Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Cold Start to Improve Market Thickness on Online Advertising Platforms: Data-Driven Algorithms and Field Experiments

Zikun Ye, Dennis J. Zhang, Heng Zhang, Renyu Zhang, Xin Chen, Zhiwei Xu

To cite this article:
Zikun Ye, Dennis J. Zhang, Heng Zhang, Renyu Zhang, Xin Chen, Zhiwei Xu (2023) Cold Start to Improve Market Thickness on Online Advertising Platforms: Data-Driven Algorithms and Field Experiments. Management Science 69(7):3838-3860. https://doi.org/10.1287/mnsc.2022.4550

https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 69, No. 7, July 2023, pp. 3838–3860
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Cold Start to Improve Market Thickness on Online Advertising Platforms: Data-Driven Algorithms and Field Experiments

Zikun Ye,[a] Dennis J. Zhang,[b] Heng Zhang,[c] Renyu Zhang,[d,*] Xin Chen,[e] Zhiwei Xu[f]

[a] Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801; [b] Olin Business School, Washington University in St. Louis, St. Louis, Missouri 63130; [c] W. P. Carey School of Business, Arizona State University, Tempe, Arizona 85287; [d] Department of Decision Sciences and Managerial Economics, CUHK Business School, The Chinese University of Hong Kong, Hong Kong, China; [e] H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30313; [f] Independent Contributor, Beijing, 100000, China
*Corresponding author

Contact: zikunye2@illinois.edu, https://orcid.org/0000-0001-9914-7966 (ZY); denniszhang@wustl.edu, https://orcid.org/0000-0002-4544-775X (DJZ); hengzhang24@asu.edu, https://orcid.org/0000-0002-6105-6994 (HZ); philipzhang@cuhk.edu.hk, https://orcid.org/0000-0003-0284-164X (RZ); xin.chen@isye.gatech.edu, https://orcid.org/0000-0002-5168-4823 (XC); rickyzhiwei@gmail.com (ZX)

Received: June 27, 2021
Revised: December 9, 2021
Accepted: February 22, 2022
Published Online in Articles in Advance: October 17, 2022

https://doi.org/10.1287/mnsc.2022.4550

Abstract. *Cold start* describes a commonly recognized challenge for online advertising platforms: with limited data, the machine learning system cannot accurately estimate the click-through rates (CTR) of new ads and, in turn, cannot efficiently price these new ads or match them with platform users. Traditional cold start algorithms often focus on improving the learning rates of CTR for new ads to improve short-term revenue, but unsuccessful cold start can prompt advertisers to leave the platform, decreasing the thickness of the ad marketplace. To address these issues, we build a data-driven optimization model that captures the essential trade-off between short-term revenue and long-term market thickness on the platform. Based on duality theory and bandit algorithms, we develop the shadow bidding with learning (SBL) algorithms with a provable regret upper bound of $O(T^{2/3}K^{1/3}(\log T)^{1/3}d^{1/2})$, where K is the number of ads and d captures the error magnitude of the underlying machine learning oracle for predicting CTR. Our proposed algorithms can be implemented in a real online advertising system with minimal adjustments. To demonstrate this practicality, we have collaborated with a large-scale video-sharing platform, conducting a novel, two-sided randomized field experiment to examine the effectiveness of our SBL algorithm. Our results show that the algorithm increased the cold start success rate by 61.62% while compromising short-term revenue by only 0.717%. Our algorithm has also boosted the platform's overall market thickness by 3.13% and its long-term advertising revenue by (at least) 5.35%. Our study bridges the gap between the theory of bandit algorithms and the practice of cold start in online advertising, highlighting the value of well-designed cold start algorithms for online advertising platforms.

History: Accepted by Gabriel Weintraub, revenue management and market analytics.
Supplemental Material: Data and the online appendices are available at https://doi.org/10.1287/mnsc.2022.4550.

Keywords: cold start problem • online advertising • contextual bandit • two-sided field experiment

1. Introduction

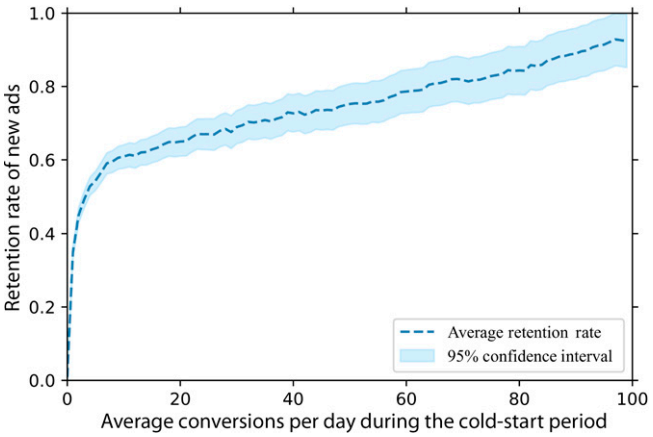
With the rapid growth of internet technology and smartphone penetration, online advertising has become an enormous industry, with a substantial impact on the entire economy. The Interactive Advertising Bureau reports that online advertising revenue in the United States increased to $124.6 billion in 2019 (16% year-over-year growth rate compared with 2018, 19% average annual growth rate since 2010), 70% of which comes from mobile advertising.[1] Facebook, TikTok, and other large online platforms monetize their gigantic user traffic primarily through online advertising. For example, in 2019, Facebook earned $69.7 billion revenue from advertising—98.53% of its total revenue.[2]

Online advertising platforms face a critical challenge called the *cold start problem* (see, e.g., Dave and Varma 2014, Choi et al. 2020). People have noted, both in the literature and in practice, that limited data history prevents online advertising platforms from accurately predicting the click-through rate (CTR) and the conversion rate (CVR) of new ads. Whereas most of the existing literature focuses on improving the statistical properties of cold start algorithms—improving the learning rates for CTR and CVR—to maximize the

short-run advertising revenue, we propose considering another important economic factor—*market thickness*—when designing cold start algorithms. Throughout this paper, we use the market thickness to represent the average number of ads competing for user impressions on an online advertising platform.[3] Specifically, we observe that in practice, throughout a whole ad campaign, advertisers especially value an ad's performance in the first few days—a bad performance with few conversions (i.e., app installs, purchases) may lead the advertiser to remove the ad from the platform. Therefore, it is crucial that platforms help new ads perform well and economically win advertisers' loyalty, maintaining the thickness of the ad pool in the auction (see Section 3 for details) while learning the CTR and CVR of these new ads during cold start. If the number of ad impressions remains the same,[4] a thicker market implies a higher revenue for the platform, with a decreasing marginal return. This is because with higher market thickness, on one hand, some user impressions that would otherwise be left unmatched can be matched with suitable ads and, on the other hand, the ads have more intensive competitions in the auctions on the platform. For an online business-to-business (B2B) auction market, Bimpikis et al. (2020) also empirically show that higher market thickness increases the platform's revenue.

To illustrate that the performance of new ads during cold start could fundamentally impact the long-term behavior of these ads, we collaborate with a large-scale online video-sharing platform (referred to as Platform O hereafter), and plot in Figure 1 the relationship between the number of conversions per day during new ads' first three days on the platform (the cold start period; on the x-axis) and the retention rate of these new ads in the subsequent two weeks on Platform O.[5] Here, we focus on the ad-level retention rate metric, the increase of which leads to higher aggregate-level market thickness in the long run, as long as the arrival rate of new ads remains stable. Two key observations emerge from Figure 1: (1) the long-term retention rate of a new ad is positively correlated with its performance during the cold start, and (2) such positive correlation is flattened when the number of conversions reaches a threshold around 10. In other words, for an ad platform to have enough market thickness and, in turn, high long-run revenue, quickly accumulating the first few conversions of each new ad is essential.[6] Not only is the ad retention dependent on the cold start, advertisers are also sensitive to whether their ads can obtain enough conversions during the cold start. On Platform O, advertisers will carefully monitor ad performance during the cold start; they may tighten the budget, reduce ad materials, or leave the platform if they are unsatisfied with the performance. Therefore, cold start performance will significantly impact the thickness of the ad marketplace.

Figure 1. (Color online) Retention Rate



On the other hand, to boost retention of new ads, platforms cannot simply provide more traffic to these new ads during cold start. This is because, as discussed earlier, the platform has less information about these new ads during cold start and, in turn, is less likely to efficiently match potential customers with these new ads. The inability to accurately predict CTR and CVR for new ads naturally brings up the exploration-exploitation trade-off between the short-term revenue generated by matching more mature ads (exploitation) and the long-term value from market thickness by matching more new ads (exploration). The fundamental trade-off in solving the cold start problem is to dynamically balance the long-term gain of successfully cold starting new ads and the short-term disutility from inefficiently matching new ads during cold start.

The main goal of this paper is to develop a new, theoretically sound and practically feasible end-to-end approach to solve the cold start problem when market thickness is important. To this end, we build a novel data-driven optimization model that integrates both the short-term revenue and the long-term cold start reward (defined as the long-term value from conversions during cold start to boost future market thickness) of an advertising platform. We develop a primal-dual-based multiarmed bandit (MAB) algorithm, denoted as the shadow bidding with learning (SBL) algorithm, which adaptively adds a shadow bid to each new ad's bidding price. Our proposed algorithm adeptly bridges theory and practice: it has a provable performance guarantee and it could be straightforwardly implemented on an online advertising platform with minimal adjustments. To demonstrate the practical value of our algorithm, we collaborated with Platform O to conduct a large-scale randomized field experiment to evaluate our algorithm.[7] Our results show that the proposed algorithm significantly increases both the cold start reward of new ads and the long-term total revenue of the entire platform. We summarize the main contributions of this paper as follows.

Optimization Model to Capture Cold Start and Market Thickness. Previous research on cold start in advertising has focused on improving CTR and CVR prediction accuracy and/or learning rates (Choi et al. 2020). We are the first in the literature to consider the economic aspect of cold start, and to explicitly model cold start as an important lever for improving market thickness. We formulate the cold start problem for online advertising platforms as a data-driven optimization model, synthesizing the linear program and MAB models in an innovative fashion. We believe our new modeling assumptions are critical, because identifying promising ads with high retention and keeping the ads market thick becomes notoriously challenging and important for not only the platform we work with but also other advertising platforms. Therefore, our modeling framework has the potential to empower other studies of the cold start problem for online advertising and recommender systems through an optimization lens.

End-to-End Solution to the Cold Start Problem for Online Advertising. To the best of our knowledge, we are the first in the literature to provide an end-to-end implementation of a new algorithm to address the cold start problem for online advertising platforms when market thickness is important. We develop a novel SBL algorithm by embedding a linear program primal-dual framework into an ϵ-greedy contextual bandit algorithm. Though theoretically compelling, existing algorithms for general contextual bandits with concave objectives (e.g., Agarwal et al. 2014, 2016) are practically infeasible on real-world online advertising platforms. This is because these algorithms rely on an underlying argmax oracle (AMO), which is unavailable or computationally intractable in practice. Our proposed bandit algorithm bridges the gap between the learning theory and advertising cold start practice with a provable performance guarantee and straightforward implementation on real online advertising platforms. The algorithm leverages the dual variables of the cold start reward constraints, the power of the advertising platform's underlying machine learning system to predict CTR and CVR, and an ϵ-greedy exploration scheme, thus yielding a provable regret bound of $O(T^{2/3}K^{1/3}(\log T)^{1/3}d^{1/2})$, where K is the number of new ads and d is the prediction error term of the underlying machine learning oracle for predicting CTR and CVR. This term characterizes the difficulty of the CTR prediction problem with the underlying machine learning (ML) oracle. The smaller the d, the simpler the CTR prediction problem, and, therefore, the more powerful the ML models to predict CTR. We also incorporate the dual mirror descent method into the SBL algorithm, which reduces its computational complexity without compromising the regret bound. Another compelling advantage of our SBL algorithm is that it enables us to minimally adjust the bidding system of an online advertising platform by simply adding a shadow bid for each ad (i.e., the dual variable of the cold start reward constraint) to its real-time bidding price.

Experimental Evaluation of Our Algorithm. We are the first in the literature to conduct two-sided randomized field experiments for bandit algorithms. In a general advertising cold start setting, the traditional one-sided experiment is invalidated by the violation of the stable unit treatment value assumption (SUTVA) (see Section 5 for more discussion on this point) and, therefore, gives rise to estimation biases as high as 120%. Such violation of SUTVA is common in the experimental evaluation of algorithms and policies on e-commerce (e.g., Facebook Marketplace; see Ha-Thuc et al. 2020) and vacation-rentals (e.g., Airbnb; see Johari et al. 2022) platforms and has caused substantially biased estimations for experiments thereof. To address such a challenge, we design and implement a novel two-sided field experiment on Platform O. Under mild assumptions, the experiment restores SUTVA and enables us to causally estimate the value of our proposed algorithm in an unbiased fashion. The new experiment framework could be applied to evaluate other algorithms and policies of recommender systems on two-sided platforms. Based on our two-sided field experiments, we find that the proposed algorithm successfully increases the cold start success rate by 61.62%. Our experiment also demonstrates that the SBL algorithm increases the average retention time of the ads and, thus, market thickness by 3.13%. Moreover, we conduct comprehensive simulation studies that show that our algorithm boosts the total (long-term) advertising revenue of the entire platform by (at least) 5.35% if the advertisers' behaviors remain the same as what we observed in the experiment even when the SBL algorithm is applied to all ads and user impressions in the long run. Such increase translates to hundreds of millions of U.S. dollars revenue boost per year for Platform O. In short, the two-sided experiments enable us to demonstrate that the SBL algorithm substantially improves the long-term revenue of an advertising platform.

In short, our study bridges the gap between the theory of bandit algorithms and the practice of cold start in online advertising, highlighting the significant value of well-designed cold start algorithms for online advertising platforms. The rest of this paper is organized as follows. In Section 2, we position our paper in the relevant literature. Section 3 discusses the business practices on which we base our model. In Section 4, we propose our algorithms and analyze the regret bound. In Section 5, we introduce our field experiment setting. Section 6 reports our experimental results. Section 7 concludes. All proofs are relegated to the online appendices.

2. Literature Review

Our paper is primarily related to three streams of literature: cold start for online advertising, bandit algorithms, and field experiments on large-scale online platforms.

Estimating the CTR of new ads is a challenging problem, because there is very little data and information to provide reliable prediction (see, e.g., Dave and Varma 2014, Choi et al. 2020). The sophisticated deep learning models developed in recent years are designed to better estimate the CTR of cold start items. For example, Zhou et al. (2018) propose the deep interest network, which incorporates data on users' historical behavior and interests to learn the CTR. Vartak et al. (2017) propose a meta-learning strategy to address the cold start problem when new items arrive continuously. In contrast, our proposed SBL algorithm avoids any extra data or different neural network architectures. Instead, we employ the ϵ-greedy random-exploration scheme and shadow bids to feed more data from new ads into the neural networks, which also substantially increases the accuracy of CTR/CVR estimations.

We study the cold start problem as a data-driven optimization model and offer efficient algorithms to tackle this challenge. Viewing the problem as ad allocation in the repeated-auction setting is in line with another stream of literature in operations management (see, e.g., Caldentey and Vulcano 2007; Balseiro et al. 2014, 2015; Hojjat et al. 2017; Balseiro and Gur 2019). In particular, Balseiro et al. (2014) adopt a dual-based bid-price control policy to study the ad allocation problem in the presence of the trade-off between short-term revenue and long-term value from delivering good spots to the (contracted) reservation ads. Whereas most of the literature on ad allocation assumes the CTR is known to the decision maker, we study a more realistic contextual bandit setting where the true CTR is unknown and is predicted by an underlying machine learning system.

Our algorithm is closely related to the literature on the stochastic contextual bandit. We compare our algorithm's properties with existing contextual bandit algorithms in Table 1, where the settings consistent with the practice of Platform O are marked in bold. At a high level, contextual bandit algorithms can be categorized into two different classes (see Simchi-Levi and Xu 2021): (1) agnostic approaches, which are model-free but require a prespecified policy set and optimization oracles; and (2) realizability-based approaches, which explicitly specify the underlying model to represent the reward as a function of contexts. As Simchi-Levi and Xu (2021, p. 1904) observe "Although many different contextual bandit algorithms (realizability-based or agnostic) have been proposed over the last 20 years, most of them have either theoretical or practical issues." There is still substantial room for improvement in this literature. It is useful to differentiate our algorithm from existing ones with both agnostic and realizability-based approaches.

The agnostic approaches for contextual bandits (e.g., Dudik et al. 2011, Agarwal et al. 2014, Agrawal et al. 2016) usually adopt a conservative exploration scheme (Bietti et al. 2021), based on an AMO and a policy set. Take, for example, the algorithm proposed by Agarwal et al. (2014) with a $\tilde{O}(\sqrt{KT\log(|\Pi|)})$ regret bound. They assume that, given the policy set Π and the set $\mathcal{S}$ of context-reward pairs $(x,r) \in X \times \mathbb{R}^K$, the AMO returns the loss-minimization policy $\pi^* = \arg\max_{\pi\in\Pi}\sum_{(x,r)\in\mathcal{S}} r(\pi(x))$. The reason to assume this oracle as a subroutine is that it is generally impractical to optimize the loss by enumerating over Π. In a practical setting such as the problem we study, this oracle is clearly infeasible. First, Platform O leverages the deep neural network, a very large policy set in which $|\Pi|$ is on the magnitude of trillions. Specifically, if the policy set Π is the collection of neural networks with fixed structure, depth, and width, even under the proper parameter discretization, its cardinality $|\Pi|$ grows exponentially with the number of parameters. In fact, without a proper realizability assumption, the AMO is computationally intractable in practice. Even if we assume the underlying data-generating process is a neural network, we are not aware of an efficient AMO for finding the optimal policy. Second, the algorithm of Agarwal et al. (2014) computes the empirical regret of a policy via the inverse propensity score (IPS) at each epoch. The IPS method gives an unbiased reward estimate of a policy and is, thus, widely used in regret analysis. However, IPS suffers from a high variance when the policy set is large and/or the past sample paths vary significantly, which is indeed the case of our implementation on Platform O. In fact, all the agnostic approaches in the contextual bandit literature suffer from the aforementioned two issues and are therefore not applicable in our context.

The core idea of our algorithm is similar to the realizability-based approaches. Some realizability-based algorithms leverage the upper confidence bound (UCB) and Thompson sampling exploration schemes, which are only tractable for reward functions parametrized in a certain way, such as linear models (Chu et al. 2011) and deep neural networks (Zhou et al. 2020). Foster et al. (2018) adopt a least-squares regression oracle for their realizability-based algorithm, which is amenable to widely used gradient-based training methods. Empirically, this algorithm works well among existing contextual bandit approaches, but it is theoretically suboptimal. In the realizability-based setting with an offline regression oracle to predict the reward, Simchi-Levi and Xu (2021) provide the first optimal black-box reduction (i.e.,

Table 1. Algorithm's Performance in the Contextual Bandit Setting

Algorithm	Bandit setting	Regret	Computational complexity
LinearUCB (Agrawal and Devanur 2014)	Linear context **Knapsack**	Optimal	Calls to offline linear regression at each round
NeuralUCB (Zhou et al. 2020)	**Neural network context** Nonknapsack	Optimal	Gradient-descent-based update of the predictor at each round
Regressor elimination (Agarwal et al. 2012)	**Realizability-based** Nonknapsack	Optimal	$\Omega(\lvert\Pi\rvert)$ intractable
ILOVETOCONBANDITS (Agarwal et al. 2014)	Agnostic Nonknapsack	Optimal	$\tilde{O}(\sqrt{KT/\log\lvert\Pi\rvert})$ calls to AMO
Algorithm adapted from ILOVETOCONBANDITS (Agarwal et al. 2016)	Agnostic **Knapsack**	Optimal	$\tilde{O}(K\sqrt{KT\log\lvert\Pi\rvert})$ calls to AMO
RegCB (Foster et al. 2018)	Realizability-based Nonknapsack	Suboptimal	$O(T^{3/2})$ calls to an offline regression oracle
FALCON (Simchi-Levi and Xu 2021)	Realizability-based Nonknapsack	Optimal	$O(\log T)$ calls to an offline regression oracle
SBL-RS/SBL-DMD (this paper)	**Neural network context** **Knapsack**	Suboptimal	Gradient-descent-based update of the predictor, $O(T^{1/3})$ calls to solving dual or dual mirror descent

Note. Boldface type indicates the same setting as our problem.

achieving the $\tilde{O}(\sqrt{T})$ theoretical lower bound) from a contextual bandit to an offline regression. The key to this reduction is a special exploration-exploitation scheme implicitly aligned with the agnostic approach proposed by Agarwal et al. (2014). This reduction works only for an objective function linear in the accumulated reward, so it is not applicable to our setting of contextual bandits with a concave objective function. It remains an open problem whether such a reduction that matches the $\tilde{O}(\sqrt{T})$ theoretical lower bound exists for contextual bandits with concave objectives (e.g., Agrawal and Devanur 2014, Agrawal et al. 2016). In this paper, we integrate a machine learning oracle into a contextual bandit model with a concave objective function, and we develop dual-based algorithms that achieve a sublinear regret. Moreover, implementing these known approaches in the literature requires substantial engineering effort, such as a complex sampling scheme for known policies, whereas our method is compatible with the existing advertising system on Platform O.

Besides the literature on contextual bandit algorithms, our work is also closely related to the growing literature on solving operations management problems with online learning. Given the intrinsically uncertain business environment, a recent trend is to combine learning theory and optimization to solve revenue management and inventory control problems. For example, Chen et al. (2019) build an algorithm to solve the joint problem of pricing and inventory control with nonparametric demand learning for nonperishable products, and they show the regret convergence result. Nambiar et al. (2019) propose an algorithm with theoretical performance guarantees to solve the dynamic pricing problem with misspecified demand models; they evaluate its performance using offline simulations. Ferreira et al. (2018) use Thompson sampling to learn the demand at each price and solve the network revenue management problem. Chen et al. (2020) build an online learning algorithm to solve the single-item inventory control problem under the periodic review, backlogging policy with unknown capacity and demand distributions. Chen and Gallego (2022) propose a primal-dual learning algorithm to learn the dual optimal solution for the personalized dynamic pricing problem with an inventory constraint. Bastani et al. (2022) propose a meta-dynamic pricing algorithm to learn the prior through experiment while solving the pricing problem. Golrezaei et al. (2019) propose learning algorithms to set reserve prices in contextual auctions. Our main contribution to this strand of literature is that we have not only proved the theoretical performance guarantee of the proposed algorithm but also implemented it on a large-scale advertising platform and tested its performance using field experiments.

Last but not least, our paper directly relates to the growing literature on field experiments on online platforms (Terwiesch et al. 2020). For example, Zhang et al. (2020) document the spillover effects across platform users in a field experiment on a retailing platform. Zeng et al. (2021) show that social nudge can boost the productivity of content providers on a social network platform through randomized field experiments. Fisher et al. (2018) leverage both modeling and field experiments to study competition-based dynamic pricing in retailing. Several other papers in the literature conduct field experiments to study platform

operations problems (e.g., Cui et al. 2019, Cui et al. 2020, Feldman et al. 2021). In the marketing literature, Schwartz et al. (2017) implement a learning algorithm to optimize the user-acquisition strategy through display advertising and conduct a field experiment to gauge its effectiveness. A growing body of literature examines the violation of SUTVA for experiments on large-scale platforms. Ha-Thuc et al. (2020) develop a new counterfactual framework for seller-side A/B testing on Facebook Marketplace and show that the new experiment framework satisfies the SUTVA. Johari et al. (2022) propose a mean-field model to show that single-sided experiments (demand-side or supply-side randomization) will result in biases in estimation for a two-sided marketplace such as Airbnb. They also propose a two-sided randomization and the associated estimator, which is unbiased when the supply and demand are extremely imbalanced. Other authors use clustering algorithms to reduce the impact of interference in experiments on networks. Rolnick et al. (2019) propose a geographical clustering algorithm (referred to as the GeoCUTS algorithm) that minimizes the interference between different geographical units while preserving the balance in cluster size. Pouget-Abadie et al. (2019) introduce a novel clustering objective and a corresponding algorithm that partitions a bipartite graph so as to maximize the statistical power of a bipartite experiment on that graph. Liu et al. (2021) document a significant cannibalization bias of one-sided A/B tests on an online ad marketplace and propose a budget-split experiment design to de-bias the estimates. Our contribution toward this stream of research is the design and implementation of a novel, two-sided randomized field experiment to causally estimate the value of our proposed bandit algorithm. The proposed experiment framework could potentially be applied to evaluating other algorithms and policies in general recommender systems of two-sided online platforms.

3. Background and Model

In this section, we first introduce the background setting of a typical demand-side platform (DSP) for online advertising, with a particular focus on its cold start problem. Based on institutional knowledge, we then develop a data-driven optimization model that integrates linear programs and multiarmed bandits to tackle the issue.

3.1. Online Advertising Platforms

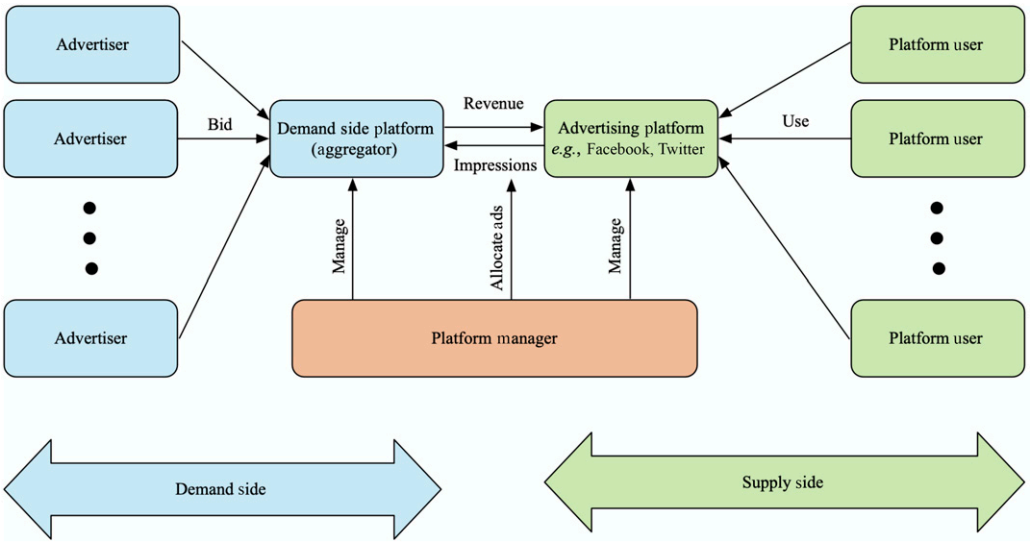
Large-scale online platforms such as Facebook and TikTok are usually equipped with a DSP, a centralized advertising system that aggregates online ads and efficiently matches the ads with users. Figure 2 summarizes the landscape of a DSP. Advertisers and platform users interact with each other on the DSP. On the demand side, advertisers set up their advertising campaigns by submitting the necessary information to the DSP: bid prices, billing options, ad content, advertising budget, and the users they wish to target. On the supply-side, platform users are exposed to ads while viewing organic content. The DSP plays a central role in allocating user impressions to different ads, with a goal of maximizing long-term revenue.

Next, we show how the DSP monetizes its user traffic. Ad impression requests from platform users continuously arrive at the DSP. For the rest of this paper, we use ad impression, user view, and user impression interchangeably. For large-scale online platforms, the DSP allocates billions of impressions to hundreds of thousands of ads each day. The decision is typically based on a large-scale auction, where hundreds of ads compete to win an ad impression based on advertisers' bids, predicted click-through rates (pCTR), and predicted conversion rates (pCVR). The user impression is allocated to the ad with the highest estimated cost per mille (eCPM) of the match between the impression and the ad, which measures the expected revenue of displaying the ad to the respective platform user a thousand times. This rule ensures that each ad impression generates the highest ex ante revenue in expectation.

Advertisers can choose from among several billing options, depending on what they wish to bid for each impression (e.g., clicks or conversions) and how the advertising fee is charged (e.g., by impressions, clicks, or conversions). Under all billing options, the bids, pCTRs, and pCVRs can be effectively converted to eCPM, based on which ads can be ranked under the same scale. See Online Appendix H for a detailed description of different auction mechanisms and billing options.

Cold Start on a DSP. The inability to accurately predict the CTR and CVR of new ads makes cold start one of the key challenges faced by platforms and advertisers alike. It is extremely difficult to strike a smart balance between boosting new ads that have great potential to enhance the long-term thickness of the platform and maximizing the short-term revenue generated by high-quality mature ads. To the best of our knowledge, most DSPs tackle the cold start problem ad hoc. For example, to increase its cold start success rate, Platform O has adopted a bid-controlling system (called the PID system; see Appendix H in the online appendix) to uniformly increase the system bidding prices for all new ads within a very short time until the preset upper bounds of the system bidding prices are met. This approach increases the probability of winning impressions for new ads, resulting in more exploration of the ads and potentially more

Figure 2. (Color online) Online Advertising on Platforms



conversions. Soon after such sharp increases in the system bidding prices, this system will adaptively lower these prices to offset the extra costs caused by the uniform bid increases. To our knowledge, this heuristic approach has no performance guarantee—fine-tuning hyperparameters is the only leverage. In the following, we formulate the cold start problem as a data-driven optimization problem, design a contextual bandit algorithm with a provable performance guarantee to address this problem, and conduct a two-sided experiment to evaluate the proposed algorithm.

3.2. The Cold Start Model

We formulate the cold start problem of a DSP as a data-driven optimization model. To highlight the key trade-off associated with the problem and avoid unnecessary complexity, we make two high-level modeling assumptions. First, the ad allocation mechanism is a first-price auction, where all advertisers bid on clicks, so they are charged once their ad is clicked. Without loss of generality and for ease of exposition, we assume CVR = pCVR = 1, that is, conversion is guaranteed upon click-through. (Later, we will show that our proposed algorithm can be easily implemented on a real DSP.) The first-price auction is more intuitive for advertisers, so there is a recent trend of switching from second-price auctions to first-price auctions in the online advertising industry. For example, Google Ad Manager moved to first-price auctions in 2019.[8] Second, the real-time system bidding price of each ad remains the same as the bid submitted by the advertiser. In the online implementation of our proposed algorithm, the model is adapted to incorporate the actual online auction mechanism and the real-time system bidding prices of the DSP we experiment on.

We consider a DSP where a set of K new ads, denoted as $A := [K] = \{1, 2, \ldots, K\}$,[9] are competing for user impressions. We only consider new ads in our base model; in Section 5.2, we discuss how our algorithm can be generalized to a setting with both new and mature ads. User impressions arrive sequentially at the DSP. We define the set of all user impressions as $[T] = \{1, 2, \ldots, T\}$. For each user impression t and ad j, there is an associated context/feature vector $x_{t,j}$. The context $x_{t,j}$ could be quite broad, containing the demographic and behavioral information of user impression t inherited from the platform, and ad information from ad j. Upon the arrival of user impression t, the DSP observes K feature vectors $x_{t,j}$, $j \in A$. For ease of exposition, we define the vector $x_t := (x_{t,1}, x_{t,2}, \ldots, x_{t,K}) \in X$, where X is a countable feature space. Suppose that ad $a_t \in A$ is chosen to be displayed. We can define a K-dimensional binary vector $v_t(a_t) \in \{0,1\}^K$ representing whether each ad is clicked. More specifically, the jth component of the vector $v_{t,j}(a_t) = 1$ only if $a_t = j$ and ad j is clicked by user t. Furthermore, we assume a stochastic contextual bandit setting, that is, the set of context vectors and the click-through vector $(x_t, \{v_t(a)\}_{a \in K})$ for $t \in [T]$ is drawn independent and identically distributed (i.i.d.) from a distribution $\mathcal{D}$ over $X \times \{0,1\}^K$, which is unknown to the DSP. And, we can observe the partial outcome $v_t(a_t)$ only at round t of the played ad a_t. Throughout this paper, we use the subscript i to denote the context index of the countable feature space X. We denote the marginal distribution of $\mathcal{D}$ over the context as $\mathcal{D}_X$, that is, for round $t \in [T]$, the context type i is drawn i.i.d. as $i \sim \mathcal{D}_X$. Given the context

information for round t and ad j, we define $c_{t,j} := \mathbb{E}[v_{t,j}(a_t) \mid a_t = j]$ as the CTR of ad j at round t. We sometimes also abuse the notation, denoting by c_{ij} the CTR of ad j under the context type i.

A core challenge faced by Platform O and other DSPs is jointly optimizing the revenue and the cold start reward of new ads. To evaluate revenue during cold start, we define $V := \sum_{t=1}^{T} v_t(a_t)$ as the accumulative click-through vector, where $V_j := \sum_{t=1}^{T} v_{t,j} \in \{0, 1, \ldots, T\}$ is the total number of clicks generated by ad j until customer T. As prescribed by the oCPC billing option (see Online Appendix H for details), the total revenue generated by the ads is given by

$$\sum_{j=1}^{K} b_j V_j,$$

where $b_j \in [0,1]$ is the bid (per click) of ad $j \in A$. To quantify the cold start reward, one may want to directly estimate the total lifetime revenue from an ad based on the number of accumulated conversions during its cold start period (the first three days for Platform O). However, such an estimation is extremely difficult, if not impossible, because we need to establish the causal effect of conversions during the cold start period on the new ad's lifetime revenue. Therefore, we take an alternative approach to approximate the aforementioned relationship between conversions in the cold start period and lifetime revenue. We observe from Figure 1 that an ad's retention rate increases linearly in the number of clicks/conversions while this number is below a certain threshold; it stays (almost) unchanged once it exceeds the threshold. Motivated by this phenomenon, we assume the cold start reward of each conversion for ad j before the number of accumulated conversions reaching the conversion target as $\beta_j \in (0,1]$. Without loss of generality, we denote the conversion target as αT, where $\alpha \in (0,1)$. Thus, the cold start reward is given by

$$\sum_{j=1}^{K} \beta_j \min\{V_j, \alpha T\}. \tag{1}$$

In practice, the conversion target αT is determined by business practice and validated by our observation in Figure 1. We specify the cold start reward per conversion β_j via two steps: (1) inherit the business practice of Platform O that $\beta_j = 2b_j$ for each ad j, and (2) conduct simulations to validate the choice of β. Our simulation results, in Online Appendix D, demonstrate that setting $\beta_j = 2b_j$ for each ad j would significantly increase the expected long-term revenue for the platform. Furthermore, our two-sided experiment shows that such a choice of β_j boosts the long-term advertising revenue of Platform O by 5.35%.

We are now ready to present the objective of the DSP for cold start, which equals the sum of revenue and cold start reward:

$$\begin{aligned}\Gamma(V) &:= \sum_{j=1}^{K} b_j V_j + \sum_{j=1}^{K} \beta_j \min\{V_j, \alpha T\} \\ &= \sum_{j=1}^{K} b_j \sum_{t=1}^{T} v_{t,j} + \sum_{j=1}^{K} \beta_j \min\left\{\sum_{t=1}^{T} v_{t,j}, \alpha T\right\},\end{aligned} \tag{2}$$

which is piecewise linear and concavely increasing in the number of conversions for each ad j.

3.3. Definition of Regret

In this subsection, we formally define the benchmark for our proposed bandit algorithms. In every round $t \in [T]$, a policy π observes the feature vector $x_t \in X$, chooses an ad/action $a_t \in A$, and observes the random outcome whether the ad is clicked. We define the history update to round t as $\mathcal{H}_t = \cup_{s=1,\ldots,t-1}\{(x_s, a_s, v_s(a_s))\}$. Let $\Delta_A := \{y \in \mathbb{R}^{|A|} : y_j \geq 0,\ \forall j \in A, \sum_{j \in A} y_j \leq 1\}$ be the set of the nonnegative weight/distribution over arms. Formally, a policy π defines a mapping from the history $\mathcal{H}_t$ and the context x_t to the set of distribution over ads Δ_A for any t.

Recall that one can express the expected reward we gain from policy π as $\mathbb{E}_{\mathcal{D}^T,\pi}[\Gamma(V)]$, where $\mathcal{D}^T$ refers to T independent copies of the distribution $\mathcal{D}$. We notice that $\Gamma(\cdot)$ is concave in V. Using Jensen's inequality, one can show that the following lemma holds.

Lemma 1. *For any policy π, the scaled expected reward can be upper-bounded as*

$$\begin{aligned}&\frac{1}{T} \cdot \mathbb{E}_{\mathcal{D}^T,\pi}[\Gamma(V)] \leq \mathsf{OPT} \\ &:= \max_{y_i \in \Delta_A, \forall i} \left\{\sum_{j=1}^{K} \mathbb{E}_{i \sim \mathcal{D}_X}[c_{ij} y_{ij} b_j] + \sum_{j=1}^{K} \beta_j \min\{\mathbb{E}_{i \sim \mathcal{D}_X}[c_{ij} y_{ij}], \alpha\}\right\}.\end{aligned}$$

Essentially, $T \cdot \mathsf{OPT}$ is the upper bound of the cold start objective function $\Gamma(\cdot)$; it can be viewed as the solution to a fluid version of our cold start problem, in which the decision variable $y_i \in \Delta_A$ is a sampling distribution over all the ads in A given user context i. Similar upper bounds are widely used in the revenue management literature (Gallego and Van Ryzin 1994, Golrezaei et al. 2014, Zhang et al. 2018), as well as the bandit learning literature (Agrawal et al. 2016, Badanidiyuru et al. 2018). With Lemma 1, one can formally define the regret for an arbitrary policy π as

$$\mathsf{Reg}(\pi) = T \cdot \mathsf{OPT} - \mathbb{E}_{\mathcal{D}^T,\pi}[\Gamma(V)]. \tag{3}$$

Our goal is to propose a novel policy that has a provably optimal performance guarantee (measured by a sublinear regret) and that can be effectively implemented on a practical DSP. As mentioned in Section 2, although the regret defined similar to (3) is common in the stochastic bandit setting with a concave objective

function (see, e.g., Agarwal et al. 2014, 2016), the existing bandit algorithms are not practically feasible in our setting, for two reasons. First, these algorithms are often built upon the AMO, which is computationally intractable for most policy classes. Furthermore, a practical DSP often relies on a machine learning system to generalize the knowledge learned from the data on observed click-throughs and conversions and make accurate predictions about future user behaviors. How to design efficient algorithms for realizability-based bandits with deep learning models and concave objectives remains an open question in the literature. Second, existing MAB algorithms usually provide an empirical estimate of OPT based on IPS (e.g., Agrawal et al. 2016). In practice, such an IPS technique may suffer from a high variance. As such, we design a novel primal-dual-based algorithm by leveraging the predictions of the machine learning system of a DSP as model inputs and adding "shadow bids" to new ads. The ad allocation policy can be easily implemented by the auction system of a real-world DSP, by adjusting the bidding prices of new ads. As we show in Section 5, one can easily implement this shadow bidding with learning algorithm on a real-world DSP, and our field experiments show significant improvements in long-term retention and revenue without much compromising short-term revenue.

4. Cold Start Algorithms

In this section, we propose novel bandit learning algorithms for our cold start problem. Our algorithms leverage the ϵ-greedy exploration strategy, the prediction power of a DSP's underlying machine learning system, and the empirically optimal dual solution to the fluid upper bound.

4.1. Shadow Bidding with Learning (SBL) Algorithms

In this subsection, we outline our primal-dual-based learning algorithm. One central difficulty is the unknown distributional information of the underlying model. In particular, the ground-truth CTRs at round t are unknown to any algorithm. Instead, we have access to an empirically estimated CTR only via the online training of predicting models on the historical data. To get an empirically optimal ad allocation policy, one can solve the following ad allocation model at round t:

$$\max_{y_i\in\Delta_A,\,\forall i\in\mathcal{I}} \sum_{i\in\mathcal{I}}\sum_{j\in A}\hat{p}_i^t\hat{c}_{ij}^t b_j y_{ij} + \sum_{j\in A}\beta_j \min\left\{\sum_{i\in\mathcal{I}}\hat{p}_i^t\hat{c}_{ij}^t y_{ij}, \alpha\right\}. \tag{4}$$

Recall that the number of contexts is countable, so we define $\mathcal{I} := \{1,2,3,\dots\}$ as the set of context types. We denote p_i as the probability that incoming context is type i. In our model and analysis, p_i is unknown in prior and is estimated using the empirical estimation $\hat{p}_i^t$ based on the historical data $\mathcal{H}_t$. Here, the empirical CTR $\hat{c}_{ij}^t$ is the estimated CTR at time t under the context i, and ad j trained on the historical data $\mathcal{H}_t$. In practice, $\hat{c}_{ij}^t$ is usually produced by a deep neural network associated with the DSP to predict the CTR/CVR of the ads facing different user contexts. By introducing an additional variable u_j for each ad j, we can transform (4) to a linear program:

$$\begin{aligned}\max_{y,u\geq 0}\ & \sum_{i\in\mathcal{I}}\sum_{j\in A}\hat{p}_i^t\hat{c}_{ij}^t b_j y_{ij} + \sum_{j\in A}\beta_j(\alpha - u_j)\\ \text{s.t.}\ & \sum_{j\in A} y_{ij}\leq 1,\ \forall i\in\mathcal{I},\ \sum_{i\in\mathcal{I}}\hat{p}_i^t\hat{c}_{ij}^t y_{ij} + u_j \geq \alpha,\ \forall j\in A\end{aligned} \tag{5}$$

We succinctly write the dual of (5) as

$$\begin{aligned}\min\ & \sum_{i\in\mathcal{I}}\hat{p}_i^t \max_{j\in A}\{\hat{c}_{ij}^t(b_j+\lambda_j)\} + \alpha\sum_{j\in A}(\beta_j - \lambda_j)\\ \text{s.t.}\ & 0\leq\lambda_j\leq\beta_j,\quad \forall j\in A,\end{aligned} \tag{6}$$

which is a nonsmooth convex program with decision variables $\lambda = (\lambda_1,\lambda_2,\dots,\lambda_K)$. Strong duality dictates that the optimal values of (5) and (6) must be the same. Utilizing such duality, we propose the following cold start algorithm.

Algorithm 1 (Shadow Bidding with Learning and Resolving (SBL-RS))

Parameters: Epoch schedule $1=\tau_1<\tau_2<\dots$ such that $\tau_m-\tau_{m-1}=\tau_{m+1}-\tau_m\leq O(T^{2/3})$. Cold start reward coefficient β. Target conversion parameter α.

Initialization: $\lambda^1\in[0,\beta]^K$, $m\leftarrow 1$.

For $t=1,2,\dots,T$, **do**

Step 1: Observe the context i_t at period t. With probability $\epsilon_t = t^{-1/3}(K\log t)^{1/3}$, the algorithm picks an ad uniformly at random. Otherwise, display an ad $a_t\in\arg\max_j \hat{c}_{i_tj}^t(b_j+\lambda_j^{\tau_m})$ with arbitrary tie-breaking rules.

Step 2: If $t=\tau_m$, we solve the dual model (6) to optimality to update λ^{τ_m} and set $m\leftarrow m+1$.

Step 3: Observe the outcome of a_t, and update the parameters of the underlying machine learning model for predicting $\hat{c}_{ij}^{t+1}$.

Several remarks are in order. First, we highlight a compelling advantage of the SBL-RS algorithm: it fits perfectly into the auction system of a real-life advertising platform, fully leveraging the predictive power of the embedded machine learning oracle to estimate the CTR of ads. This makes our algorithm generalizable and implementable for any large-scale DSP. In particular, we periodically resolve the optimization problem (6) to produce the dual vector λ^{τ_m}. With the most recent λ^{τ_m}, and the most up-to-date CTR estimation given the context, $\hat{c}_{ij}^t$, the SBL-RS algorithm picks ad j

with the highest adjusted eCPM, $\hat{c}_{ij}^t(b_j + \lambda_j^{\tau_m})$ (ties broken arbitrarily), which can be easily implemented in practice by adding $\lambda_j^{\tau_m}$ to the bidding price of ad j in the auction system (see the oSBL algorithm presented in Section 5). Solving the dual problem with a carefully chosen epoch will save computing resources without hurting algorithm performance.

Second, the term shadow bidding comes from the ad-selection rule. We pick the ad with the largest adjusted eCPM upon the arrival of each user, which is the sum of the bid price, b_j, and the shadow price for the cold start reward, $\lambda_j^{\tau_m}$, multiplied by the predicted CTR. The original bidding process of the DSP seeks only to maximize the short-term revenue by picking ad j with the highest eCPM $= \hat{c}_{ij}^t b_j$, whereas we add $\lambda_j^{\tau_m}$, the shadow price associated with the constraint $\sum_{i\in\mathcal{I}} \hat{p}_i^t \hat{c}_{ij}^t y_{ij} + u_j \geq \alpha$ in (5), to the bid b_j in order to capture the long-term cold start reward of displaying ad j immediately. In effect, we use the solution in the dual space to characterize the correct assignment in the primal space, which gives a fast and simple allocation rule. Similar dual-based strategies are used in the online linear program under stochastic input or random permutation (Li et al. 2020), and the noncontextual knapsack bandit setting (Badanidiyuru et al. 2018). One may also wonder whether existing algorithms that directly solve the primal problem (e.g., Agarwal et al. 2014, 2016) would also work in our cold start setting. In fact, though theoretically possible, directly implementing the primal solutions on a practical DSP is very hard, if not impossible. Specifically, solving the primal problem amounts to dictating an ad-assignment scheme. However, the primal space of the problem is extremely large—its solution has a cardinality of the number of impressions multiplied by the number of ads, which is on the order of trillions. The cardinality of the associated dual space, however, is the number of ads, which is on the order of hundreds of thousands. Therefore, in practice, working with the dual space is substantially simpler than working with the primal space.

Third, in each round t, we explore new ads with probability $t^{-1/3}(K \log t)^{1/3}$, and exploit, with probability $1 - t^{-1/3}(K \log t)^{1/3}$, following the dual-based policy (6). This exploration-exploitation schedule is common for ϵ-greedy algorithms in the bandit learning literature. The novelty of our algorithm lies in the well-designed, dual-based exploitation scheme and the integration of an MAB algorithm with the underlying machine learning oracle of a DSP. One may also consider other exploration-exploitation strategies such as upper confidence bound or Thompson sampling. We leave it to future researchers to study the optimal exploration scheme for our cold start problem. Moreover, an ϵ-greedy based algorithm (such as SBL-RS) can be naturally embedded into a DSP in practice without much engineering change, as we will show in Section 5.

Although the SBL-RS algorithm achieves a sublinear regret bound and is therefore asymptotically optimal (see Theorem 1(a)), it needs to solve the dual program (6) $O(T^{\frac{1}{3}})$ times which may be computationally costly. Furthermore, the theoretical guarantee in Theorem 1(a) holds if the dual program (6) is solved to optimality. Although achieving the exact optimality is generally hard, there are polynomial-time algorithms to achieve an arbitrarily small optimality gap ϵ. For example, the subgradient descent method with box constraint projection has an $O(\epsilon^{-2})$ computational complexity for our Lipschitz continuous convex object function (6) (see, e.g., theorem 3.2.2 in Nesterov 2014). However, this approach is computationally inefficient for a large-scale auction in practice, because it is costly to obtain the subgradient in this setting. Such computational complexity also motivates us to consider a variant of our SBL-RS algorithm, which incorporates the dual mirror descent (DMD) optimization method (e.g., Balseiro et al. 2022) into our SBL algorithmic framework. Using DMD, it suffices to update the shadow bids without solving the dual program throughout the algorithm. Specifically, let $\varphi(\cdot)$ be a σ-strongly convex function with respect to the ℓ_1-norm (i.e., $\varphi(\lambda) \geq \varphi(\lambda_0) + \langle \nabla\varphi(\lambda_0), \lambda - \lambda_0\rangle + \frac{\sigma}{2}\|\lambda - \lambda_0\|_1^2$ for any λ and λ_0, where $\nabla\varphi(\cdot)$ is the gradient of $\varphi(\cdot)$) and define the Bregman divergence associated with $\varphi(\cdot)$ as follows:

$$D_\varphi(\lambda_1, \lambda_2) := \varphi(\lambda_1) - \varphi(\lambda_2) - \langle \nabla\varphi(\lambda_2), \lambda_1 - \lambda_2\rangle. \quad (7)$$

We are now ready to present the SBL algorithm incorporated with dual mirror descent.

Algorithm 2 (Shadow Bidding with Learning and Dual Mirror Descent (SBL-DMD))

Parameters: Cold start reward coefficients β, target conversion parameter α, and learning rate η.

Initialization: $\lambda^1 \in [0, \beta]^K$.

For $t = 1, 2, \ldots, T$, **do**

Step 1: Observes the context i_t at period t. With probability $\epsilon_t = t^{-1/3}(K \log t)^{1/3}$, the algorithm picks an ad uniformly at random. Otherwise, display an ad $a_t \in \arg\max_j \hat{c}_{i_t j}^t (b_j + \lambda_j^t)$ with arbitrary tie-breaking rules.

Step 2: Updating λ via the online dual mirror descent. Let $s_t(\lambda) = -\sum_{j\in[K]\setminus a_t} \alpha\lambda_j + (\hat{c}_{i_t a_t}^t - \alpha)\lambda_{a_t}$. Let $z_t \in \partial_\lambda s_t(\lambda)$ be a subgradient and assign

$$\lambda^{t+1} \leftarrow \arg\min_{0\leq\lambda_j\leq\beta_j, \forall j\in A} \langle z_t, \lambda\rangle + \frac{1}{\eta} D_\varphi(\lambda, \lambda^t). \quad (8)$$

Step 3: Observe the outcome of a_t, and update the parameters of the underlying machine learning model for predicting $\hat{c}_{ij}^{t+1}$.

Incorporating dual mirror descent into our SBL framework frees our algorithm from solving the empirical dual (6). Instead, the dual variables are

updated with the help of the Bergman divergence $D_\varphi(\cdot,\cdot)$ via a convex optimization (8). In particular, if the strongly convex function $\varphi(\cdot)$ is properly chosen, the dual variable update (8) has a closed-form solution and, thus, can be obtained very efficiently. We show in Theorem 1 that, by setting the learning rate to $\eta = \Theta(1/\sqrt{T})$, the regret of SBL-DMD is of the same order as SBL-RS. In our implementation on Platform O, we adapt SBL-RS to the practical online advertising system (the oSBL algorithm presented in Section 5).

4.2. Analysis of the Regret Bound

Notice that in running the SBL-RS algorithm, we are effectively solving

$$\mathrm{OPT}^t = \min_{0\le\lambda_j\le\beta_j,\ \forall j\in A} \sum_{i\in\mathcal{I}} \hat{p}_i^t \max_{j=1,2,\dots,K} (\hat{c}_{ij}^t(b_j+\lambda_j)) + \alpha\sum_{j=1}^{K}(\beta_j-\lambda_j), \tag{9}$$

where $\hat{c}_{ij}^t$ is the estimate of c_{ij} produced by the underlying prediction model prior to round t, and $\hat{p}_i^t$ denotes the empirical distribution of contexts prior to round t. Before formally presenting the results of our main regret analysis, we address two basic issues regarding this formulation. First, we need to bound the gap between optimal empirical primal allocation and our optimal empirical dual allocation. By strong duality, this gap is induced by tie-breaking in Steps 1 and 2 of the SBL algorithms. As we will show in Online Appendix B, adding an arbitrarily small perturbation to the CTR estimate $\hat{c}_{ij}^t$ will ensure that the tie breaking in Step 1 will only induce an arbitrarily small additional regret. To bound the gap from tie breaking in Step 2, we make the following assumption.

Assumption 1. *For each context $i\in\mathcal{I}$, each ad $j\in A$, and each period t, it holds that*

$$\hat{p}_i^t\hat{c}_{ij}^t \le O(T^{-\frac{1}{3}}(\log T)^{\frac{1}{3}}K^{-\frac{5}{3}}).$$

Assumption 1 states that the empirically estimated probability of a user with context i clicking ad j is negligible. This assumption is introduced to guarantee that the error from tie breaking in Step 1 of the SBL algorithms is small. Similar assumptions are made for other primal-dual settings (e.g., Devanur and Hayes 2009, Agrawal et al. 2014). We note that a typical online DSP faces hundreds of millions of different users, each of whom can be regarded as a unique context. Therefore, Assumption 1 is made without loss of generality in practice. We remark that Step 2 of SBL-RS generally incurs the computational/optimization error in practice as we have discussed. To demonstrate the effectiveness of our method, subgradient descent algorithm together with the arbitrary tie-breaking rule for our ad allocation model, in Online Appendix E, we report a numerical experiment that shows that the solution of method produces negligible error compared with the exact solution in the primal space solved by the simplex method.

The second source of regret for the SBL algorithms is the prediction error associated with the underlying machine learning model to estimate CTR. Clearly, the performance of our algorithms depends on that of the underlying predictor for estimating CTR. In practice, the underlying predictor returns the predicted CTR in period t, $\hat{c}_{ij}^t$, by training from a class of functions $\mathcal{X}$ that estimates the CTR of each context facing each ad ($X\times A \longmapsto [0,1]$). The CTR predictor may take the form of linear regressors, regression trees, and neural networks, the last of which are the actual case with a practical advertising system like Platform O. To bound the prediction error of the underlying machine learning model, we make the following prediction oracle assumption.

Assumption 2 (Prediction Oracle). *For each ad $j\in A$ with n_j^t observed i.i.d. contexts drawn from the distribution $\mathcal{D}_X$ before round t and the corresponding click-through outcomes of showing ad j to these contexts, with probability at least $1-\delta$, for any context i, the estimate $\hat{c}_{ij}^t$ satisfies $|\hat{c}_{ij}^t - c_{ij}| \le O\left(\sqrt{\log(1/\delta)d/n_j^t}\right)$, where d parameterizes the prediction error of the underlying machine learning oracle and only depends on the function class $\mathcal{X}$.*

Assumption 2 is made regardless of the total number of contexts $m = |\mathcal{I}|$. Instead, it assumes that as long as ad j is displayed for a total of n_j^t times with the user contexts drawn in an i.i.d. fashion from the distribution $\mathcal{D}_X$, the error of estimating its CTR has an order of $O\left(\sqrt{1/n_j^t}\right)$ with a high probability, regardless of which contexts the ad is displayed to at round t. A similar assumption with i.i.d. data and a function-class-dependent prediction error is made by Simchi-Levi and Xu (2021) for a wide class of $\mathcal{X}$, such as kernel methods, random forests, and deep neural networks. An alternative interpretation of Assumption 2 is that the platform has a machine learning oracle for predicting the ad CTRs with reasonable generalization error, that is, it is capable of learning from training data and makes accurate predictions on unseen data. In particular, the error decreases with the training sample size n_j^t and is impacted by the prediction error characterized by d.

We emphasize that Assumption 2 could be satisfied for general prediction models such as linear regression, regression trees, and neural networks. We first note that, in the noncontextual setting (i.e., X is a singleton), Assumption 2 is reduced to the standard Hoeffding's inequality with $d=1$. If $\mathcal{X}$ is the set of linear regressors and the true data-generating process is

indeed a linear function, the prediction oracle assumption holds with the ridge regression and the prediction error term d defined as the context dimension (Hsu et al. 2014). For the standard ridge regression model, it requires $n_j^t \geq \Omega(d \log(d/\delta))$ in general to achieve this $O\left(\sqrt{1/n_j^t}\right)$ error bound. In our context, as long as we additionally assume $T > Kd$, extra exploration of $O(Kd \log(d/\delta))$ periods before the start of the SBL still suffices to achieve the same regret bound in Theorem 1. Because we are most interested in the dependence of regret on T when T is sufficiently large, this condition is naturally satisfied in this regime. Also notice that in the literature, the dependence of error bound on the prediction error term is either d or $\sqrt{d}$, depending on the regularity conditions of the features (Chu et al. 2011, Yang and Wang 2020, Zhou et al. 2020). We explicitly specify those conditions for neural networks in Online Appendix G. For the regression tree predictor, it has been well established in the literature (e.g., Wager and Walther 2015) that an adaptive regression tree with each child node containing at least η fraction of the data points in its parent node and each leaf node containing q training samples has a more general convergence rate of $\sqrt{\log(n_j^t)\log(d)/q \log((1-\eta)^{-1})}$. Therefore, Assumption 2 is satisfied under the mild condition that the regression trees have a fixed depth and $q = \Omega(n_j^t)$, which commonly holds in practice. For a large-scale practical DSP such as Platform O, $\mathcal{X}$ is the set of fully connected neural networks with the rectified linear unit (ReLU) activation function. Assumption 2 holds in this setting with certain parameterization of very wide neural networks. Specifically, the error parameter d will be very large in this setting (i.e., $d = O(m^8)$), which may be impractical in practice. We defer a detailed discussion of the DNN prediction oracle to Online Appendix G. In practice, the number of contexts is large, but thanks to the enormous wealth of data the platform can access, advanced neural network algorithms can extract useful information with small generalization errors. We are now ready to state our main theoretical result in the following theorem.

Theorem 1 ($\tilde{O}(T^{2/3})$ Regret Bound). *Suppose Assumptions 1 and 2 hold:*

a. *The expected regret of the SBL-RS algorithm is upper-bounded by* $O(T^{2/3}K^{1/3}(\log T)^{1/3}d^{1/2})$.

b. *The expected regret of the SBL-DMD algorithm is upper-bounded by*

$$O(T^{\frac{2}{3}}K^{\frac{1}{3}}(\log T)^{\frac{1}{3}}d^{\frac{1}{2}}) + \mathbb{E}\left[\sum_{t=1}^{T} s_t(\lambda) + \frac{2\eta}{\sigma}T + \frac{1}{\eta}D_\varphi(\lambda, \lambda^1)\right],$$

for any $\lambda \in \prod_{j \in A}[0, \beta_j]$.

Thus, by taking $\lambda = 0$ *and* $\eta = \sqrt{\frac{\sigma \bar{D}}{2T}}$, *where* $\bar{D} := D_\varphi(0, \lambda^1)$, *we have the expected regret of the SBL-DMD algorithm is bounded by* $O(T^{2/3}K^{1/3}(\log T)^{1/3}d^{1/2})$.

Theorem 1 shows that our proposed SBL-RS and SBL-DMD algorithms both have an expected regret of order $\tilde{O}(T^{2/3})$, which is consistent with the ϵ-greedy type algorithms for contextual bandits. Furthermore, the bound depends on the prediction error term of the predictor by $\sqrt{d}$. This is a natural and necessary price we have to pay with the SBL algorithms, which relies on the underlying machine learning model to predict CTR. If the underlying CTR prediction is easy (hard) so that the prediction error term of the predictor is small (large), our algorithm can achieve a sharper (looser) regret bound. Theorem 1 presents our regret bound in the expected regret, but we can easily extend it to a high-probability-type bound using the Azuma-Hoeffding inequality. We also remark that the analysis of our ϵ-greedy based SBL algorithms is more difficult than the standard contextual bandit algorithms with a linear reward function (e.g., Chu et al. 2011). This is because the cold start reward depends on the aggregated click-though outcomes over all T periods.

The proof of Theorem 1 relies on carefully mapping the total rewards into the dual space. We first establish, by Lemma 2 (in Online Appendix B), the approximate complementary slackness and bound the duality gap between the empirical primal and the empirical dual due to tie-breaking in Step 1 of the SBL algorithms by $O(T^{-1/3}K^{1/3}(\log T)^{1/3})$. Then, we build an auxiliary history-independent reward process: each click of ad j generates a reward of $b_j + \beta_j$, regardless of whether the threshold αT is met. Based on the approximate complementary slackness and Hoeffding's inequality, Lemma 3 (in Online Appendix B) bounds, under the SBL-RS algorithm, the gap between the auxiliary reward process and the optimal reward by $O(T^{2/3}K^{1/3}(\log T)^{1/3}d^{1/2})$. Finally, in Lemma 4 (in Online Appendix B), we bound the gap between the auxiliary reward process and the true reward process by $O(\sqrt{KT \log T})$. Because T is usually several orders of magnitude larger than K, putting all bounds together yields the desired regret bound of order $O(T^{2/3}K^{1/3}(\log T)^{1/3}d^{1/2})$ for the SBL-RS algorithm. For the SBL-DMD algorithm, a key property of dual mirror descent (Proposition 1 in Online Appendix B.4) implies that, compared with SBL-RS, it incurs only an additional regret of a lower order, $O(\sqrt{T})$. Therefore, the SBL-DMD algorithm also has a regret bound of $\tilde{O}(T^{2/3})$.

5. Field Experiment Design and Algorithm Implementation

To demonstrate the practical value of our SBL algorithm, we conduct a two-sided randomized field

experiment to causally evaluate its impact on both the revenue and the cold start reward/success rate. In this section, we first discuss our field setting and introduce our two-sided experiment design, then we present the online implementation of our algorithm (i.e., the oSBL algorithm).

5.1. Two-Sided Experiment Design

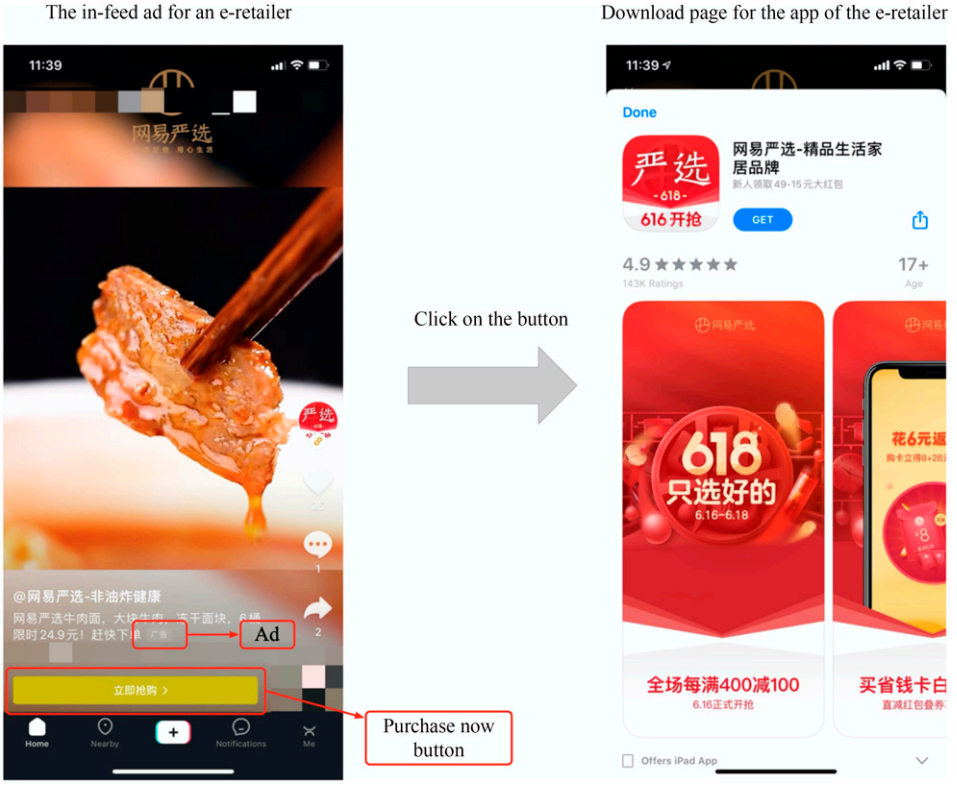
We collaborate with a large-scale online video-sharing platform (Platform O), where in-feed advertising contributes to more than half of its revenue. Platform O features interactive short videos (whose length is typically no more than 30 seconds) and in-feed ads—more akin to TikTok than to YouTube. On Platform O and other social media platforms, in-feed ads are presented in a short-wide format (with an "Ad" label) and intertwined with other organic content updates. As a user swipes up on the screen, a new organic video or an in-feed ad will be shown. Unlike on YouTube, where users have to watch a certain length of an ad before skipping it, Platform O users can swipe up to skip an ad at any time. Users interested in an ad may click the button that directs them to external sites such as AppStore to download the smartphone app, an e-commerce website for online shopping, and so on. Users are converted if they finish the target action set by the advertiser, such as downloading the app or purchasing the product. Figure 3 illustrates this process.

One may want to test the effectiveness of our algorithm by either randomly assigning new ads or randomly assigning user views into treatment and control groups, as shown in Figure 4, panels (a) and (b), respectively. However, both designs would violate the stable unit treatment value assumption (SUTVA, see Imbens and Rubin 2015), thus causing biased estimates for the effect of the new algorithm (Blake and Coey 2014, Johari et al. 2022). Figure 4(a) illustrates the ad-side randomization design, in which new ads are randomly assigned to treatment and control groups. The SBL algorithm is applied to all ads in the treatment group, and the baseline cold start algorithm of the DSP (the real-time bidding prices generated by the PID controller; see Section 3.1 and online Appendix H) is applied to ads in the control group. In this ad-randomization setting, the ads using our new algorithm will compete with those using the baseline algorithm on the same set of impressions, so the global effect of the SBL algorithm (i.e., the effect of the algorithm applied to all the ads on the platform) will be overestimated due to the cannibalization effect. This bias has been confirmed by our numerical simulations, which show that the ad-side experiment overestimates the cold start success rate by as much as 120% (see Table 8 in Online Appendix D.2).

Alternatively, one may conduct an experiment that randomizes over user views (UVs), in which users are randomly assigned to treatment and control groups each using different algorithms (SBL for treatment and the baseline algorithm for control). See Figure 4(b) for an illustration. Such UV-side randomization design has been widely applied in other online platform contexts (see, e.g., Schwartz et al. 2017). For our setting, however, the UV-side randomization design is also invalidated, again due to the violation of SUTVA. Both the SBL algorithm and the baseline algorithm are applied to the same new ads, through which the effect of SBL will spill over to the control group. Specifically, under this experiment design, the SBL algorithm is applied to all new ads so that the underlying machine learning model could produce better CTR estimates for all new ads, which are also served by the baseline algorithm. Therefore, the effect of the baseline algorithm will be overestimated. Due to such spillover effect, directly comparing the outcomes of the treatment- and control group users under the UV-side randomization will result in underestimates for the effect of our algorithm. Our simulation studies have confirmed that such underestimation bias under UV-side randomization could be as high as 40% (see Table 8 in Online Appendix D.2).

To address the aforementioned SUTVA violation issue under one-sided experiments, we design a novel two-sided field experiment to evaluate our SBL algorithm. A similar two-sided experimental framework has also been studied by Johari et al. (2022) from a theoretical perspective. Liu et al. (2021) study a similar two-sided randomization design with proportionally split budgets, and implement it on an online advertising marketplace. The major difference between our setting and theirs is that we measure the outcomes of both the UV side and the ad side, which motivates us to implement and analyze the novel two-sided design. Our experiment was conducted from May 23, 2020, to May 30, 2020; the experiment design is illustrated in Figure 5. Specifically, we randomly assigned 33% platform UVs into the treatment group and another 33% UVs into the control group. On the ad side, we randomly assigned 20% of the new ads to the treatment group and 20% to the control group. The rest of the UVs and ads are referred to as the nonexperiment UVs and nonexperiment ads. The ad-side randomization is independent from the UV-side randomization. The SBL algorithm is applied if both the UV and ad are in the treatment group (cell B11 in Figure 5), whereas the baseline algorithm is applied if both the UV and ad are in the control group (cell B22 in Figure 5). The salient feature of this design is that the treatment (control) ads can bid only on the treatment (control) UVs; they are not allowed to bid on the control (treatment) UVs. Implementation-wise, we set the bids in cells B12, B21, B31, and B32 in Figure 5 to 0. For the nonexperiment new ads, we applied the baseline cold start algorithms

Figure 3. (Color online) How Ads Are Displayed to Users[10]



regardless of UVs (cells B13, B23, and B33 in Figure 5). Finally, we keep the bidding algorithm for the mature ads (cells B14, B24, and B34 in Figure 5) unchanged.

Through such a two-sided randomization design, the SUTVA condition is restored to the greatest extent we can. First, our two-sided design avoids the direct competitions between treatment and control ads on the same UVs and therefore removes the cannibalization effect of experimenting with ad-side randomization only. Furthermore, blocking the control UV impressions for treatment ads and the treatment UV impressions for control ads (B21 and B12 in Figure 5, respectively) confines the treatment ads to our SBL algorithm and the control ads to the baseline algorithm, thus removing the spillover effect of the experiment only with UV-side randomization. Specifically, the CTR estimates, produced by the underlying machine learning system, for the treated ads will be affected only by our SBL algorithm,

Figure 4. (Color online) One-Sided Randomization Experiments

(a) Experiment with ad-side randomization

	Treatment new ads	Control new ads	Non-experiment new ads	Mature ads
100% UV	Treatment condition	Control condition		

(b) Experiment with UV-side randomization

	100% new ads	Mature ads
Treatment UV	Treatment condition	
Control UV	Control condition	
Non-experiment UV		

Figure 5. (Color online) Two-Sided Experiment Design

	20% treatment new ads	20% control new ads	60% non-experiment new ads	Mature ads
33% treatment UV	B11	B12	B13	B14
33% control UV	B21	B22	B23	B24
33% non-experiment UV	B31	B32	B33	B34

Notes. The new ads in cell B11 are bid with the SBL algorithm (i.e., the shadow bids λ^* will be added to the real-time bidding prices). The new ads in cells B21, B31, B12, and B32 are forbidden to join the auction. All other ads in uncolored cells join the auction following the real-time bidding prices without shadow bids.

whereas those for the control ads only by the baseline algorithm. Thus, the difference between the treated ads and control ads shall be causally attributed to the effect of our new algorithm compared with the baseline one. To fully ensure SUTVA for our two-sided experiment, we make two additional assumptions during our experiment period: (1) the CTR and CVR distributions of the mature ads are not affected by the cold start algorithms applied to different UVs; and (2) the number of ad impressions displayed to a user is not affected by the cold start algorithms applied to different ads. We report the verification of both assumptions in Online Appendix C.1. Although our two-sided experiment design helps us to tease out cannibalization bias in a two-sided experiment and generates unbiased causal estimates for the effect of our algorithm, we acknowledge a limitation of our experimental framework is that it may introduce another potential bias by reducing the competition in the auctions for both the treatment and control UVs (i.e., 20% of new ads are blocked for the experimented UVs; see Figure 5). In other words, even though the internal validity of the experiment is secured, the external validity may be affected by this added competition level. However, we remark that, because each ad is targeted to a specific set of platform users, blocking 20% of new ads only reduces the number of ads competing for the experimented UVs by 6%, suggesting a marginal reduction in ad competition in our two-sided experiment. In Online Appendix D, we build a simulation system[11] and conduct simulation studies and sensitivity analysis of our two-sided experiment to demonstrate that the experiment results are close to the ground truth of the overall treatment effect if the algorithm is applied to all UVs and all ads. A theoretically justified approach to fully de-bias the estimates from our two-sided experiment to evaluate the treatment effect of the algorithm applied to all UVs and all ads should be a promising direction for future research.

Some other recent developments on experiment design and analysis have also addressed the violation of SUTVA in a two-sided setting (e.g., Pouget-Abadie et al. 2019, Rolnick et al. 2019). This line of research focuses on developing cluster-level randomization and the associated algorithms to improve the power of statistical inferences in this setting. We, however, take a different approach, proposing a new two-sided experimental framework that causally evaluates an advertising algorithm for a large-scale DSP.

5.2. Online Implementation of the Algorithm

We highlight a key advantage of our SBL algorithms—they can be easily adapted into the infrastructure of Platform O's DSP. Such convenience has enabled us to actually implement the algorithm online. The implemented version of the algorithm is online shadow bidding with learning (oSBL), as detailed next.

Algorithm 3 (Online Shadow Bidding with Learning (oSBL))

Parameters: Set epoch schedule $1 = \tau_1 < \tau_2 < \cdots < \tau_m = T$ with fixed, one-hour intervals; the cold start reward coefficient $\beta_j = 2b_j$; and the conversion target $\alpha T = 10$ for new ads.

Initialization: $\lambda^1 \leftarrow 0$, $m \leftarrow 1$.

For $t = 1, 2, \ldots, T$, **do**

Step 1: Observe the context i_t at round t. Choose the top 150 ads (including new and mature ads, ranked by a preranking model[12]), together with 15 randomly picked new ads, to join the auction.

Step 2: Get $\hat{c}^t_{i_t j}$, the estimate of pCTR × pCVR. Display the ad $a^*_t = \arg\max_{j \in [K_t]} \hat{c}^t_{i_t j}(b^t_j + \lambda^{\tau_m}_j)$, where b^t_j is the system bidding price calculated by a real-time PID system for ad j at period t, and $[K_t] = [K^n_t] \cup [K^m_t]$ is the set of 165 ads that join the auction at time t, with $[K^n_t]$ ($[K^m_t]$) as the set of new (mature) ads.

Step 3: If $t = \tau_m$, construct the history data set $\mathcal{H}_t$ by randomly sampling 4% of the auctions in the past hour, the auction/round index set of which is denoted by $\mathcal{T}^t$. Update $m \leftarrow m + 1$, and $\lambda^{\tau_m}_j$ for each ad j by solving the following dual program,

where the shadow bidding price for mature ads is set as 0, that is, $\lambda_j = 0, \ \forall j \in [K_\tau^m]$,

$$\min \sum_{\tau \in \mathcal{T}^t} \max_{j \in [K_\tau]} \{\hat{c}_{i_\tau j}^{\tau}(\lambda_j + b_j^s)\} + \alpha |\mathcal{T}^t| \sum_{j \in [K_\tau]} (\beta_j - \lambda_j)$$

$$\text{s.t.} \quad \lambda_j \in [0, \beta_j], \ \forall j \in [K_\tau^n], \ \lambda_j = 0, \ \forall j \in [K_\tau^m].$$

Step 4: Observe the outcome of ad a_t^*, and update the parameters of the neural networks. The advertiser will be charged based on the real-time system bidding price b_j^t, instead of the total bid $b_j^t + \lambda_j^{\tau_m}$.

Several aspects of the oSBL algorithm are different from the original SBL-RS and SBL-DMD. First, oSBL accounts for both new and mature ads, with the shadow bids for mature ads fixed at 0. Second, the bid for each ad j in oSBL will follow the system bidding prices generated by the PID system (so the bid will change over time), to which the shadow bids are adaptively added. Furthermore, the two-sided experiments (Figure 5) can also be easily implemented online by adjusting the shadow bids according to which cell the ad-UV pair belongs to. Third, the exploration of the oSBL algorithm is to randomly add 15 new ads into the final auction for each impression, instead of the ϵ-greedy scheme proposed in SBL-RS and SBL-DMD. This adjustment is mainly driven by the fact that we inherit the 10%-exploration heuristic that has already been implemented by Platform O's DSP. Making minimum changes to the online system of the DSP will ensure the robustness of our new algorithm. Fourth, when computing the shadow bids λ_j^* for each ad j, we sample 4% of the total auctions for user impressions. Such a downsampling approach could further reduce the computational burden of the oSBL algorithm. As a matter of fact, our algorithm could produce robust shadow bids even with a sampling rate of only 1%, as shown by our robustness check results in Online Appendix F.

We also set the fixed, one-hour epoch schedule interval in oSBL to update the shadow bids. On the one hand, this makes the pace of the algorithm consistent with other online systems of the DSP, such as the predictive models for pCTR and pCVR, and the PID controller. On the other hand, it alleviates the computational burden of the algorithm so that the shadow bids can be generated in a timely manner. Also, due to the engineering constraint, the pCTR and pCVR data cannot be accessed in real time by our field experiment in our actual implementation on Platform O—directly resolving the optimal dual variables for the empirical allocation problem will be more robust than updating via the mirror descent procedure. This is why oSBL is implemented in a resolving fashion on Platform O.

Finally, we set the cold start reward coefficient $\beta_j = 2b_j$ and the conversion target $\alpha T = 10$ mainly because of the business practice of Platform O's DSP. Furthermore, as shown by our simulation results in Section 6.3 with $\beta_j = 2b_j$, the oSBL algorithm would yield a substantial (at least 5.35%) increase in Platform O's long-term total advertising revenue. The more sophisticated choice of the cold start reward coefficient, therefore, will boost the long-term revenue even higher.

6. Field Experiment Results

In this section, we present the results of our two-sided field experiment. The randomization check (see Online Appendix C.2) confirms that both the treatment ads and the control ads in our sample are comparable, implying that any difference between groups after the experiment started should be attributed to whether our new oSBL algorithm has been implemented. In the following subsections, we document three sets of results to demonstrate the value of our proposed algorithm: (1) the short-term impact, (2) the long-term impact, and (3) the global treatment effect on advertising revenue. This experiment, together with a comprehensive simulation study, shows that advertising revenue from our oSBL algorithm increased at least 5.35%. For a large-scale platform such as Platform O, such an increase would translate to hundreds of millions of U.S. dollars per year.

6.1. Short-Term Performance of Our oSBL Algorithm

We present the model-free results here. See Online Appendix C.4 for the robustness checks with regression models. We base our analysis on the following metrics during the experiment period:

1. Cold Start Success Rate. This metric is defined as the proportion of ads whose total number of conversions exceeds the conversion target, that is, $\sum_{j=1}^{K} \mathbb{I}_{\{V_j \geq \alpha T\}}/K$, where K is the total number of new ads assigned to the respective experimental group. For Platform O, the cold start period of any ad is the first three days, whereas the conversion target during the cold start period is 10 conversions, that is, $\alpha T = 10$. New ads that arrive in the last three days of the experiment do not pass the entire cold start period, but this will not affect comparisons between the treatment and control groups.

2. Cold Start Reward. This metric is clustered at each ad j, that is, $\beta_j \min\{V_j, \alpha T\}$.

3. Short-Term Revenue. This metric is clustered at the UV level for all (old and mature) ads in different experiment groups on the DSP.

4. Ratio Between Real and Target Cost per Conversion. This metric is clustered at the ad level to evaluate the impact of our oSBL algorithm on the controllability of advertisers' costs. If this ratio is significantly larger than 1, it implies that our new algorithm substantially increases the cost per conversion for the advertisers, which may cause them to complain or even leave the platform.

Table 2. Short-Term Effects of oSBL

Panel A: Effects on the cold start at the ad level

	Time window: May 23–30, 2020		
Dependent variable	Treatment (1)		Control (2)
Number of impressions	29,866 (268,139)		20,605 (232,143)
Relative effect size		44.94%****	
Number of clicks	2,352 (24,088)		1,738 (30,780)
Relative effect size		35.34%**	
Number of conversions	9.38 (108.93)		5.86 (68.67)
Relative effect size		59.67%****	
Observations	34,605		34,076
Cold start success rate	0.0438 (0.204)		0.0271 (0.162)
Relative effect size		61.62%****	
Cold start reward	261.6 (958.1)		177.1 (930.5)
Relative effect size		47.71%****	
Observations	34,605		34,076

Panel B: Effects of the algorithm on short-term revenue and the objective value

	Time window: May 23–30, 2020		
Dependent variable	Treatment (1)		Control (2)
Revenue per user	1.439 (37.81)		1.448 (37.83)
Relative effect size of total revenue		−0.717%**	
Observations	240,308,309		240,538,298
Total objective value	354,856,325		354,334,315
Relative effect size		0.147%****	

Panel C: Effects of the algorithm on advertiser costs

$\frac{\text{Real Cost}}{\text{Target Cost}} - 1$	−30% ~ 30% (1)	30% ~ 100% (2)	>100% (3)
Proportion of ads (treatment condition)	0.665	0.041	0.059
	(0.472)	(0.198)	(0.235)
Proportion of ads (control condition)	0.648	0.026	0.045
	(0.477)	(0.159)	(0.209)
p-value of t-test	0.74	0.45	0.58

Notes. Mean values are reported in this table. To protect sensitive data, the impressions, clicks, conversions, and revenues are linearly scaled. The cutoff ranges of panel C—[−30%, 30%], [30%, 100%], and > 100%—are adopted in consistency with Platform O's business practice. Standard errors in panels A and C are clustered at the ad level and reported in parentheses.

*$p < 0.1$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

We summarize our experimental results on the short-term impact of the oSBL algorithm in Table 2, which can also be replicated by regression analysis presented in Online Appendix C.4. It is evident from panel A that our oSBL algorithm has substantially increased both the cold start success rate and the cold start reward of new ads by 61.62% and 47.41% (p-values ≤ 0.0001). Such improvements result from the shadow bids produced by oSBL that are added to the real-time system bidding prices, which also give rise to a 44.94% increase in ad impressions, a 35.34% increase in ad clicks, and a 59.67% increase in ad conversions (p-values ≤ 0.01) during the cold start period of new ads.

Thanks to the shadow bids, our oSBL algorithm significantly improved the new ad cold start performance, but it could also have cannibalized the impressions and conversions of mature ads. As a consequence, the algorithm could have reduced total short-term revenue during the experiment. Comparing the per-UV revenue of the treatment and control groups on the UV side, panel B of Table 2 confirms this intuition, by quantifying that the oSBL algorithm will decrease short-term revenue by 0.717% (with a p-value less than 0.01). This small relative decrease in short-term revenue is both within our expectation and acceptable for Platform O. And, as we articulate in Section 6.3, a short-term loss (−0.717%) can be well compensated for by the long-term revenue boost (5.35%) of the oSBL algorithm.

Are the improvements in success rate and reward offset by increased advertiser costs? Panel C of Table 2 addresses this question by examining the distribution of the relative gap between the real cost of an ad and

its target cost gap (measured by $\frac{\text{Real Cost}}{\text{Target Cost}} - 1$). It shows that there is no significant difference between the distribution of the relative gap for ads in the treatment group and that for ads in the control group. Specifically, the treatment and control groups are similar in the proportion of new ads whose relative cost gap is in each of the following ranges [−30%, 30%], [30%, 100%], and > 100%. The results show that the oSBL algorithm does not boost advertisers' cost to increase the cold start success rate and the cold start reward of their new ads.

Last but not least, our oSBL algorithm has substantially increased the prediction accuracy for the CTR of new ads. Specifically, our two-sided experiments show that the area under the curve (AUC) of new ad CTR prediction in the treatment group is 7.48% larger than the one in the control condition, with the p-value of t-test being 0.017. (To protect sensitive data, we only report the relative difference here.)

6.2. Long-Term Performance of Our oSBL Algorithm

We next examine the long-term impact of oSBL on both ads and advertisers after the cold start period, when the shadow bids are set to zero. Regarding ads, we evaluate how oSBL influences the lifetime performance of an ad after the cold start period by comparing the following postexperiment metrics of the treatment and control ads: (a) retention days (number of days that an ad is active after the cold start period), (b) lifetime number of impressions, (c) lifetime revenue, (d) CTR/CVR, and (e) average system bidding price. Because the distribution of impression and lifetime advertising revenue after cold start is heavy tailed, we perform the t-test after taking log-transformation of or winsorizing the revenue at the 99% level. Regarding advertisers, we investigate whether oSBL changes advertiser behaviors, especially their bid prices and the length of time they wish to keep their ads active on Platform O.

Panel A of Table 3 documents the effects on ads. The results show that our algorithm significantly increased—by 3.13%—the average number of active days and, thus, the average market thickness (defined as the average number of ads competing for each user impression). Figure 6 plots the scaled (to protect sensitive data) total number of ads in different experiment conditions that remained active each day after the experiment—it shows that our proposed algorithm significantly increased market thickness—by 7.21% on average—especially during the first two weeks after cold start.

Comparing the lifetime revenues of the treatment and control ads reveals further insights. The oSBL algorithm boosted postexperiment revenue after cold start by 34.02%. This benefit is driven by the fact that the algorithm not only thickens the market by retaining the ads longer but also successfully identifies high-quality ads with 11.14% higher CTRs. By algorithmically adding the optimal shadow bids to the system bidding prices, the oSBL algorithm automatically awards more user traffic to the new ads with higher CTR potential, thus significantly increasing the CTRs of the treatment ads after the cold start period. To this regard, the benchmark PID algorithm under-explores the new ads so that it is unable to identify the ads with the highest CTR performance in the long run with a high confidence. We also observe that our proposed algorithm had no significant impact on the CVR and average system bidding prices of an ad. In summary, the oSBL algorithm substantially improved the market thickness and CTR of the ads after cold start. In Section 6.3, we build a simulation model to demonstrate that such long-term effects on ads could be translated into a significant global treatment effect of our algorithm on the advertising revenue of a DSP.

Given that the oSBL algorithm significantly improves ads' CTRs and, thus, revenue performance, would advertisers also respond to such improvements by changing their behaviors on Platform O (such as bidding prices)? In particular, if an advertiser increases its expectation on cold start performance, would the effectiveness of our algorithm be weakened? To address these questions, we next examine whether advertisers would behave differently after oSBL is adopted. To this end, we adopt a two-stage least squares (2SLS) specification in Equation (10) to identify whether the total number of conversions during the cold start period will change an advertiser's behavior, where X_j are the advertiser-specific features such as industry fixed effects, bidding prices, budge, and target strategy. We define exp ratio as the proportion of treatment ads among all the ads in our experiment for each advertiser and adopt it as the instrumental variable. The endogenous variable is the total number of conversions by an advertiser during the cold start period and the experiment, whereas the dependent variable may take different forms such as average bidding prices, the total number of impressions/conversions, and the average number of retention days for all the ads of an advertiser. Note that exp ratio is a valid instrument in this setting. One the one hand, the p-values of the weak instrument tests are smaller than 10^{-5}, so the strong first-stage assumption holds. On the other hand, it is unlikely that exp ratio could impact an advertiser's behavior through a channel other than conversions, so the exclusion restriction also holds.

$$\begin{aligned}&\text{First Stage :}\\&\text{Cold Start Conversion}_j = \alpha_0 + \alpha_1 \text{exp ratio}_j + X_j + \epsilon\\&\text{Second Stage :}\\&\text{Dependent Variable}_j = \beta_0 + \beta_1 \text{Cold Start Conversion}_j + X_j + \epsilon\end{aligned} \tag{10}$$

Finally, panel B of Table 3 reports the estimation results for the 4,340 advertisers we study. After controlling for industry fixed effects, bidding prices, budget, and target

Table 3. The Long-Term Effects of oSBL

Panel A: Effects of oSBL on ads			
	Time window: May 31–August 31, 2020		
Dependent variable	Treatment (1)	Control (2)	Relative increase (3)
Retention days	10.20 (11.03)	9.89 (10.81)	3.13%**
Log (Impressions)	8.02 (3.18)	7.46 (3.00)	****
99% Winsorized impressions	107,424 (379,251)	63,981 (242,250)	67.90%****
Log (Revenue)	2.60 (2.69)	2.13 (2.56)	****
99% winsorized revenue	485 (1,914)	362 (1,526)	34.02%****
CTR	0.054 (0.059)	0.049 (0.056)	11.14%****
CVR	0.023 (0.132)	0.024 (0.138)	$p > 0.1$
Bid prices	57.14 (62,62)	57.19 (61.30)	$p > 0.1$
Observations	34,605	34,076	

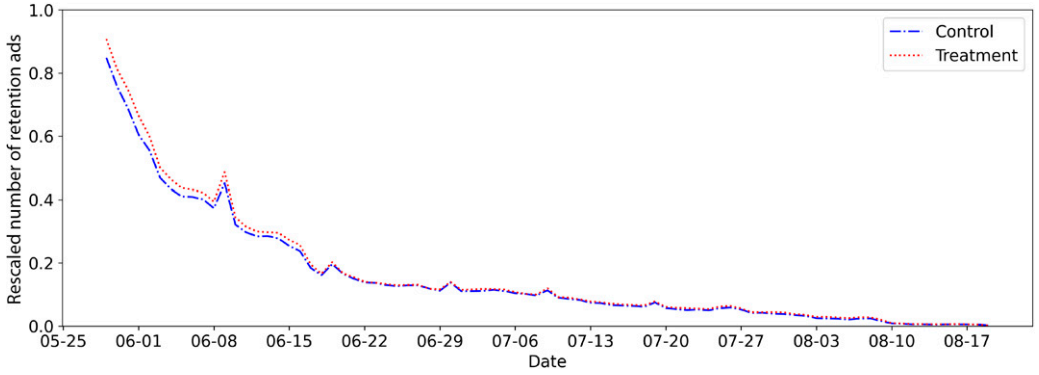
Panel B: Effects of oSBL on advertiser behaviors				
	Time window: May 31–June 25, 2020			
	Dependent variable:			
	Bidding prices (1)	Impressions (2)	Conversions (3)	Retention days (4)
Treatment–Control	29.57	5568	6.35	−0.0023
	(334.04)	(13,180)	(20.10)	(0.073)
p-value	0.93	0.67	0.75	0.75
Industry fixed effects	Yes	Yes	Yes	Yes
Bidding price		Yes	Yes	Yes
Budget	Yes	Yes	Yes	Yes
Target strategy	Yes	Yes	Yes	Yes

Notes. Mean values are reported in panel A. To protect sensitive data, all metrics are linearly scaled. For panel B, we report only the coefficient and its standard error of the endogenous variable (i.e., the number of conversions during the cold start period and the experiment). Standard errors in panel A (panel B) are clustered at the ad (advertiser) level and reported in parentheses.

*$p < 0.1$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$.

strategy, we find no evidence that implies our oSBL algorithm significantly changed advertisers' long-term behaviors on Platform O. Notice that the analysis of the long-term effect of our oSBL algorithm is based on a one-time experiment that only lasted for eight days. The advertisers' behaviors on the platform might change if the oSBL algorithm is applied to all their ads for a longer period of time, thus invalidating the aforementioned benefits of the algorithm such as increasing the market thickness and identifying the ads with a high CTR. The oSBL algorithm boosts the ad retention and market thickness by awarding more user traffic and, consequently, more conversions to new ads. In the long run, however, advertisers might perceive such additional traffic and conversions for their new ads and, thus, increase their expectation of new ad performance correspondingly. This will in turn change advertiser behavior and weaken the effectiveness of

Figure 6. (Color online) Effect of oSBL on Market Thickness



the algorithm. We acknowledge this limitation of our study, which can be addressed with running a high-cost long-run hold-out experiment similar to ours.

6.3. Global Treatment Effect of Our oSBL Algorithm on Advertising Revenue

Our experiment cannot directly observe the effect of oSBL on the long-term revenue change, because all the experimental new ads would flow into the pool of mature ads and join the auction under both treatment and control UVs. One solution is to use the two-sided experiment with blocking for those experimented new ads after they mature. Though this design can better deal with cannibalization and spillover effects when estimating the long-term effect of the algorithm on the revenue, it is much more costly, due to the lifetime blocking of all ads for a substantial portion of UVs.

In this subsection, we seek to quantify, through simulation, the global treatment effect of our oSBL algorithm on advertising revenue. Specifically, based on our empirical results on the long-term impact of the algorithm (Table 3, panel A), our oSBL could substantially improve the length of ads' retention time by 3.13% and CTR by 11.14% without negatively affecting their CVR and average system bidding prices. This motivates us to build a simulation model to translate such positive impact into the long-term revenue boost for the platform.

To estimate the long-term revenue increase our algorithm could generate, we use data with 12 million impressions between April 9 and April 30, 2020. We randomly sampled 1.2 million impressions with replacement for each simulation and replicated 10 times via Bootstrap—the following results all pass the t-test with p-values smaller than 10^{-3}. In our simulation, we apply the oSBL for the new ads in the treatment condition. After the cold start, new ads flowed into the pool of mature ads. As we documented in Section 6.2, the ads under the oSBL have a higher CTR and longer retention after the cold start period. To model this oSBL effect, we assume the CTR of the treated ads will increase by Δ_{CTR} (relative changes) and their retention time length will increase by Δ_r (relative changes). As shown next, because the values of Δ_{CTR} and Δ_r may change once the algorithm is applied to all ads and the entire user traffic, we perform sensitivity analysis by varying the values of Δ_{CTR} and Δ_r.

We first validate our model by replicating our experimental results so that the short-term revenue will decrease when oSBL is applied to 20% of new ads (see panel B of Table 2) during the experiment period (May 23–30, 2020). We use the nonexperiment impressions (B13, B23, B33, B14, B24, and B34 in Figure 5) in the simulation. We compare the total revenue of these eight days for two cases: where oSBL is applied to 20% of new ads and where the baseline algorithm is applied to all new ads. We find that the average short-term revenue decrease is 0.583%, which is consistent with our regression-based result that oSBL (applied to 20% of the ads) will decrease short-term revenue by 0.592%. Therefore, our simulation model is fairly accurate in predicting the short-term revenue loss caused by oSBL.

The estimation of two key parameters in our simulation model Δ_r (which refers to the average relative increase of the retention length for mature ads) and Δ_{CTR} (which refers to the average relative increase of the CTR for mature ads) relies on the two-sided experiment where only 20% of ads and 33% of UVs are included in the treatment group. Therefore, it is challenging to extrapolate their estimates to the counterfactual setting, where the oSBL algorithm is applied to the entire ad set and user population. To obtain a complete picture on the global treatment effect of our algorithm, we conduct a sensitivity analysis with our simulation model by varying Δ_r from 0% to 5% and varying Δ_{CTR} from 0% to 20%, assuming that oSBL is applied to all ads and UVs. Baseline revenue is denoted by R_0 and revenue under the oSBL algorithm is denoted by $R(\Delta_r, \Delta_{CTR})$ (so $R_0 = R(0,0)$). We are interested in the relative advertising revenue increase associated with oSBL:

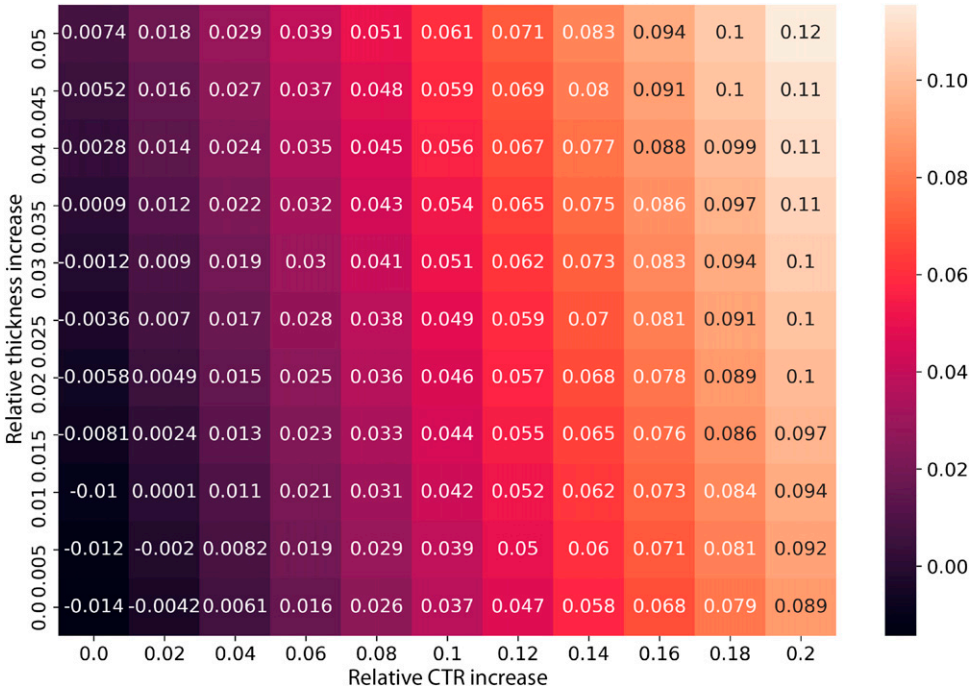
$$\Xi(\Delta_r, \Delta_{CTR}) = \frac{R(\Delta_r, \Delta_{CTR}) - R_0}{R_0} \times 100\%.$$

Figure 7 demonstrates that for a wide range of potential values for Δ_r and Δ_{CTR}, the relative revenue increase $\Xi(\Delta_r, \Delta_{CTR})$ is significantly above 0, implying that our oSBL algorithm could substantially boost the long-term advertising revenue of a DSP. In particular, if we linearly extrapolate the estimates from our two-sided experiment, $\Delta_r = 3.13\%$ and $\Delta_{CTR} = 11.14\%$, to all ads and the entire population, the advertising revenue increase from our algorithm is at least 5.35%. For a large-scale platform such as Platform O, such an increase would translate to hundreds of millions of U.S. dollars per year. Finally, the previous simulations do not fully capture, if any, the long-term behavior changes of the advertisers when the new oSBL algorithm are applied to all their ads for a long period of time as we discussed in the previous subsection. This is an inevitable limitation of our simulation study, because our experiment only lasted for eight days.

7. Discussion and Conclusion

We close by discussing several promising directions for future research. In Sections 3 and 5, we detail the cost-control problem brought by complicated auction mechanisms and bidding/payment methods in a real DSP. Future research could examine the cold start algorithm under real-time bidding. Furthermore, our cold start algorithm, in principle, could be embedded

Figure 7. (Color online) Global Treatment Effect of oSBL on Advertising Revenue



into a recommender system for general content as well. In this setting, ranking is a prominent data-driven decision. Thus, an interesting extension of our work would be integrating MAB algorithms and state-of-the-art ranking models such as Learning2Rank. Finally, future research could test other ways of conducting two-sided field experiments and quantify the biases that may be introduced by the violation of SUTVA in two-sided platforms.

Acknowledgments

The authors thank Department Editor Prof. Gabriel Weintraub, the anonymous associate editor, and three referees for their very helpful and constructive comments, which have led to significant improvements in both the content and exposition of this study. The authors are also indebted to Prof. Hengchen Dai for her constructive feedback on the initial draft of this work. They also thank the industry partner for their support on sharing the data, implementing the algorithm, and conducting the experiment.

Endnotes

[1] See https://www.iab.com/insights/internet-advertising-revenue-fy2019-q12020/ for more details.

[2] See the financial report of Facebook: https://www.sec.gov/ix?doc=/Archives/edgar/data/1326801/000132680120000013/fb-12312019x10k.htm.

[3] Note that one advertiser can launch multiple ad campaigns, so we may also define market thickness as the average number of advertisers on the platform. These two thickness metrics clearly have strong positive correlations, and our paper focuses on the ad's perspective. Furthermore, one can split the platform into multiple submarkets according to the industry the advertiser belongs to (e.g., e-commerce, online gaming, and FinTech). Such segmentation will allow for individualized implementation and analysis for each submarket.

[4] It is indeed the case with our industry partner, Platform O, that the total number of user impressions for ads is not significantly affected by the advertisement algorithm the platform adopts.

[5] The retention rate in these two weeks is defined as the number of ads that have exposure to users every day in these two weeks divided by the total number of ads. To protect the platform's identity and sensitive data, we re-scale the y-axis value to [0,1]. The curve pattern remains the same if we vary the duration from one day to 14 days.

[6] Obviously, Figure 1 only shows the correlation between new ads' conversions during the cold start and their long-run retention rates. In Online Appendix C.1, we also conduct extensive empirical analysis to provide causal evidence that gaining 10 conversions during the cold start will significantly boost ad retention by 15.03%.

[7] Platform O's area under the curve (AUC) of new ad CTR prediction is 5.77% smaller than that of mature ads, a sizable gap for a large-scale platform that indicates it is more difficult to predict the CTR of new ads than that of mature ones. This prediction inaccuracy is amplified by the sparsity of conversions. Along these lines, Facebook recommends that its advertisers earmark enough budget for at least 50 conversions to successfully bring their ads out of the initial learning phase (i.e., the cold start period). See https://www.facebook.com/business/help/112167992830700?id=561906377587030.

[8] See https://www.blog.google/products/admanager/rolling-out-first-price-auctions-google-ad-manager-partners/.

[9] For ease of reading, we summarize all the notations in Table 4 in Online Appendix A.

[10] To protect the Platform O's identity, we created the screenshots of an in-feed ad on another large online advertising platform in Figure 3, whose interface is similar to Platform O.

[11] See the GitHub repository at https://github.com/zikunye2/cold_start_to_improve_market_thickness_simulation for the code of our simulation system.

[12] On Platform O's DSP, there are two stages before an ad enters the final auction—filtering and pre-ranking—both of which adopt deep neural network models to rule out the ads not suitable for the user impression.

References

Agrawal S, Devanur NR (2014) Bandits with concave rewards and convex knapsacks. *Proc. 15th ACM Conf. Econom. Comput.* (Association for Computing Machinery), 989–1006.

Agrawal S, Devanur NR, Li L (2016) An efficient algorithm for contextual bandits with knapsacks, and an extension to concave objectives. *29th Annual Conf. Learn. Theory*, vol. 49 (PMLR), 4–18.

Agrawal S, Wang Z, Ye Y (2014) A dynamic near-optimal algorithm for online linear programming. *Oper. Res.* 62(4):876–890.

Agarwal A, Dudík M, Kale S, Langford J, Schapire R (2012) Contextual bandit learning with predictable rewards. *Proc. 15th Internat. Conf. Artificial Intelligence Statist.*, vol. 22 (PMLR), 19–26.

Agarwal A, Hsu D, Kale S, Langford J, Li L, Schapire R (2014) Taming the monster: A fast and simple algorithm for contextual bandits. *Proc. 31st Internat. Conf. Machine Learn* 32(2):1638–1646.

Badanidiyuru A, Kleinberg R, Slivkins A (2018) Bandits with knapsacks. *J. ACM* 65(3):1–55.

Balseiro SR, Gur Y (2019) Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Sci.* 65(9):3952–3968.

Balseiro SR, Besbes O, Weintraub GY (2015) Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Sci.* 61(4):864–884.

Balseiro S, Lu H, Mirrokni V (2022) The best of many worlds: Dual mirror descent for online allocation problems. *Oper. Res.*, ePub ahead of print May 23, https://doi.org/10.1287/opre.2021.2242.

Balseiro SR, Feldman J, Mirrokni V, Muthukrishnan S (2014) Yield optimization of display advertising with ad exchange. *Management Sci.* 60(12):2886–2907.

Bastani H, Simchi-Levi D, Zhu R (2022) Meta dynamic pricing: Transfer learning across experiments. *Management Sci.* 68(3):1865–1881.

Bietti A, Agarwal A, Langford J (2021) A contextual bandit bake-off. *J. Machine Learn. Res.* 22(133):1–49.

Bimpikis K, Elmaghraby WJ, Moon K, Zhang W (2020) Managing market thickness in online business-to-business markets. *Management Sci.* 66(12):5783–5822.

Blake T, Coey D (2014) Why marketplace experimentation is harder than it seems: The role of test-control interference. *Proc. 15th ACM Conf. Econom. Comput.* (Association for Computing Machinery), 567–582.

Caldentey R, Vulcano G (2007) Online auction and list price revenue management. *Management Sci.* 53(5):795–813.

Chen N, Gallego G (2022) A primal-dual learning algorithm for personalized dynamic pricing with an inventory constraint. *Math. Oper. Res.* ePub ahead of print February 10, https://doi.org/10.1287/moor.2021.1220.

Chen B, Chao X, Ahn HS (2019) Coordinating pricing and inventory replenishment with nonparametric demand learning. *Oper. Res.* 67(4):1035–1052.

Chen W, Shi C, Duenyas I (2020) Optimal learning algorithms for stochastic inventory systems with random capacities. *Production Oper. Management* 29(7):1624–1649.

Choi H, Mela CF, Balseiro SR, Leary A (2020) Online display advertising markets: A literature review and future directions. *Inform. Systems Res.* 31(2):556–575.

Chu W, Li L, Reyzin L, Schapire R (2011) Contextual bandits with linear payoff functions. *Proc. 14th Internat. Conf. Artificial Intelligence Statist.* (PMLR), 208–214.

Cui R, Li J, Zhang DJ (2020) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on Airbnb. *Management Sci.* 66(3):1071–1094.

Cui R, Zhang DJ, Bassamboo A (2019) Learning from inventory availability information: Evidence from field experiments on Amazon. *Management Sci.* 65(3):1216–1235.

Dave K, Varma V (2014) Computational advertising: Techniques for targeting relevant ads. *Foundations Trends Inform. Retrieval* 8(4–5):263–418.

Devanur NR, Hayes TP (2009) The adwords problem: Online keyword matching with budgeted bidders under random permutations. *Proc. 10th ACM Conf. Electronic Commerce* (Association for Computing Machinery, New York), 71–78.

Dudik M, Hsu D, Kale S, Karampatziakis N, Langford J, Reyzin L, Zhang T (2011) Efficient optimal learning for contextual bandits. *Proc. 27th Conf. Uncertainty Artificial Intelligence*, 169–178.

Feldman J, Zhang DJ, Liu X, Zhang N (2021) Customer choice models vs. machine learning: Finding optimal product displays on Alibaba. *Oper. Res.* 70(1):309–328.

Ferreira KJ, Simchi-Levi D, Wang H (2018) Online network revenue management using Thompson sampling. *Oper. Res.* 66(6):1586–1602.

Fisher M, Gallino S, Li J (2018) Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Sci.* 64(6):2496–2514.

Foster D, Agarwal A, Dudík M, Luo H, Schapire R (2018) Practical contextual bandits with regression oracles. Dy J, Krause A, eds. *Proc. 35th Internat. Conf. Machine Learn.*, vol. 80 (PMLR), 1539–1548.

Gallego G, Van Ryzin G (1994) Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management Sci.* 40(8):999–1020.

Golrezaei N, Javanmard A, Mirrokni V (2019) Dynamic incentive-aware learning: Robust pricing in contextual auctions. *Adv. Neural Inform. Processing Systems* 32:9759–9769.

Golrezaei N, Nazerzadeh H, Rusmevichientong P (2014) Real-time optimization of personalized assortments. *Management Sci.* 60(6):1532–1551.

Ha-Thuc V, Dutta A, Mao R, Wood M, Liu Y (2020) A counterfactual framework for seller-side a/b testing on marketplaces. *Proc. 43rd Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (Association for Computing Machinery), 2288–2296.

Hojjat A, Turner J, Cetintas S, Yang J (2017) A unified framework for the scheduling of guaranteed targeted display advertising under reach and frequency requirements. *Oper. Res.* 65(2):289–313.

Hsu D, Kakade SM, Zhang T (2014) Random design analysis of ridge regression. *Foundations Comput. Math.* 14(3):569–600.

Imbens GW, Rubin DB (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge University Press, Cambridge, UK).

Johari R, Li H, Weintraub G (2022) Experimental design in two-sided platforms: An analysis of bias. *Management Sci.* ePub ahead of print January 25, https://doi.org/10.1287/mnsc.2021.4247.

Li X, Sun C, Ye Y (2020) Simple and fast algorithm for binary integer and online linear programming. *Adv. Neural Inform. Processing Systems* 33:9412–9421.

Liu M, Mao J, Kang K (2021) Trustworthy online marketplace experimentation with budget-split design. *Proc. 27th ACM SIGKDD Conf. Knowledge Discovery Data Mining* (Association for Computing Machinery), 3319–3329.

Nambiar M, Simchi-Levi D, Wang H (2019) Dynamic learning and pricing with model misspecification. *Management Sci.* 65(11):4980–5000.

Nesterov Y (2014) *Introductory Lectures on Convex Optimization: A Basic Course*, vol. 87 (Springer Science & Business Media, New York).

Pouget-Abadie J, Aydin K, Schudy W, Brodersen K, Mirrokni V (2019) Variance reduction in bipartite experiments through correlation clustering. *Adv. Neural Inform. Processing Systems* 32:13309–13319.

Rolnick D, Aydin K, Pouget-Abadie J, Kamali S, Mirrokni V, Najmi A (2019) Randomized experimental design via geographic clustering. *Proc. 25th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (Association for Computing Machinery), 2745–2753.

Schwartz EM, Bradlow ET, Fader PS (2017) Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Sci.* 36(4):500–522.

Simchi-Levi D, Xu Y (2021) Bypassing the monster: A faster and simpler optimal algorithm for contextual bandits under realizability. *Math. Oper. Res.* 47(3):1904–1931.

Terwiesch C, Olivares M, Staats BR, Gaur V (2020) OM Forum—A review of empirical operations management over the last two decades. *Manufacturing Service Oper. Management* 22(4):656–668.

Vartak M, Thiagarajan A, Miranda C, Bratman J, Larochelle H (2017) A meta-learning perspective on cold-start recommendations for items. *Adv. Neural Inform. Processing Systems* 30:6907–6917.

Wager S, Walther G (2015) Adaptive concentration of regression trees, with application to random forests. Preprint, submitted March 22, https://doi.org/10.48550/arXiv.1503.06388.

Yang L, Wang M (2020) Reinforcement learning in feature space: Matrix bandit, kernels, and regret bound. *Proc. 37th Internat. Conf. Machine Learn.*, vol. 119 (PMLR), 10746–10756.

Zeng Z, Dai H, Zhang D, Zhang H, Zhang R, Xu Z, Shen ZJM (2021) The impact of social nudges on user-generated content for social network platforms. *Management Sci.* Forthcoming.

Zhang H, Rusmevichientong P, Topaloglu H (2018) Multiproduct pricing under the generalized extreme value models with homogeneous price sensitivity parameters. *Oper. Res.* 66(6): 1559–1570.

Zhang DJ, Dai H, Dong L, Qi F, Zhang N, Liu X, Liu Z, Yang J (2020) The long-term and spillover effects of price promotions on retailing platforms: Evidence from a large randomized experiment on Alibaba. *Management Sci.* 66(6):2589–2609.

Zhou D, Li L, Gu Q (2020) Neural contextual bandits with UCB-based exploration. *Proc. 37th Internat. Conf. Machine Learn.*, vol. 119 (PMLR), 11492–11502.

Zhou G, Zhu X, Song C, Fan Y, Zhu H, Ma X, Yan Y, Jin J, Li H, Gai K (2018) Deep interest network for click-through rate prediction. *Proc. 24th ACM SIGKDD Internat. Conf. Knowledge Discovery Data Mining* (Association for Computing Machinery), 1059–1068.