., 2017; Jamieson and Jain, 2018; Ju et al., 2019) data collection procedures, Bayesian inference has historically been viewed as *immune* to optional stopping rules (Lindley, 1957; Edwards et al., 1963; Savage, 1972; Good, 1991). Nevertheless, a recent debate has emerged concerning the effects of optional stopping on frequentist properties and interpretation of Bayes estimators (Yu et al., 2014; Sanborn and Hills, 2014; Rouder, 2014; Dienes, 2016; Deng

5.8 Practical considerations and extensions

We conclude the experimental design discussing some practical considerations that arise in implementation and ways in which the design can be extended to accommodate variations in the experimentation environment and advertiser goals.

5.8.1 Regret minimization versus best-arm identification

We implement BITS under a regret minimization framework based on the viewpoint that the advertiser seeks to maximize her payoffs from auction participation during the experiment. We could alternatively cast the problem as one of pure best-arm identification as studied by [Bubeck et al. \(2009\)](#), for example. In the best-arm identification formulation, the problem is cast in terms of pure exploration, so the role of adaptive experimentation is to obtain information efficiently before committing to a final decision involving the best-arm identified with that information.¹⁷ To leverage Propositions 1 and 2, what we need is a MAB framework to recover the arm with highest expected reward, so the core idea behind our proposed approach ports in a straightforward way to this alternative formulation of the experimental objective.

5.8.2 Parametric assumptions and alternative algorithms

More flexible parametric specifications can be used instead of (11) and (14).¹⁸ A more flexible distribution may be especially desirable for FPAs due to the explicit dependence of the *CATEs* on the distribution of B_{CP} via equation (8). The cost of more flexibility is that the researcher has to employ more complex MCMC methods, which may be slower than the Gibbs sampling algorithm presented above. If the updating becomes slow, the induced latency may form an impediment to implementation in practical ad-tech settings. This is because conditional conjugacy is likely to fail under alternative parametric specifications. Furthermore, any algorithm with convergence guarantees could in theory be used instead

et al., 2016; Schönbrodt et al., 2017; Wagenmakers et al., 2019; Tendeiro et al., 2019; de Heide and Grünwald, 2020; Rouder and Haaf, 2020; Hendriksen et al., 2021). We do not attempt to resolve this debate in this paper. In several simulations we ran, we found that the practical impact of optional stopping was minimal in our setting.

¹⁷Russo (2020) provides an adaptation of TS to best-arm identification. For an example of a study that adopts this approach to identify an optimal treatment assignment policy, see [Kasy and Sautmann \(2021\)](#).

¹⁸For a discussion of how more flexible parameterizations can be used for Bayesian estimation of treatment effects, see, for example, [Heckman et al. \(2014\)](#).

of the proposed BITS algorithm if the practitioner is not comfortable with making specific distributional assumptions, which may not be required for these other methods.

5.8.3 Obtaining draws from posterior distribution

The method we presented requires the researcher to employ Gibbs sampling in each round, which becomes slower as the number of rounds increases. This is because it requires the posterior to be updated conditioning on the data collected from the beginning of the experiment. The size of these data grows as we increase the rounds. If the procedure becomes too slow, Sequential Monte Carlo (SMC) or particle filtering methods could instead be used to speed up the sampling.¹⁹ In SMC, one updates conditioning on the data collected in the most recent round, rather than from the beginning of the experiment. SMC is also attractive if the practitioner chooses to use more flexible parametric specifications.

5.8.4 Using additional data on competing bids

It is straightforward to adapt the BITS procedure if different types of auction data are made available. For SPAs, we have assumed the advertiser only observes B_{CP} when she effectively has to pay this amount; otherwise, all she knows is that it is bounded below by b . On the other hand, for FPAs the advertiser never observes B_{CP} . These assumptions characterize the most stringent data limitations in these auction environments.

In some scenarios, the data limitations may be less stringent. For instance, if the transaction price from the auction is made public by the AdX to auction participants, the advertiser can possibly obtain a more precise lower bound on B_{CP} whenever the transaction price is larger than b in SPAs. This yields a new definition of \bar{B}_{CP} . Accommodating this does not require any modification to the BITS procedure, but does require us to assume that the transaction price is also independent from the potential outcomes conditional on x . For FPAs, disclosure of the transaction price would imply that the advertiser would observe B_{CP} whenever she lost the auction, which would give rise to an analogous procedure as the one adopted above for SPAs.

Finally, in the event B_{CP} itself is made public by the AdX, the algorithm simplifies

¹⁹For an application of SMC methods to MAB problems, see [Cherkassky and Bornn \(2013\)](#).

further since the censoring problem vanishes. Hence, the practitioner can update the posterior distribution over the parameters δ_{CP} and σ_{CP}^2 analytically, without the need to use the Gibbs sampling procedures because of the exact conjugacy implied by (11) and (14).

5.8.5 Budget constraints

The current formulation does not explicitly incorporate budget constraints, which can be present and relevant in practice for running experiments. Nevertheless, budget constraints are implicitly considered in the design by the use of a stopping rule for the experiment as discussed in Section 5.7. While there are more general RL tools to perform bid optimization in the presence of budget constraints, such as Cai et al. (2017), these methods do not address causal inference tasks. Part of the complication is that the optimal bidding policy becomes a function both of x and of the remaining budget, and linking it to CATEs becomes non-trivial. Formal incorporation of budget constraints would therefore need extending the algorithm beyond its current scope, and is left as a topic for future research.

5.8.6 Choosing priors to address the cold start problem

While priors always play an important role in Bayesian inference, they can become even more important in the context of experimentation as a way to deal with a “cold start” problem. Well-informed priors might situate the algorithm at a good starting point, shortening the duration of the experiment and, consequently, decreasing its costs. On the other hand, poorly specified priors might have the opposite effect and become inferior even to diffuse, non-informative priors. We discuss briefly how historical data that may be available to the advertiser can be used to inform the choice of the prior parameters. We assume the experimenter has access to a historical data set $W_n = \{b_i, X_i, D_i, Y_i, \bar{B}_{CP,i}\}_{i=1}^n$ for SPAs and $W_n = \{b_i, X_i, D_i, Y_i\}_{i=1}^n$ for FPAs.

To leverage the historical data, we can equate the means and variances of the prior distributions to the approximate means and variances of estimators of δ_1 , δ_0 , δ_{CP} , σ_1^2 , σ_0^2 and σ_{CP}^2 , where the last estimator is only required for SPAs. It is straightforward to develop a maximum likelihood estimator (MLE) for the prior parameters, which, in this case,

are \sqrt{n} -consistent and asymptotically normal. The MLE sets, for $k \in \{1, 0, CP\}$,

$$\begin{aligned}
\frac{\alpha_k}{\beta_k} &= \hat{\sigma}_k^{-2} \\
\frac{\alpha_k}{\beta_k^2} &= \frac{1}{n} A\hat{var} \left[\sqrt{n} \left(\hat{\sigma}_k^{-2} - \sigma_k^{-2} \right) \right] = \frac{\hat{\sigma}_k^8}{n} A\hat{var} \left[\sqrt{n} \left(\hat{\sigma}_k^2 - \sigma_k^2 \right) \right] \\
\mu_{\delta_k} &= \hat{\delta}_k \\
A_k &= n\hat{\sigma}_k^2 \left\{ A\hat{var} \left[\sqrt{n} \left(\hat{\delta}_k - \delta_k \right) \right] \right\}^{-1}.
\end{aligned} \tag{34}$$

Notice that for the parameters of the Gamma distributions this is equivalent to:

$$\begin{aligned}
\alpha_k &= n\hat{\sigma}_k^{-4} \left\{ A\hat{var} \left[\sqrt{n} \left(\hat{\sigma}_k^{-2} - \sigma_k^{-2} \right) \right] \right\}^{-1} = n\hat{\sigma}_k^4 \left\{ A\hat{var} \left[\sqrt{n} \left(\hat{\sigma}_k^2 - \sigma_k^2 \right) \right] \right\}^{-1} \\
\beta_k &= n\hat{\sigma}_k^{-2} \left\{ A\hat{var} \left[\sqrt{n} \left(\hat{\sigma}_k^{-2} - \sigma_k^{-2} \right) \right] \right\}^{-1} = n\hat{\sigma}_k^6 \left\{ A\hat{var} \left[\sqrt{n} \left(\hat{\sigma}_k^2 - \sigma_k^2 \right) \right] \right\}^{-1}.
\end{aligned} \tag{35}$$

For the purposes of this estimation we assume that $D_i \perp\!\!\!\perp Y_i(1), Y_i(0) | X_i$. Since we maintain Assumption 1, the only potential source of dependence between D_i and the potential outcomes is b_i . The validity of this assumption is not a concern when the algorithm is implemented for actual experimentation since the bids b_i are under the control of the experimenter. If the historical data come from an experiment in which bids were randomized, this condition is also satisfied. If the assumption is violated in the historical data, the resulting priors have some bias; however, the experimentation algorithm will still consistently recover the true *CATEs*. Further details about the estimators leveraging historical data and their computation are relegated to Appendix B.

6 Simulation evaluations of proposed approach

This section provides simulation results documenting the performance of the proposed approach. We first discuss the results for SPAs, and then for FPAs. Within each, we first show that the algorithm works as intended in a non-contextual setup, and then demonstrate performance under the more involved contextual setup. Performance on the economic goal is assessed using cumulative pseudo-regret as the performance metric. Performance on the inference goal is assessed using the Mean Squared Error (*MSE*) as the performance metric. The algorithm is compared to several alternative experimental designs for each scenario. Overall, the results show that the proposed approach performs

well and is superior to the other considered alternatives.

6.1 SPAs

6.1.1 Non-contextual case

We begin by considering the non-contextual case, in which $P = 1$ so we can ignore x . Our goal is to show that the BITS algorithm works in practice in recovering the true ATE . Additionally, we assess its performance in achieving both the inference and economic goals relative to alternative methods.

DGP, simulation details and grids of bids

For these simulations, we assume that:

$$\begin{bmatrix} \log Y_i(1) \\ \log Y_i(0) \\ \log B_{CP,i} \end{bmatrix} \stackrel{iid}{\sim} N \left(\begin{bmatrix} 0.809 \\ 0.22 \\ 0.4 \end{bmatrix}, \begin{bmatrix} 0.49 & 0 & 0 \\ 0 & 0.81 & 0 \\ 0 & 0 & 0.25 \end{bmatrix} \right), \quad (36)$$

where the value of δ_1 is chosen so that $ATE = 1$. These values are chosen for the sake of illustration.

We simulate 1,000 different epochs. Each epoch has $T = 100$ rounds, each of which with 50 new observations, so that $n_t = 50$ for $1 \leq t \leq 100$. We keep these values fixed and change the grid of bids we use to assess how the performance of our algorithm changes. In particular, we consider three different grids:

1. 3 arms: $b \in \{0.6, 1.0, 1.5\}$
2. 5 arms: $b \in \{0.6, 0.8, 1.0, 1.25, 1.5\}$
3. 10 arms: $b \in \{0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5\}$

Since the width of bids and the number of observations are kept fixed while the number of arms increases, we expect the performance of BITS to deteriorate with finer grids.

Approaches under consideration

We consider the following experimental approaches to estimate the *ATE*:

1. *A/B test (A/B)*: to randomize treatment, the A/B test simply randomizes with equal probability the bid placed from the same grid of bids used by *BITS*. In other words, the A/B test is a design that implements equal allocation of experimental traffic to various arms non-adaptively. Once the data are collected, the *ATE* is estimated by running a regression of Y on D using the experimental sample. Under pure bid randomization, the estimated slope coefficient from this regression is consistent for the *ATE* due to Assumption 1.
2. *Explore-then-commit (ETC)*: this approach proceeds as an A/B test for the first half of the experiment, that is, for the first 50 rounds; it then collects all data from these 50 rounds and runs a regression of Y on D as above to estimate the *ATE*. In the second half of the experiment, this approach places this estimate as the bid for arriving impressions, thus committing to what was learned in the first part of the experiment to exploitation of that information in the latter part.
3. *Off-the-shelf Thompson Sampling (TS)*: this is the basic implementation for the TS algorithm to this setting. It assumes that the rewards obtained from each arm are independent draws from arm-specific normal distributions. It then updates the mean and variance of each such distribution in the usual way employing only observations obtained from that arm. We use the conjugate normal-gamma priors with uninformative parameters for all arms. To estimate the probabilities that each arm is optimal, we take 1,000 draws from each normal distribution and use Monte Carlo integration.
4. *Bidding Thompson Sampling (BITS)*: to implement our algorithm, we make use of non-informative priors for all parameters, that is, we set $\alpha_k = \beta_k = \mu_{\delta_k} = A_k = 0$ for $k \in \{1, 0, CP\}$. Every time we run Gibbs sampling, we set the initial values to $\delta_1^{(0)} = \delta_0^{(0)} = \delta_{CP}^{(0)} = 0$ and $\sigma_1^{2,(0)} = \sigma_0^{2,(0)} = \sigma_{CP}^{2,(0)} = 1$. We take $Q = 1,000$ draws, drop the first half and use only the multiples of 10 (510, 520, etc.) to estimate the optimality probabilities to mitigate possible dependence between the draws.

Criteria of comparison

To compare the performance of the aforementioned methods we consider two metrics. The first, *cumulative pseudo-regret*, represents the economic goal. Following [Bubeck and Cesa-Bianchi \(2012\)](#), the pseudo-regret from method ι on round t is given by:

$$\text{Pseudo-regret}_{\iota,t} = \bar{\pi}(b^*) - \sum_{r=1}^R \Pr_{\iota,t}(b_r \text{ is pulled}) \times \bar{\pi}(b_r), \quad (37)$$

and therefore cumulative pseudo-regret at round t is given by:

$$\begin{aligned} \text{Cumulative pseudo-regret}_{\iota,t} &= \sum_{\tau=1}^t \text{Pseudo-regret}_{\iota,\tau} \\ &= t \times \bar{\pi}(b^*) - \sum_{\tau=1}^t \sum_{r=1}^R \Pr_{\iota,\tau}(b_r \text{ is pulled}) \times \bar{\pi}(b_r). \end{aligned} \quad (38)$$

Notice that for the A/B test it follows that $\Pr_{\iota,t}(b_r \text{ is pulled}) = 1/R$ for all t , and therefore it exhibits constant pseudo-regret and linear cumulative pseudo-regret. Moreover, we note that for ETC the bid placed when $t > 50$ does not belong to the original grid, but is rather the OLS estimate from the regression of Y on D using the data gathered on the first 50 rounds. Finally, we note that for both TS and BITS the probability $\Pr_{\iota,t}(\cdot)$ is not known exactly; instead, it is estimated via Monte Carlo integration.

The second metric, *MSE*, represents the inference goal. We define the estimated *MSE* of the *ATE* of method ι as:

$$\widehat{MSE}_{\iota,ATE} = \frac{1}{E} \sum_{e=1}^E \left(\widehat{ATE}_{\iota,e} - ATE \right)^2, \quad (39)$$

where e indexes the epoch and E is the total number of epochs, which, in our simulations, is 1,000. The term $\widehat{ATE}_{\iota,e}$ is the estimate of *ATE* obtained at the end of epoch e by method ι . For the A/B test this estimate corresponds to the OLS estimate of the slope coefficient of the regression of Y on D using data from all 100 rounds, while for ETC it corresponds to the same object using data from the first 50 rounds. In turn, for TS and BITS we use the final optimality probabilities to average over all bids in the grid, that is, we have that $\widehat{ATE}_{\iota,e} = \sum_{r=1}^R \hat{\psi}_{\iota,T,e}(b_r) \times b_r$.

Results

We now present the results from our simulation exercises. Before comparing BITS to the

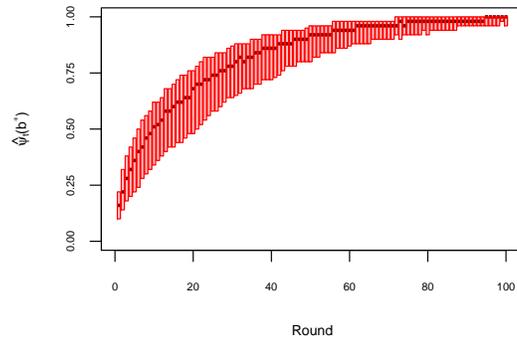
other methods, we first demonstrate it succeeds in recovering the optimal bid and, therefore, the *ATE*. To do so, for each round t we create a boxplot of $\hat{\psi}_t(b^*) = \hat{\Pr}_t(b^* \text{ is optimal})$ across the 1,000 epochs, where the edges of the box correspond to the interquartile range and the darker stripe corresponds to the median. For the algorithm to work we require that $\lim_{t \rightarrow +\infty} \hat{\psi}_t(b^*) = 1$. Figure 1 displays results using each of the three grids of bids described above.

Looking at the figure, we see that the BITS algorithm converges to the true optimal bid as data accumulate, which is most easily seen in Figure 1a due to the simplicity of its design since it considered only three arms. Unsurprisingly, as the number of arms increases the speed of convergence of the algorithm diminishes, as seen in Figures 1b and especially 1c. Notice that this is not only a function of the number of arms, but also because we kept its width fixed. Making the grid finer implies that the arm-specific expected rewards are closer to one another, so that the algorithm requires more data to precisely estimate and distinguish them.

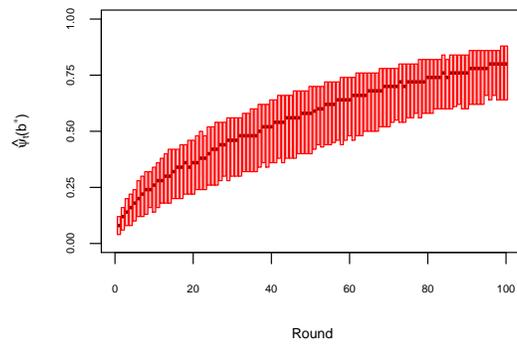
Having shown that the BITS algorithm consistently recovers the optimal bid, we now proceed to compare it to the aforementioned alternative methods. We begin by displaying the evolution of cumulative pseudo-regret averaged over the 1,000 epochs. Figure 2 displays the results for each method separately for each grid of bids.

Looking at the figure, we see that the BITS algorithm dominates the alternative methods in terms of average cumulative pseudo-regret. This is because BITS incorporates regret minimization as an explicit goal and fully exploits the structure of the data. As explained above, cumulative pseudo-regret is linear for an A/B test. For ETC, it increases much more slowly after the first 50 rounds. This is because 50 rounds worth of data, which in this cases corresponds to 2,500 observations, is enough to obtain a precise estimate of $b^* = ATE$. It is interesting to note that the off-the-shelf TS algorithm converges much more slowly than BITS, illustrating the consequence of ignoring the structure of the data. Under our specification, the off-the-shelf TS algorithm does not even overcome the simple ETC policy in terms of average cumulative pseudo-regret. Finally, notice that adding more bids to the grid while keeping its width fixed has a beneficial effect for minimizing regret. This is because while having more bids slows down the convergence of the algorithm, it considers bids whose expected rewards are closer to the optimal one.

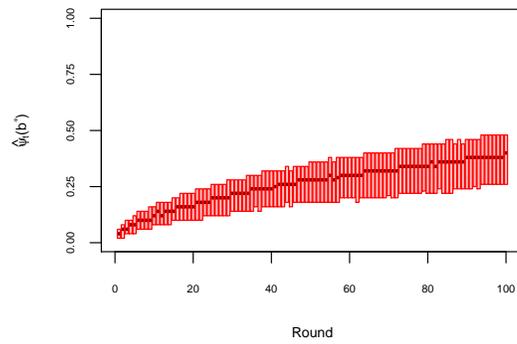
To conclude this analysis, we now present the results for $\widehat{MSE}_{t,ATE}$ as defined above. Once again, we compute different results for each one of the three grids of bids we con-



(a) 3 arms

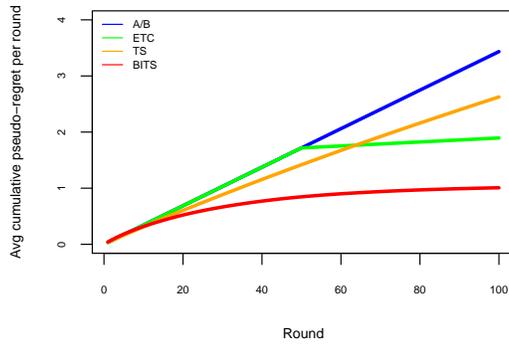


(b) 5 arms

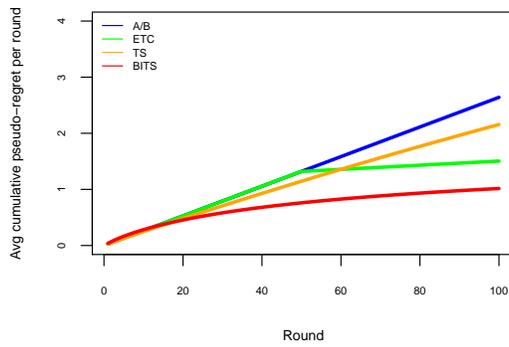


(c) 10 arms

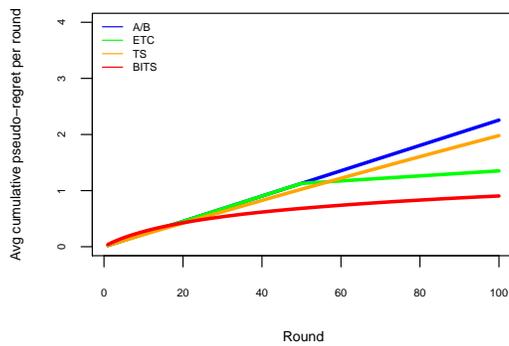
Figure 1: Convergence of BITS algorithm for SPAs



(a) 3 arms



(b) 5 arms



(c) 10 arms

Figure 2: Average cumulative pseudo-regret per method for SPAs

sider. Results are displayed in Table 2.

Table 2: $\widehat{MSE}_{l,ATE}$ for each estimation method for SPAs

Estimation method (l)	3 arms	5 arms	10 arms
A/B	0.007	0.007	0.007
ETC	0.012	0.013	0.016
TS	0.031	0.023	0.019
BITS	0.002	0.005	0.006

A few patterns become apparent. First, the ordering of performance does not change across the grids: BITS is the best at minimizing $\widehat{MSE}_{l,ATE}$, followed by A/B, ETC and then TS. Second, the performance of A/B does not seem to be altered by the inclusion of additional arms, while ETC is only slightly affected.

The effects of including additional arms is more interesting when we consider the adaptive methods, TS and BITS. Since the width of the arms is fixed, these results reflect the tradeoff between two elements: the addition of more arms with expected rewards that are closer to the optimal and the speed of convergence. For a slow algorithm such as TS, while the inclusion of additional arms further slows it down, it also forces the algorithm to consider options that are closer to the optimal one. As a result, at the end of 100 rounds the algorithm puts relatively more mass on closer-to-optimal arms, which diminishes its $\widehat{MSE}_{l,ATE}$. On the other hand, the end result is reversed for a fast algorithm such as BITS: the decrease in the speed of convergence offsets the consideration of more near-to-optimal alternatives, increasing the $\widehat{MSE}_{l,ATE}$.

Summary

The results displayed above showcase the qualities of the BITS algorithm for SPAs. As expected, it provides a consistent approach to recovering the optimal bid as shown in Figure 1 and it dominates typical alternative approaches to estimate the ATE in terms of regret as shown in Figure 2. In addition, we also document that the algorithm performs well on the inference goal via the $\widehat{MSE}_{l,ATE}$, which is illustrated in Table 2. While these results are a function of the specification we chose for the simulation exercises, they indicate that BITS

is an attractive option to achieve the advertiser’s dual objectives.

6.1.2 Contextual case

Having established the validity of BITS for the non-contextual case, we now proceed to analyze its performance for the contextual case. For brevity, we do not compare BITS to alternative methods here (the qualitative nature of the comparisons remain unaltered). Instead, we simply document its performance in recovering multiple bids, and thus *CATEs*, at the same time.

DGP, simulation details and grids of bids

For these simulations, we assume that:

$$\begin{bmatrix} \log Y_i(1) \\ \log Y_i(0) \\ \log B_{CP,i} \end{bmatrix} \Big|_{X_i} \stackrel{iid}{\sim} N \left(\begin{bmatrix} 0.81 & 1.04 & 1.25 & 1.43 & 1.57 \\ 0.20 & 0.31 & 0.45 & 0.59 & 0.70 \\ 0.25 & 0.33 & 0.40 & 0.47 & 0.55 \end{bmatrix} X_i, \begin{bmatrix} 0.36 & 0 & 0 \\ 0 & 0.64 & 0 \\ 0 & 0 & 0.25 \end{bmatrix} \right). \quad (40)$$

Hence, we have $P = 5$ contexts. The vector δ_1 is chosen so that the *ATEs* equals 1.00, 1.50, 2.00, 2.50 and 3.00. We assume that the contexts are equiprobable, so that the unconditional *ATE* is 2.00.

We simulate 1,000 different epochs. Each epoch has $T = 100$ rounds, each of which with 100 new observations, so that $n_t = 100$ for $1 \leq t \leq 100$, divided equally across the 5 contexts. Each context has its specific grid of bids under consideration. In particular, we consider the following grids:

1. Context 1: $b \in \{0.25, 0.50, 1.00, 1.25, 1.50\}$
2. Context 2: $b \in \{0.50, 1.00, 1.50, 2.00, 2.50\}$
3. Context 3: $b \in \{1.00, 1.50, 2.00, 2.50, 3.00\}$
4. Context 4: $b \in \{1.00, 1.75, 2.50, 3.25, 4.00\}$
5. Context 5: $b \in \{1.00, 2.00, 3.00, 4.00, 5.00\}$

Results

To demonstrate the validity of BITS to the contextual case we display the evolution of two objects. First, to show that the BITS consistently recovers $b^*(x)$ for all x , we plot the interquartiles and median across the 1,000 epochs of the lowest estimated optimality probability of the true optimal arm for each round across the different contexts, that is, $\min_{x \in \mathbb{X}} \hat{\psi}_t(b^*(x)|x)$. The convergence of this object to 1 implies that all best arms are concurrently recovered. Figure 3a shows that this is achieved.

Second, we show that the unconditional pseudo-regret per round, averaged over the different contexts, converges to 0. This is a consequence of BITS identifying the best arm for all contexts concurrently. Results are displayed in Figure 3b.

6.2 FPAs

6.2.1 Non-contextual case

We again begin with the non-contextual case. The structure of this exercise is very similar to the one implemented for SPAs.

DGP, simulation details and grids of bids

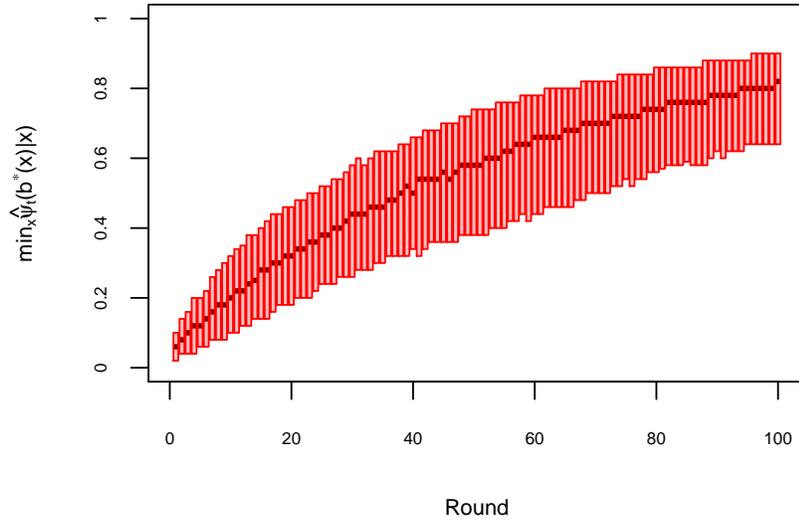
For these simulations, we assume that:

$$\begin{bmatrix} \log Y_i(1) \\ \log Y_i(0) \\ \log B_{CP,i} \end{bmatrix} \stackrel{iid}{\sim} N \left(\begin{bmatrix} 0.736 \\ 0.22 \\ 0.481 \end{bmatrix}, \begin{bmatrix} 0.49 & 0 & 0 \\ 0 & 0.81 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right), \quad (41)$$

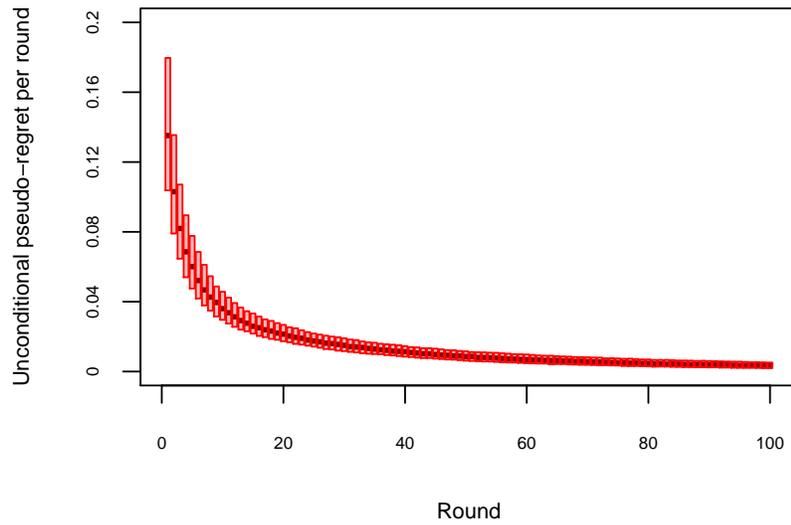
where the value of δ_1 is chosen so that $ATE = 0.8$ and the value of δ_{CP} is chosen so that $b^* = 0.5$. Again, these values are chosen for the sake of illustration.

We simulate 1,000 different epochs. Each epoch has $T = 100$ rounds, each of which with 50 new observations, so that $n_t = 50$ for $1 \leq t \leq 50$. We keep these values fixed and change the grid of bids we use to assess how the performance of our algorithm changes. In particular, we consider three different grids:

1. 3 arms: $b \in \{0.1, 0.5, 1.0\}$



(a) $\min_{x \in \mathcal{X}} \hat{\psi}_t(b^*(x)|x)$



(b) Unconditional pseudo-regret per round

Figure 3: Convergence of BITS algorithm for contextual case for SPAs

2. 5 arms: $b \in \{0.1, 0.3, 0.5, 0.75, 1.0\}$
3. 10 arms: $b \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$

As before, since the width of bids and the number of observations are kept fixed while the number of arms increases, we expect the performance of BITS to deteriorate as the number of arms increases.

Results

To demonstrate that the BITS algorithm recovers the optimal bid for FPAs we replicate Figure 1 in Figure 4. Once again, for the algorithm to work we require that $\lim_{t \rightarrow +\infty} \hat{\psi}_t(b^*) = 1$, which is indeed obtained as seen from the figure.

We now compare the performance of the BITS algorithm to that of A/B, ETC and TS using the same criteria as above. We start by replicating Figure 2. Results are displayed below in Figure 5. They are qualitatively identical to the previous ones in that BITS performs best compared to the other methods. One qualitative difference from the SPA case is noticeable: now the off-the-shelf TS outperforms the ETC policy, which uses the collected data to choose the best bid as the one whose observations yield highest sample profit, at least for the duration of this experiment. We note that the gap between these two methods diminishes as more arms are added to the grid.

As above, we also compare these methods in terms of \widehat{MSE} . However, we cannot perform this comparison with the off-the-shelf TS. Since this method does not exploit the link to auction payoffs, Proposition 2 no longer holds, so estimating the *ATE* leveraging this relationship is not viable. In turn, the estimator of *ATE* for BITS is $\widehat{ATE}_{t,e} = \sum_{r=1}^R \hat{\psi}_{t,T,e}(b_r) \times \hat{\chi}_T(b_r)$. Results comparing against the other methods are displayed in Table 3.

Unlike before, BITS is now dominated by A/B and ETC on the inference goal. This is not surprising. Since Proposition 1 no longer holds, the pursuit of the inference and the economic goals are not perfectly aligned for FPAs. Hence, pursuit of the economic goal (in which BITS beat the other methods) may come at the cost of performance of the economic goal. Another way to see the reduced performance on the inference goal in FPAs compared to SPAs is that in FPAs we have two sources of uncertainty in estimating the *ATE*, which lead to a higher \widehat{MSE} : the uncertainty over the optimal arm, as before

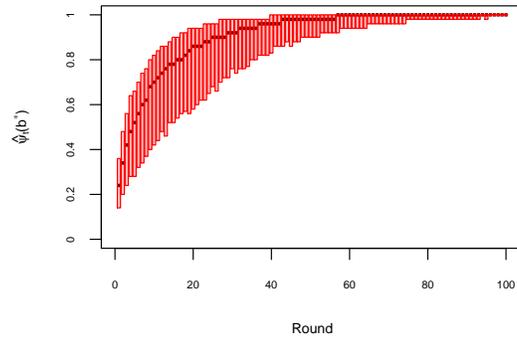
Table 3: $\widehat{MSE}_{\iota,ATE}$ for each estimation method for FPAs

Estimation method (ι)	3 arms	5 arms	10 arms
A/B	0.004	0.003	0.003
ETC	0.007	0.007	0.007
BITS	0.010	0.023	0.011

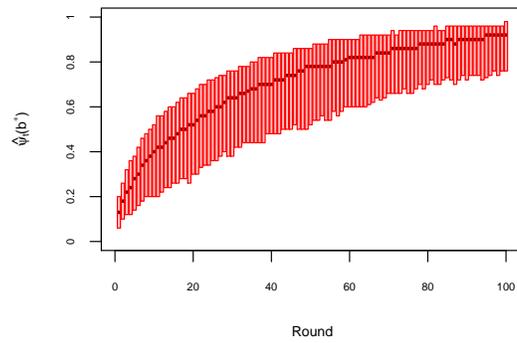
in SPAs, plus the uncertainty arising from the MCMC draws used to construct the ATE . Nevertheless, we note that the \widehat{MSE} s are still low, demonstrating that BITS can perform well in estimating the ATE for FPAs.

Finally, unlike in SPAs, the recovery by BITS of the true best bid in an FPA does not automatically imply it recovers the true ATE without bias. While solving the MAB problem consistently recovers b^* , to estimate the ATE in FPAs we also need to consistently estimate $\frac{F_{CP}(\cdot)}{f_{CP}(\cdot)}$ as seen in equation (8). One could wonder how the reversed hazard rate of B_{CP} is identified with FPAs, when one never actually observes B_{CP} . The intuition for the identification is that in each FPA auction we participate in, we observe a bound on B_{CP} corresponding to the bid we place. When we win an auction with a context x , the bid placed b is an upper bound on B_{CP} . When we lose an auction with a context x playing bid b , the bid placed b is a lower bound on B_{CP} . Thus, each auction yields an observation on the upper and lower bounds on B_{CP} . Under Assumption 1, these observations are from the truncated above or below marginal distributions of B_{CP} , and under Assumption 2, there will be observations of upper and lower bounds of B_{CP} for all b within the bid-grid for all x . Hence, the distribution of B_{CP} is identified for each x , as is its reversed hazard rate and, consequently, the $CATE$ s.

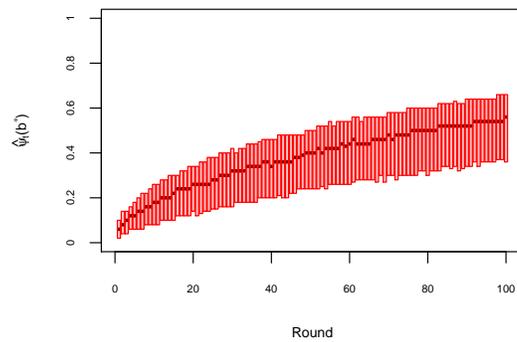
To demonstrate that the ATE is recovered without bias, Figure 6 shows the estimated density across epochs of the ATE estimated by BITS. Specifically, we estimate the ATE via equation (29) for each epoch, and use a Gaussian kernel and Silverman’s rule-of-thumb bandwidth. Looking at the figure, we see that the distribution is centered at $ATE = 0.8$, the true value, for all grid sizes, although, as expected, convergence is quicker when the number of arms is smaller.



(a) 3 arms

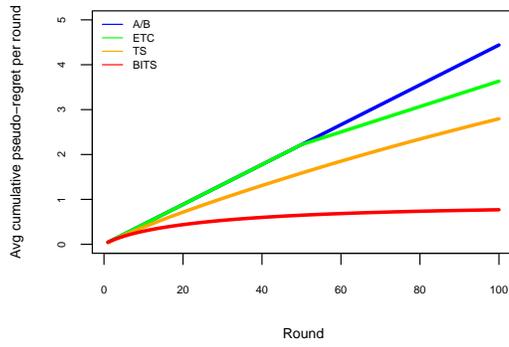


(b) 5 arms

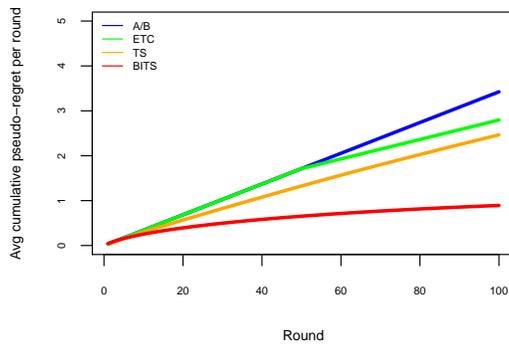


(c) 10 arms

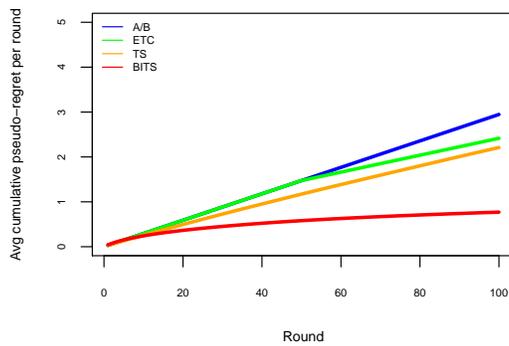
Figure 4: Convergence of BITS algorithm for FPAs



(a) 3 arms

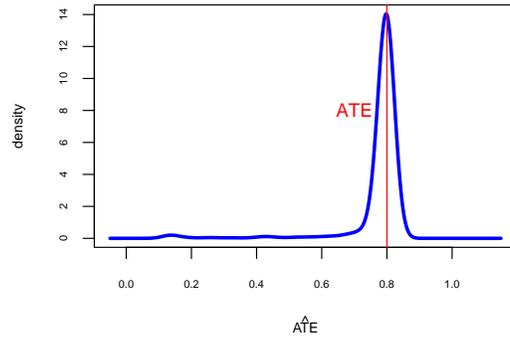


(b) 5 arms

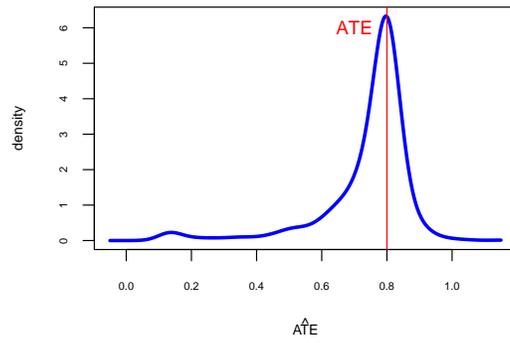


(c) 10 arms

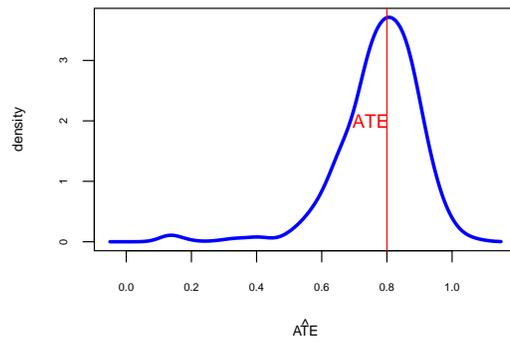
Figure 5: Average cumulative pseudo-regret per method for FPAs



(a) 3 arms



(b) 5 arms



(c) 10 arms

Figure 6: BITS estimate of ATE for FPAs

6.2.2 Contextual case

We now replicate the results for the contextual case for FPAs.

DGP, simulation details and grids of bids

For these simulations, we assume that:

$$\begin{bmatrix} \log Y_i(1) \\ \log Y_i(0) \\ \log B_{CP,i} \end{bmatrix} \Big| X_i \stackrel{iid}{\sim} N \left(\begin{bmatrix} 0.65 & 0.75 & 0.85 & 0.95 & 1.05 \\ 0.20 & 0.30 & 0.40 & 0.50 & 0.60 \\ -1.38 & -0.50 & 0.20 & 0.80 & 1.31 \end{bmatrix} X_i, \begin{bmatrix} 1.44 & 0 & 0 \\ 0 & 1.21 & 0 \\ 0 & 0 & 1.00 \end{bmatrix} \right). \quad (42)$$

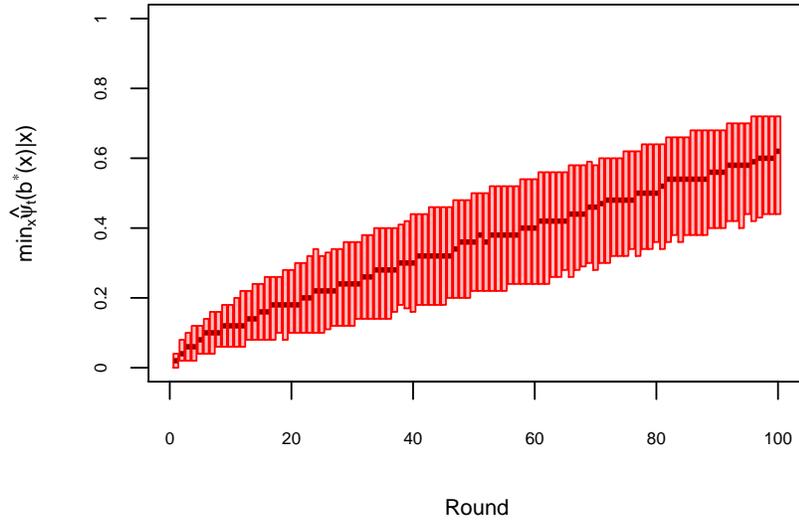
Hence, we have $P = 5$ contexts. The vector δ_{CP} is chosen so that b^* equals 0.50, 0.75, 1.00, 1.25 and 1.50. We assume the contexts are equiprobable.

We simulate 1,000 different epochs. Each epoch has $T = 100$ rounds, each of which with 100 new observations, so that $n_t = 100$ for $1 \leq t \leq 100$, divided equally across the 5 contexts. Each context has its specific grid of bids under consideration. In particular, we consider the following grids:

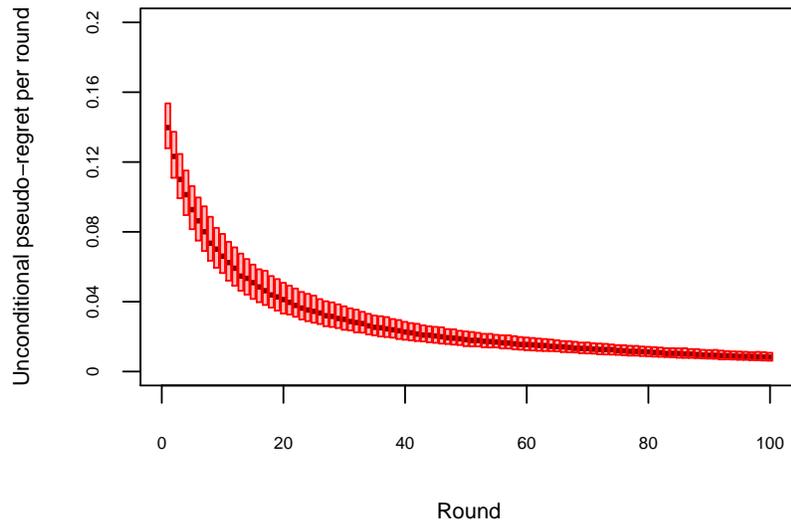
1. Context 1: $b \in \{0.25, 0.375, 0.50, 0.625, 0.75\}$
2. Context 2: $b \in \{0.25, 0.50, 0.75, 1.00, 1.25\}$
3. Context 3: $b \in \{0.40, 0.70, 1.00, 1.30, 1.60\}$
4. Context 4: $b \in \{0.45, 0.85, 1.25, 1.65, 2.05\}$
5. Context 5: $b \in \{0.50, 1.00, 1.50, 2.00, 2.50\}$

Results

We replicate Figure 3 obtained for SPAs, with analogous results displayed in Figure 7. Looking at the figure, we see we obtain the same qualitative patterns, demonstrating that the BITS algorithm succeeds in concurrently recovering the optimal bid for different contexts when the auction mechanism is an FPA.



(a) $\min_x \hat{\psi}_t(b^*(x)|x)$



(b) Unconditional pseudo-regret per round

Figure 7: Convergence of BITS algorithm for contextual case

7 Concluding remarks

An online experimental design for causal inference for RTB advertising is presented. The experimental design leverages the theory of optimal bidding under sealed-bid SPAs and FPAs to align the twin goals of obtaining economic payoff maximization and inference on the expected effect of advertising for varying subpopulations. The algorithm is framed as a contextual bandit implemented via a modified TS that is adaptively updated via MCMC. The SPA environment is historically the most popular auction format for RTB ads and the proposed experimental design perfectly aligns the economic and inference goals of the advertiser in this environment. Extensions to the more complex FPA environment, which has become more recently popular, are also presented.

Some broader implications of the experimental design beyond RTB advertising are worth mentioning. First, the ideas presented here can be useful in other situations outside of ad-auctions where there is a cost to obtaining experimental units and where managing these costs is critical for the viability of the experiment. Another implication is that, in business experiments, incorporating the firm's payoff or profit maximization goal into the allocation and acquisition of experimental units is helpful. Given the burgeoning utilization of experimentation by firms, we believe that leveraging this perspective in business experiments more broadly has value. Finally, another takeaway from the proposed approach is that it demonstrates the utility of embedding experimental design in the micro-foundations of the problem, which enables leveraging economic theory to make progress on running experiments more efficiently. This aspect could be utilized in other settings where large-scale experimentation is feasible and where economic theory puts structure on the behavior of agents and associated outcomes.

References

- Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.
- Amemyia, T. (1984). Tobit models: A survey. *Journal of Econometrics*, 24(1–2):3–61.
- Athey, S. and Haile, P. A. (2002). Identification of standard auction models. *Econometrica*, 70(6):2107–2140.
- Austin, D., Seljan, S., Moreno, J., and Tzeng, S. (2016). Reserve price optimization at scale. In Zaiane, O. R. and Matwin, S., editors, *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 528–536, New York, USA. IEEE.
- Balseiro, S. R., Besbes, O., and Weintraub, G. Y. (2015). Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Science*, 61(4):864–884.
- Balseiro, S. R., Golrezaei, N., Mahdian, M., Mirrokni, V., and Schneider, J. (2019). Contextual bandits with cross-learning. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, pages 9679–9688, New York, USA. Curran Associates, Inc.
- Balseiro, S. R. and Gur, Y. (2019). Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968.
- Bareinboim, E., Forney, A., and Pearl, J. (2015). Bandits with unobserved confounders: A causal approach. In Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pages 1342–1350, New York, USA. Curran Associates, Inc.
- Bastani, H. and Bayati, M. (2020). Online decision-making with high-dimensional covariates. *Operations Research*, 68(1):276–294.
- Bergemann, D. and Välimäki, J. (2008). Bandit problems. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics: Volume 1 – 8*, pages 336–340. Palgrave Macmillan UK, London, UK.
- Blake, T., Nosko, C., and Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, 83(1):155–174.

- Block, H. W., Savits, T. H., and Singh, H. (1998). The reversed hazard rate function. *Probability in the Engineering and Informational Sciences*, 12(1):69–90.
- Bubeck, S. and Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122.
- Bubeck, S., Munos, R., and Stoltz, G. (2009). Pure exploration in multi-armed bandits problems. In Gavalda, R., Lugosi, G., Zeugmann, T., and Zilles, S., editors, *International Conference on Algorithmic Learning Theory (ALT 2009)*, pages 23–37, Berlin, Germany. Springer.
- Cai, H., Ren, K., Zhang, W., Malialis, K., Wang, J., Yu, Y., and Guo, D. (2017). Real-time bidding by reinforcement learning in display advertising. In de Rijke, M. and Shokouhi, M., editors, *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM '17)*, pages 661–670, New York, USA. ACM.
- Cesa-Bianchi, N., Gentile, C., and Mansour, Y. (2014). Regret minimization for reserve prices in second-price auctions. *IEEE Transactions on Information Theory*, 61(1):549–564.
- Chan, J., Koop, G., Poirier, D. J., and Tobias, J. L. (2020). *Bayesian Econometric Methods*. Cambridge University Press.
- Chawla, S., Hartline, J., and Nekipelov, D. (2016). A/B testing of auctions. *arXiv preprint arXiv:1606.00908*.
- Cherkassky, M. and Bornn, L. (2013). Sequential Monte Carlo bandits. *arXiv preprint arXiv:1310.1404*.
- Chib, S. (1992). Bayes inference in the Tobit censored regression model. *Journal of Econometrics*, 51(1–2):79–99.
- Chib, S. and Hamilton, B. H. (2000). Bayesian analysis of cross-section and clustered data treatment models. *Journal of Econometrics*, 97(1):25–50.
- Choi, H., Mela, C., Balseiro, S., and Leary, A. (2020). Online display advertising markets: A literature review and future directions. *Information Systems Research*, 31(2):556–575.
- de Heide, R. and Grünwald, P. D. (2020). Why optional stopping can be a problem for Bayesians. *Psychonomic Bulletin & Review*, forthcoming.
- Deng, A., Lu, J., and Chen, S. (2016). Continuous monitoring of A/B tests without pain: Optional stopping in Bayesian testing. In Zaiane, O. R. and Matwin, S., editors, *2016*

- IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 243–252, New York, USA. IEEE.
- Despotakis, S., Ravi, R., and Sayedi, A. (2019). First-price auctions in online display advertising. *SSRN:3485410*.
- Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72:78–89.
- Dikkala, N. and Tardos, É. (2013). Can credit increase revenue? In Chen, Y. and Immorlica, N., editors, *9th International Conference on Web and Internet Economics (WINE 2013)*, pages 121–133, Berlin, Germany. Springer.
- Dimakopoulou, M., Zhou, Z., Athey, S., and Imbens, G. (2018). Estimation considerations in contextual bandits. *arXiv preprint arXiv:1711.07077*.
- Edwards, W., Lindman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, 70(3):193–242.
- Feit, E. M. and Berman, R. (2019). Test & roll: Profit-maximizing A/B tests. *Marketing Science*, 38(6):1038–1058.
- Feng, Z., Podimata, C., and Syrgkanis, V. (2018). Learning to bid without knowing your value. In Tardos, É., editor, *Proceedings of the 2018 ACM Conference on Economics and Computation (EC '18)*, pages 505–522, New York, USA. ACM.
- Forney, A., Pearl, J., and Bareinboim, E. (2017). Counterfactual data-fusion for online reinforcement learners. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, pages 1156–1164, Sidney, Australia. PMLR.
- Geng, T., Lin, X., and Nair, H. S. (2020). Online evaluation of audiences for targeted advertising via bandit experiments. In Rossi, F., editor, *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI '20)*, pages 13273–13279, Palo Alto, USA. AAAI Press.
- Goldenshluger, A. and Zeevi, A. (2013). A linear response bandit problem. *Stochastic Systems*, 3(1):230–261.
- Good, I. J. (1991). A comment concerning optional stopping. *Journal of Statistical Computation and Simulation*, 39(3):191–192.

- Gopalan, A., Mannor, S., and Mansour, Y. (2014). Thompson sampling for complex online problems. In Xing, E. P. and Jebara, T., editors, *Proceedings of the 31st International Conference on Machine Learning (ICML 2014)*, pages 100–108, Beijing, China. PMLR.
- Gordon, B. R., Jerath, K., Katona, Z., Narayanan, S., Shin, J., and Wilbur, K. C. (2021). Inefficiencies in digital advertising markets. *Journal of Marketing*, 85(1):7–25.
- Hadad, V., Hirshberg, D. A., Zhan, R., Wager, S., and Athey, S. (2021). Confidence intervals for policy evaluation in adaptive experiments. *arXiv preprint arXiv:1911.02768*.
- Han, Y., Zhou, Z., Flores, A., Ordentlich, E., and Weissman, T. (2020a). Learning to bid optimally and efficiently in adversarial first-price auctions. *arXiv preprint arXiv:2007.04568*.
- Han, Y., Zhou, Z., and Weissman, T. (2020b). Optimal no-regret learning in repeated first-price auctions. *arXiv preprint arXiv:2003.09795*.
- Haoyu, Z. and Wei, C. (2020). Online second price auction with semi-bandit feedback under the non-stationary setting. In Rossi, F., editor, *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI '20)*, pages 6893–6900, Palo Alto, USA. AAAI Press.
- Heckman, J. J., Lopes, H. F., and Piatek, R. (2014). Treatment effects: A Bayesian perspective. *Econometric Reviews*, 33(1–4):36–67.
- Hendriksen, A., de Heide, R., and Grünwald, P. D. (2021). Optional stopping with Bayes factors: A categorization and extension of folklore results, with an application to invariant situations. *Bayesian Analysis*, forthcoming.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960.
- Imbens, G. W. and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *Annals of Statistics*, 25(1):305–327.
- Iyer, K., Johari, R., and Sundararajan, M. (2014). Mean field equilibria of dynamic auctions with learning. *Management Science*, 60(12):2949–2970.
- Jamieson, K. G. and Jain, L. (2018). A bandit approach to sequential experimental design with false discovery control. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)*, pages 3664–3674, New York, USA. Curran Associates, Inc.

- Jin, J., Song, C., Li, H., Gai, K., Wang, J., and Zhang, W. (2018). Real-time bidding with multi-agent reinforcement learning in display advertising. In Cuzzocrea, A., editor, *Proceedings of the 27th ACM International Conference on Information and Knowledge Management (CIKM '18)*, pages 2193–2201, New York, USA. ACM.
- Johari, R., Pekelis, L., and Walsh, D. J. (2019). Always valid inference: Bringing sequential analysis to A/B testing. *Operations Research*, forthcoming.
- Johnson, G. A., Lewis, R. A., and Nubbemeyer, E. I. (2017). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, 54(6):867–884.
- Ju, N., Hu, D., Henderson, A., and Hong, L. (2019). A sequential test for selecting the better variant: Online A/B testing, adaptive allocation, and continuous monitoring. In Culpepper, J. S. and Moffat, A., editors, *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19)*, pages 492–500, New York, USA. ACM.
- Kallus, N. (2018). Instrument-armed bandits. In Janoos, F., Mohri, M., and Sridharan, K., editors, *International Conference on Algorithmic Learning Theory (ALT 2018)*, pages 529–546, Lanzarote, Spain. PMLR.
- Kanoria, Y. and Nazerzadeh, H. (2021). Incentive-compatible learning of reserve prices for repeated auctions. *Operations Research*, forthcoming.
- Kasy, M. and Sautmann, A. (2021). Adaptive treatment assignment in experiments for policy choice. *Econometrica*, 89(1):113–132.
- Koop, G. and Poirier, D. J. (1997). Learning about the across-regime correlation in switching regression models. *Journal of Econometrics*, 78(2):217–227.
- Lattimore, F., Lattimore, T., and Reid, M. D. (2016). Causal bandits: Learning good interventions via causal inference. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, pages 1181–1189, New York, USA. Curran Associates, Inc.
- Lewis, R. and Wong, J. (2018). Incrementality bidding & attribution. *SSRN:3129350*.
- Lindley, D. V. (1957). A statistical paradox. *Biometrika*, 44(1/2):187–192.
- McAfee, R. P. (2011). The design of advertising exchanges. *Review of Industrial Organization*, 39(3):169–185.

- Milgrom, P. R. and Weber, R. J. (1982). A theory of auctions and competitive bidding. *Econometrica*, 50(5):1089–1122.
- Misra, K., Schwartz, E. M., and Abernethy, J. (2019). Dynamic online pricing with incomplete information using multi-armed bandit experiments. *Marketing Science*, 38(2):226–252.
- Mohri, M. and Medina, A. M. (2016). Learning algorithms for second-price auctions with reserve. *Journal of Machine Learning Research*, 17(1):2632–2656.
- Muthukrishnan, S. (2009). Ad exchanges: Research issues. In Leonardi, S., editor, *5th International Conference on Web and Internet Economics (WINE 2009)*, pages 1–12, Berlin, Germany. Springer.
- Nie, X., Tian, X., Taylor, J., and Zou, J. (2018). Why adaptively collected data have negative bias and how to correct for it. In Storkey, A. and Perez-Cruz, F., editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics (AISTATS 2018)*, pages 1261–1269, Lanzarote, Spain. PMLR.
- Olsen, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the Tobit model. *Econometrica*, 46(5):1211–1215.
- Ostrovsky, M. and Schwarz, M. (2016). Reserve prices in internet advertising auctions: A field experiment. *Working paper, Stanford University*.
- Pearl, J. (2009). *Causality*. Cambridge University Press.
- Pouget-Abadie, J., Mirrokni, V., Parkes, D. C., and Airoidi, E. M. (2018). Optimizing cluster-based randomized experiments under monotonicity. In Guo, Y. and Farooq, F., editors, *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*, pages 2090–2099, New York, USA. ACM.
- Rhuggenaath, J., Akcay, A., Zhang, Y., and Kaymak, U. (2019). Optimizing reserve prices for publishers in online ad auctions. In Ishibuchi, H. and Zhao, D., editors, *2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFER)*, pages 1–8, New York, USA. IEEE.
- Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. John Wiley & Sons.
- Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, 21(2):301–308.

- Rouder, J. N. and Haaf, J. M. (2020). Optional stopping and the interpretation of Bayes factor. *PsyArXiv Preprints*.
- Roughgarden, T. and Wang, J. R. (2019). Minimizing regret with multiple reserves. *ACM Transactions on Economics and Computation*, 7(3):17:1–17:18.
- Russo, D. J. (2020). Simple Bayesian algorithms for best-arm identification. *Operations Research*, 68(6):1625–1647.
- Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., and Wen, Z. (2018). A tutorial on Thompson sampling. *Foundations and Trends® in Machine Learning*, 11(1):1–96.
- Sahni, N. S., Narayanan, S., and Kalyanam, K. (2019). An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *Journal of Marketing Research*, 56(3):401–418.
- Sanborn, A. N. and Hills, T. T. (2014). The frequentist implications of optional stopping on Bayesian hypothesis tests. *Psychonomic Bulletin & Review*, 21(2):283–300.
- Savage, L. J. (1972). *The Foundations of Statistics*. Courier Corporation.
- Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., and Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2):322–339.
- Schwartz, E. M., Bradlow, E. T., and Fader, P. S. (2017). Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4):500–522.
- Scott, S. L. (2015). Multi-armed bandit experiments in the online service economy. *Applied Stochastic Models in Business and Industry*, 31(1):37–45.
- Simonov, A., Nosko, C., and Rao, J. M. (2018). Competition and crowd-out for brand keywords in sponsored search. *Marketing Science*, 37(2):200–215.
- Tendeiro, J. N., Kiers, H. A. L., and van Ravenzwaaij, D. (2019). Mathematical evidence for the adequacy of Bayesian optional stopping. *PsyArXiv Preprints*.
- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.
- Tunuguntla, S. and Hoban, P. R. (2021). A near-optimal bidding strategy for real-time display advertising auctions. *Journal of Marketing Research*, 58(1):1–21.

- Vijverberg, W. P. M. (1993). Measuring the unidentified parameter of the extended Roy model of selectivity. *Journal of Econometrics*, 57(1–3):69–89.
- Villar, S. S., Bowden, J., and Wason, J. M. S. (2015). Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statistical Science*, 30(2):199–215.
- Wagenmakers, E.-J., Gronau, Q. F., and Vandekerckhove, J. (2019). Five Bayesian intuitions for the stopping rule principle. *PsyArXiv Preprints*.
- Weed, J., Perchet, V., and Rigollet, P. (2016). Online learning in repeated auctions. In Feldman, V., Rakhlin, A., and Shamir, O., editors, *29th Annual Conference on Learning Theory*, pages 1562–1583, New York, USA. PMLR.
- Wu, D., Chen, C., Yang, X., Chen, X., Tan, Q., Xu, J., and Gai, K. (2018). A multi-agent reinforcement learning method for impression allocation in online display advertising. *arXiv preprint arXiv:1809.03152*.
- Xu, M., Qin, T., and Liu, T.-Y. (2013). Estimation bias in multi-armed bandit algorithms for search advertising. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pages 2400–2408, New York, USA. Curran Associates, Inc.
- Yang, F., Ramdas, A., Jamieson, K. G., and Wainwright, M. J. (2017). A framework for Multi-A(rmed)/B(andid) testing with online FDR control. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 5957–5966, New York, USA. Curran Associates, Inc.
- Yu, E. C., Sprenger, A. M., Thomas, R. P., and Dougherty, M. R. (2014). When decision heuristics and science collide. *Psychonomic Bulletin & Review*, 21(2):268–282.

Appendix

A Full conditional distributions for Gibbs sampling

This section outlines the specific full conditional distributions used to sample from the posterior of the Markov chain induced by the BITS algorithm. In Section 5.4.3, because of the parametric forms in (11) and (14), we have that,

$$\begin{aligned}
& \sigma_{CP}^{-2,(q)} \Big| \log \tilde{B}_{CP,t}, X_t, \mu_{\delta_{CP}}, A_{CP}, \alpha_{CP}, \beta_{CP} \\
& \quad \sim \Gamma \left(\alpha_{CP} + \frac{N_t}{2}, \beta_{CP} + \frac{1}{2} \left[(\log \tilde{B}_{CP,t} - X_t \hat{\delta}_{CP,t})' (\log \tilde{B}_{CP,t} - X_t \hat{\delta}_{CP,t}) \right. \right. \\
& \quad \quad \left. \left. + (\hat{\delta}_{CP,t} - \mu_{\delta_{CP}})' X_t' X_t (A_{CP} + X_t' X_t)^{-1} A_{CP} (\hat{\delta}_{CP,t} - \mu_{\delta_{CP}}) \right] \right) \\
& \sigma_1^{-2,(q)} \Big| \log \tilde{Y}_t(1), X_t, \mu_{\delta_1}, A_1, \alpha_1, \beta_1 \\
& \quad \sim \Gamma \left(\alpha_1 + \frac{N_t}{2}, \beta_1 + \frac{1}{2} \left[(\log \tilde{Y}_t(1) - X_t \hat{\delta}_{1,t})' (\log \tilde{Y}_t(1) - X_t \hat{\delta}_{1,t}) \right. \right. \\
& \quad \quad \left. \left. + (\hat{\delta}_{1,t} - \mu_{\delta_1})' X_t' X_t (A_1 + X_t' X_t)^{-1} A_1 (\hat{\delta}_{1,t} - \mu_{\delta_1}) \right] \right) \tag{A.1} \\
& \sigma_0^{-2,(q)} \Big| \log \tilde{Y}_t(0), X_t, \mu_{\delta_0}, A_0, \alpha_0, \beta_0 \\
& \quad \sim \Gamma \left(\alpha_0 + \frac{N_t}{2}, \beta_0 + \frac{1}{2} \left[(\log \tilde{Y}_t(0) - X_t \hat{\delta}_{0,t})' (\log \tilde{Y}_t(0) - X_t \hat{\delta}_{0,t}) \right. \right. \\
& \quad \quad \left. \left. + (\hat{\delta}_{0,t} - \mu_{\delta_0})' X_t' X_t (A_0 + X_t' X_t)^{-1} A_0 (\hat{\delta}_{0,t} - \mu_{\delta_0}) \right] \right)
\end{aligned}$$

and

$$\begin{aligned}
& \delta_{CP}^{(q)} \Big| \sigma_{CP}^{2,(q)} \log \tilde{B}_{CP,t}, X_t, \mu_{\delta_{CP}}, A_{CP} \\
& \quad \sim N \left((A_{CP} + X_t' X_t)^{-1} (X_t' \log \tilde{B}_{CP,t} + A_{CP} \mu_{\delta_{CP}}), \sigma_{CP}^2 (A_{CP} + X_t' X_t)^{-1} \right) \\
& \delta_1^{(q)} \Big| \sigma_1^{2,(q)} \log \tilde{Y}_t(1), X_t, \mu_{\delta_1}, A_1 \\
& \quad \sim N \left((A_1 + X_t' X_t)^{-1} (X_t' \log \tilde{Y}_t(1) + A_1 \mu_{\delta_1}), \sigma_1^2 (A_1 + X_t' X_t)^{-1} \right) \\
& \delta_0^{(q)} \Big| \sigma_0^{2,(q)} \log \tilde{Y}_t(0), X_t, \mu_{\delta_0}, A_0 \\
& \quad \sim N \left((A_0 + X_t' X_t)^{-1} (X_t' \log \tilde{Y}_t(0) + A_0 \mu_{\delta_0}), \sigma_0^2 (A_0 + X_t' X_t)^{-1} \right),
\end{aligned} \tag{A.2}$$

where

$$\begin{aligned}
\hat{\delta}_{CP,t} &= (X_t' X_t)^{-1} X_t' \log \tilde{B}_{CP,t} \\
\hat{\delta}_{1,t} &= (X_t' X_t)^{-1} X_t' \log \tilde{Y}_t(1) \\
\hat{\delta}_{0,t} &= (X_t' X_t)^{-1} X_t' \log \tilde{Y}_t(0),
\end{aligned} \tag{A.3}$$

B Maximum likelihood estimators used on historical data

This section describes in more detail the maximum likelihood estimators (MLEs) that can be used to choose the parameters of the prior distributions. The assumptions we make on the historical data are the same as the ones discussed in Section 5.8.6.

B.1 Potential outcomes

We begin by describing how we use historical data to pick the parameters for the prior distributions associated with the potential outcomes. Because we are maintaining the assumption that $D_i \perp\!\!\!\perp Y_i(1), Y_i(0) | X_i$, we can simply use OLS on historical data to pick the parameters since it is equivalent to the MLE. In particular, define $X_{i1} \equiv D_i X_i$. It follows that:

$$\hat{\delta}_1 = \left(\frac{1}{n} \sum_{i=1}^n X_{i1} X_{i1}' \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_{i1} \log Y_i \right),$$

$$\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n (\log Y_i - X_{i1}' \hat{\delta}_1)^2$$

and

$$\sqrt{n} \begin{bmatrix} \hat{\delta}_1 - \delta_1 \\ \hat{\sigma}_1^2 - \sigma_1^2 \end{bmatrix} \xrightarrow{d} N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 (\mathbb{E} [X_{i1} X_{i1}'])^{-1} & 0 \\ 0' & 2\sigma_1^2 \end{bmatrix} \right).$$

Hence,

$$\hat{Avar} [\sqrt{n} (\hat{\delta}_1 - \delta_1)] = \hat{\sigma}_1^2 \left(\frac{1}{n} \sum_{i=1}^n X_{i1} X_{i1}' \right)^{-1}$$

and

$$A\hat{var} \left[\sqrt{n} \left(\hat{\sigma}_1^2 - \sigma_1^2 \right) \right] = 2\hat{\sigma}_1^4.$$

The estimators $\hat{\delta}_0$ and $\hat{\sigma}_0^2$ are analogous to the ones above, with $X_{i0} \equiv (1 - D_i)X_i$ replacing X_{i1} , so we omit them for brevity.

B.2 Highest competing bid for SPAs

Even though we maintain the assumption of treatment exogeneity, we still have to account for censoring of the highest competing bid. Given the normality assumption, the censoring characterizes a standard Tobit model. To make its MLE more computationally manageable, we first reparametrize the model so that the log-likelihood function becomes globally concave as first shown by [Olsen \(1978\)](#). Let $\aleph_{CP} \equiv \sigma_{CP}^{-1}\delta_{CP}$ and $\beth_{CP} \equiv \sigma_{CP}^{-1}$. The log-likelihood of the data is then given by:

$$\begin{aligned} \log L(W_n | \aleph_{CP}, \beth_{CP}) = & \frac{1}{n} \sum_{i=1}^n \{ D_i \log [\beth_{CP} \phi (\beth_{CP} \log b_{CP,i} - X_i' \aleph_{CP})] \\ & + (1 - D_i) \log [\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)] \}. \end{aligned}$$

We use the Newton-Raphson algorithm to compute the estimator. This requires us to compute the first and second derivatives of the log-likelihood function. We have that:

$$\begin{aligned}
\frac{\partial \log L}{\partial \aleph_{CP}} &= \frac{1}{n} \sum_{i=1}^n \left\{ D_i (\beth_{CP} \log b_{CP,i} - X_i' \aleph_{CP}) + (1 - D_i) \frac{\phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)}{\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)} \right\} X_i \\
\frac{\partial \log L}{\partial \beth_{CP}} &= \frac{1}{n} \sum_{i=1}^n \left\{ D_i \left[\frac{1}{\beth_{CP}} - \log b_{CP,i} (\beth_{CP} \log b_{CP,i} - X_i' \aleph_{CP}) \right] \right. \\
&\quad \left. - (1 - D_i) \frac{\phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)}{\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)} \log b_i \right\} \\
\frac{\partial^2 \log L}{\partial \aleph_{CP} \partial \aleph_{CP}'} &= -\frac{1}{n} \sum_{i=1}^n \left\{ D_i - (1 - D_i) \frac{\phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)}{\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)} \right. \\
&\quad \left. \times \left[(X_i' \aleph_{CP} - \beth_{CP} \log b_i) - \frac{\phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)}{\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)} \right] \right\} X_i X_i' \\
\frac{\partial^2 \log L}{\partial \aleph_{CP} \partial \beth_{CP}} &= \frac{1}{n} \sum_{i=1}^n \left\{ D_i \log b_{CP,i} + (1 - D_i) \frac{\phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)}{\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)} \right. \\
&\quad \left. \times \left[(X_i' \aleph_{CP} - \beth_{CP} \log b_i) - \frac{\phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)}{\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)} \right] \log b_i \right\} X_i \\
\frac{\partial^2 \log L}{\partial \beth_{CP}^2} &= -\frac{1}{n} \sum_{i=1}^n \left\{ D_i \left(\frac{1}{\beth_{CP}^2} + (\log b_{CP,i})^2 \right) - (1 - D_i) \frac{\phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)}{\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)} \right. \\
&\quad \left. \times \left[(X_i' \aleph_{CP} - \beth_{CP} \log b_i) - \frac{\phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)}{\Phi (X_i' \aleph_{CP} - \beth_{CP} \log b_i)} \right] (\log b_i)^2 \right\}.
\end{aligned}$$

Letting

$$H(\aleph_{CP}, \beth_{CP}) = \begin{bmatrix} \frac{\partial^2 \log L(\aleph_{CP}, \beth_{CP})}{\partial \aleph_{CP} \partial \aleph_{CP}'} & \frac{\partial^2 \log L(\aleph_{CP}, \beth_{CP})}{\partial \aleph_{CP} \partial \beth_{CP}} \\ \frac{\partial^2 \log L(\aleph_{CP}, \beth_{CP})}{\partial \aleph_{CP}' \partial \beth_{CP}} & \frac{\partial^2 \log L(\aleph_{CP}, \beth_{CP})}{\partial \beth_{CP}^2} \end{bmatrix},$$

it then follows that

$$\sqrt{n} \begin{bmatrix} \hat{\aleph}_{CP} - \aleph_{CP} \\ \hat{\beth}_{CP} - \beth_{CP} \end{bmatrix} \xrightarrow{d} N \left(0, -\text{plim}_{n \rightarrow \infty} \left[H(\hat{\aleph}_{CP}, \hat{\beth}_{CP})^{-1} \right] \right).$$

To convert the parameters back to the original ones, we use the delta method. We

have that:

$$j(\aleph_{CP}, \beth_{CP}) = \begin{bmatrix} \beth_{CP}^{-1} \aleph_{CP} \\ \beth_{CP}^{-2} \end{bmatrix},$$

which implies that

$$\nabla j(\aleph_{CP}, \beth_{CP}) = \begin{bmatrix} \frac{\partial j(\aleph_{CP}, \beth_{CP})}{\partial \aleph_{CP}} & \frac{\partial j(\aleph_{CP}, \beth_{CP})}{\partial \beth_{CP}} \end{bmatrix} = \begin{bmatrix} \beth_{CP}^{-1} I_p & -\beth_{CP}^{-2} \aleph_{CP} \\ 0 & -2\aleph_{CP}^{-3} \end{bmatrix},$$

where I_p is the identity matrix with dimension p . Thus, by the delta method:

$$\sqrt{n} \begin{bmatrix} \hat{\delta}_{CP} - \delta_{CP} \\ \hat{\sigma}_{CP}^2 - \sigma_{CP}^2 \end{bmatrix} \xrightarrow{d} N \left(0, -\text{plim}_{n \rightarrow \infty} \left[\nabla j(\hat{\aleph}_{CP}, \hat{\beth}_{CP}) H(\hat{\aleph}_{CP}, \hat{\beth}_{CP})^{-1} \nabla j(\hat{\aleph}_{CP}, \hat{\beth}_{CP})' \right] \right).$$

Finally, $A\hat{v}ar[\sqrt{n}(\hat{\delta}_{CP} - \delta_{CP})]$ and $A\hat{v}ar[\sqrt{n}(\hat{\sigma}_{CP}^2 - \sigma_{CP}^2)]$ are obtained by picking the block diagonal elements of the matrix

$$-\nabla j(\hat{\aleph}_{CP}, \hat{\beth}_{CP}) H(\hat{\aleph}_{CP}, \hat{\beth}_{CP})^{-1} \nabla j(\hat{\aleph}_{CP}, \hat{\beth}_{CP})'.$$

B.3 Highest competing bid for FPAs

When the data come from FPAs, the highest competing bid, B_{CP} , is never observed. Given our assumptions and the necessity to normalize $\sigma_{CP}^2 = 1$, to recover δ_{CP} we need to estimate a standard Probit model. The log-likelihood of the data is then given by:

$$\begin{aligned} \log L(W_n | \delta_{CP}) &= \frac{1}{n} \sum_{i=1}^n \left\{ D_i \log \Phi(\log b_i - X_i' \delta_{CP}) \right. \\ &\quad \left. + (1 - D_i) \log [1 - \Phi(\log b_i - X_i' \delta_{CP})] \right\}. \end{aligned}$$

Once again, to use the Newton-Raphson algorithm we need to compute the first and second derivatives of the log-likelihood function. We have that:

$$\frac{\partial \log L}{\partial \delta_{CP}} = -\frac{1}{n} \sum_{i=1}^n \left\{ D_i \frac{\phi(\log b_i - X_i' \delta_{CP})}{\Phi(\log b_i - X_i' \delta_{CP})} - (1 - D_i) \frac{\phi(\log b_i - X_i' \delta_{CP})}{1 - \Phi(\log b_i - X_i' \delta_{CP})} \right\} X_i$$

and

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \delta_{CP} \partial \delta'_{CP}} &= -\frac{1}{n} \sum_{i=1}^n \phi(\log b_i - X'_i \delta_{CP}) \{ \Phi(\log b_i - X'_i \delta_{CP}) [1 - \Phi(\log b_i - X'_i \delta_{CP})] \}^{-2} \\ &\quad \times \{ [D_i - \Phi(\log b_i - X'_i \delta_{CP})] \Phi(\log b_i - X'_i \delta_{CP}) [1 - \Phi(\log b_i - X'_i \delta_{CP})] \\ &\quad \times (\log b_i - X'_i \delta_{CP}) + \phi(\log b_i - X'_i \delta_{CP}) [D_i - 2D_i \Phi(\log b_i - X'_i \delta_{CP}) \\ &\quad + \Phi^2(\log b_i - X'_i \delta_{CP})] \} X_i X'_i. \end{aligned}$$

We therefore have that:

$$\sqrt{n} (\hat{\delta}_{CP} - \delta_{CP}) \xrightarrow{d} N \left(0, -\text{plim}_{n \rightarrow \infty} \left(\frac{\partial^2 \log L}{\partial \delta_{CP} \partial \delta'_{CP}} \right)^{-1} \right).$$

C Gibbs sampling when potential outcomes are correlated

We now present a more general Gibbs sampling procedure that accommodates the possibility that $\rho \neq 0$. When $\rho \neq 0$, the missing values $\log Y_i^{miss}$ depend on the observed values $\log Y_i$ even conditional on D_i , which requires us to change the priors and the procedure accordingly. To do so, we combine the Bayesian estimator for the standard Tobit model introduced by [Chib \(1992\)](#) for SPAs or the Bayesian estimator for the Probit model introduced by [Albert and Chib \(1993\)](#) for FPAs with the approach to estimate the parameters in a seemingly unrelated regressions (SUR) model where all equations have the same set of regressors with data augmentation in a single Gibbs sampling algorithm.²⁰ We now present these adaptations in detail.

C.1 Prior distributions

For $k \in \{1, 0\}$ we replace (14) with

$$\begin{aligned} \Sigma^{-1} &\equiv \begin{bmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_0 \\ \rho \sigma_1 \sigma_0 & \sigma_0^2 \end{bmatrix}^{-1} \sim \mathcal{W}(\nu, \Xi^{-1}) \\ \delta &\equiv \text{vec}(\Delta) = \begin{bmatrix} \delta_1 \\ \delta_0 \end{bmatrix} \sim N(\mu_\delta, \Sigma \otimes A_\delta^{-1}) \end{aligned} \tag{C.1}$$

²⁰See Section 2.8.5 of [Rossi et al. \(2005\)](#) and Section 14.11 of [Chan et al. \(2020\)](#) for more details.

where $\mathcal{W}(\cdot, \cdot)$ denotes the Wishart distribution, ν is a non-negative scalar, Ξ is a 2-by-2 matrix, $\mu_\delta = [\mu'_{\delta_1}, \mu'_{\delta_0}]'$ is a $2P$ -by-1 vector and A_δ is a P -by- P matrix.²¹ We will also use the following P -by-2 matrix: $M_\delta \equiv [\mu_{\delta_1}, \mu_{\delta_0}]$.

C.2 Distributions of missing values, data augmentation and completion

Instead of (16) and (17) it now follows that:

$$\begin{aligned} \log Y_i^{miss}(1) \Big| D_i = 0, \log Y_i, \log \bar{B}_{CP,i}, \log b_i, X_i, \theta &\stackrel{d}{=} \\ \log Y_i^{miss}(1) \Big| D_i = 0, \log Y_i, X_i, \delta, \Sigma &\sim N \left(X_i' \delta_1 + \frac{\rho \sigma_1}{\sigma_0} (\log Y_i - X_i' \delta_0), (1 - \rho^2) \sigma_1^2 \right) \end{aligned} \quad (\text{C.2})$$

and

$$\begin{aligned} \log Y_i^{miss}(0) \Big| D_i = 1, \log Y_i, \log \bar{B}_{CP,i}, \log b_i, X_i, \theta &\stackrel{d}{=} \\ \log Y_i^{miss}(0) \Big| D_i = 1, \log Y_i, X_i, \delta, \Sigma &\sim N \left(X_i' \delta_0 + \frac{\rho \sigma_0}{\sigma_1} (\log Y_i - X_i' \delta_1), (1 - \rho^2) \sigma_0^2 \right), \end{aligned} \quad (\text{C.3})$$

while (15) and (24) remain the same. We can redefine δ_i^{miss} and $\sigma_i^{2,miss}$ as

$$\delta_i^{miss} = D_i \left(X_i' \delta_0 + \frac{\rho \sigma_0}{\sigma_1} (\log Y_i - X_i' \delta_1) \right) + (1 - D_i) \left(X_i' \delta_1 + \frac{\rho \sigma_1}{\sigma_0} (\log Y_i - X_i' \delta_0) \right) \quad (\text{C.4})$$

and

$$\sigma_i^{2,miss} = (1 - \rho^2) \left[D_i \sigma_0^2 + (1 - D_i) \sigma_1^2 \right], \quad (\text{C.5})$$

respectively, and combine them into

$$\log Y_i^{miss} \Big| \log Y_i, D_i, X_i, \delta, \Sigma \sim N \left(\delta_i^{miss}, \sigma_i^{2,miss} \right). \quad (\text{C.6})$$

The completion process in (21) remains unchanged.

²¹We maintain independent priors for the parameters associated with $\{Y(1), Y(0)\}$ and B_{CP} because of Assumption 1. Should this assumption be relaxed, we could then express (14) including δ_{CP} into δ and Δ and the same for σ_{CP}^2 and the correlations between $\log B_{CP}$ and $\log Y(1)$ and $\log Y(0)$ into the matrix Σ .

C.3 Drawing from posterior distribution

We once again condition on the “complete” data, \tilde{W}_t , and on the parameters of the prior distributions. In addition to the previously defined objects, we will also use the following N_t -by-2 matrix, $\log \tilde{Y}_{PO,t} \equiv [\log \tilde{Y}_t(1), \log \tilde{Y}_t(0)]$, as well as

$$\tilde{\Delta}_t = (X_t' X_t + A_\delta)^{-1} (X_t' \log \tilde{Y}_{PO,t} + A_\delta M_\delta) \quad (\text{C.7})$$

and

$$SSR_t = (\log \tilde{Y}_{PO,t} - X_t \tilde{\Delta}_t)' (\log \tilde{Y}_{PO,t} - X_t \tilde{\Delta}_t) + (\tilde{\Delta}_t - M_\delta)' A_\delta (\tilde{\Delta}_t - M_\delta). \quad (\text{C.8})$$

To draw new values for σ_{CP}^2 (for SPAs only) and δ_{CP} we still utilize expressions (A.1) and (A.2), respectively. However, instead of using these expressions to draw new values for Σ and δ , we now leverage the following results:

$$\Sigma^{-1,(q)} \Big| \theta^{(q-1)}, \theta_{\text{prior}}, \tilde{W}_t \stackrel{d}{=} \Sigma^{-1,(q)} \Big| \log \tilde{Y}_{PO,t}, X_t, \nu, \Xi, \mu_\delta, A_\delta \quad (\text{C.9})$$

and

$$\delta^{(q)} \Big| \Sigma^{(q)}, \sigma_{CP}^{2,(q)}, \delta^{(q-1)}, \delta_{CP}^{(q-1)}, \theta_{\text{prior}}, \tilde{W}_t \stackrel{d}{=} \delta^{(q)} \Big| \Sigma^{(q)}, \log \tilde{Y}_{PO,t}, X_t, \mu_\delta, A_\delta. \quad (\text{C.10})$$

For completeness, given the parametric assumptions we made it follows that:

$$\Sigma^{-1,(q)} \Big| \log \tilde{Y}_{PO,t}, X_t, \nu, \Xi, \mu_\delta, A_\delta \sim \mathcal{W} \left(\nu + N_t, (\Xi + SSR_t)^{-1} \right) \quad (\text{C.11})$$

and

$$\delta^{(q)} \Big| \Sigma^{(q)}, \log \tilde{Y}_{PO,t}, X_t, \mu_\delta, A_\delta \sim N \left(\text{vec}(\tilde{\Delta}_t), \Sigma^{(q)} \otimes (X_t' X_t + A_\delta)^{-1} \right). \quad (\text{C.12})$$

C.4 Summary

We summarize this adapted Gibbs sampling procedure below.

Algorithm 3: Gibbs sampling when $\rho \neq 0$

- 1 Set $\theta^{(0)}$ and θ_{prior} .
- for** $(q = 1, \dots, Q)$ **do**
- 2 Draw $\left\{ \log Y_i^{\text{miss},(q)}(1), \log Y_i^{\text{miss},(q)}(0), \log B_{CP,i}^{\text{miss},(q)} \right\}_{i=1}^{N_t}$ using (15), and (24) for FPA's, and (C.2)–(C.6).
- 3 Construct $\left\{ \log \tilde{Y}_i^{(q)}(1), \log \tilde{Y}_i^{(q)}(0), \log \tilde{B}_{CP,i}^{(q)} \right\}_{i=1}^{N_t}$ according to (21).
- 4 Draw $\left\{ \Sigma^{-1,(q)}, \delta^{(q)}, \sigma_{CP}^{-2,(q)}, \delta_{CP}^{(q)} \right\}$ according to (A.1)–(A.2) and (C.7)–(C.12).
- end**