# Enhancing Position Auctions in Retail Media

Most Recent Version: https://tinyurl.com/JMPSidPrusty


Siddharth Prusty

Duke University

Carl F. Mela

Duke University

Hana Choi

University of Rochester


August 16, 2025


Siddharth Prusty is a Ph.D. candidate at Fuqua School of Business, Duke University, Durham, NC, 27708; email: siddharth.prusty@duke.edu. Carl F. Mela is the T. Austin Finch Foundation Professor and Professor of Marketing at Fuqua School of Business, Duke University, Durham, NC 27708; email: mela@duke.edu. Hana Choi is an Assistant Professor of Marketing at Simon School of Business, University of Rochester, Rochester, NY 14627; email: hana.choi@rochester.edu.


The paper is based on the first author's dissertation. Special thanks to our adtech partner for their support in providing data. Authors thank Srinivas Tunuguntla, Wilfred Amaldoss, Allan Collard-Wexler, and David Brown for helpful comments. All errors are our own.

**Abstract**

Retail media, a fast-growing channel for digital advertising, surpassed $55 billion of ad spend in 2024. A common retail media format involves position auctions, in which advertisers bid for higher placements on a retailer's product listing page. Advertiser bids are combined with a retailer-set quality score to determine the allocation of sponsored slots and the resulting payments. Quality scores boost certain advertisers' positions and reduce their per-click price. Unlike search engine advertising, retail media position auctions can monetize sales commissions as well as clicks. This paper develops a quality score approach to effectively balance these monetization options.

To connect quality scores to retail revenues, the paper develops a structural model linking advertiser bids and revenues to the retailer's quality score choices coupled with a machine learning model of consumer behavior. These models are estimated using auction-advertiser level data from a quality score experiment conducted at a mid-size US based retail marketplace. Policy simulations show that a quality score approach that balances clicks and commissions improves retailer profits by 7% and advertiser surplus by 42% over click-based approaches typically used by retailers, leading to a win-win outcome for both.

**Keywords:** retail media, sponsored listings, generalized second price auctions, quality scores, mechanism design, envy free equilibria, position effects.

## 1 Introduction

Retail media, or digital advertising operated by retail marketplaces (e.g. Amazon, Walmart), exceeded $55 billion in advertising spend in 2024, accounting for nearly 23% of all digital ad spend.[1] Driven by these systematic shifts in ad spending, nine of the ten largest retailers in the US, including Amazon, Walmart, Target, and Kroger, have established their own retail media networks.[2]

Digital retail media advertising comprises two forms; sponsored listings and display advertising. Our emphasis lies on the former, which accounts for 62% of retail media ad spend.[3] Sponsored listings enable advertisers to increase the visibility of their products on the retailer's product listing page, thereby enhancing the likelihood that their products will be noticed. Sponsored listings are sold via a position auction, in which advertisers bid for favorable positions. Advertisers are not directly ranked by their bids, but by the descending order of their bid rank, which is the product of their bid and a quality score. For example, if an advertiser bids a dollar, and their quality score is two, then the effective advertiser bid (or bid rank) is two dollars. Consequently, quality scores serve as a means for the retailer to put "a foot on

[1] https://www.emarketer.com/chart/261200/us-digital-retail-media-ad-spending-2019-2024-billions-change-of-digital-ad-spending.

[2] https://www.forbes.com/sites/daviddoty/2022/04/26/walmart-target-and-other-mega-retailers-leverage-first-party-data-to-become-new-media-giants/

[3] https://www.emarketer.com/insights/guide-retail-media/.

the scale" of the advertiser's bid. Retailers favor advertisers whose products are more likely to be clicked or sold, as retail profits are tied to these two consumer actions. Despite the growth and popularity of this advertising format, research is scarce on how retailers should set quality scores to improve their total profits (the sum of revenues from clicks and purchases) in retail position auctions. This paper aims to fill that gap.

Position auctions (sometimes called sponsored listings) in retail media differ from standard position auctions in search advertising (Yao & Mela, 2011; Narayanan & Kalyanam, 2015; Ghose & Yang, 2009; S. Yang & Ghose, 2010; Amaldoss et al., 2016; Berman & Katona, 2013; Katona & Zhu, 2017), as monetization accrues through both clicks and purchases, not just clicks. For example, in search engine advertising, Google chooses quality scores based on click-through rate, relevance to search query, and landing page quality, as Google seeks to enhance the consumer experience and increase click revenue.[4] However, a retail media platform that focuses solely on click revenues overlooks downstream revenues from commissions.

To shed light on how click and commission revenue affect quality scores in retail media, we collect auction-advertiser-level data from a mid-sized retail marketplace based in the US and combine this data with product listing information scraped from the website. The auction data includes experimental variation in quality scores within a product, which helps identify rank effects, central to understanding the efficacy of position auctions. We use these data to estimate a structural model of advertiser bidding behavior and a machine learning model of consumer clicking and purchasing behavior. The advertiser model extends the literature on search engine position auctions (e.g., Edelman et al., 2007; Varian, 2007; Athey & Nekipelov, 2011; Lahaie & Pennock, 2007) to the context of retail media. The consumer model uses deep learning to predict clicks and gradient-boosted decision trees to predict purchases.

We find that a quality score rule that balances click and commission revenue can improve total retailer profits by 7% over click-based approaches and improve advertiser surplus by around 42%, leading to a win-win for retailers and advertisers. One reason a click-only based quality score rule leads to worse overall (combined retailer and advertiser) outcomes is that it intensifies advertiser competition, lowering advertiser outcomes. Moreover, such a rule ignores potential retailer gains from high-commission products. In contrast, a commission-only quality score reduces the likelihood of a click on an ad, lowering overall revenue for the retailer and advertiser. This paper shows that a careful balance of the two revenue sources can improve outcomes for all.

[4]https://support.google.com/google-ads/answer/6167118?hl=en. Also, https://www.semrush.com/blog/quality-score/. Note that landing page quality is not important in a retail media context since clicking on sponsored listings in retail media does not redirect a consumer to outside the retail website.

Our research advances two major research streams, one related to advertiser behavior in position auctions and the other related to publisher behavior. We discuss each stream in depth next, but note that the stream on advertiser research does not focus on the publisher problem, and the stream on publisher behavior focuses on concerns other than position auctions in retail media.

Examples of the first stream include Edelman et al., 2007; Varian, 2007; Lahaie, 2006; Yao & Mela, 2011; Amaldoss et al., 2015; Katona & Zhu, 2017; Berman & Katona, 2013; Lu et al., 2015. Building on Edelman et al. (2007) and Varian (2007), we develop a structural model of advertiser bidding and use it to predict advertisers' response to changes in quality scores. Furthermore, much of this research is theoretical, while we assume an empirical focus.

Regarding the second stream, much of the theoretical literature in retail media has considered different strategic retailer decisions, such as information design (Long et al., 2022), self-preferencing (Long & Amaldoss, 2024), reselling format (Abhishek et al., 2016, Long, 2023), or endemic and non-endemic ads (W. J. Choi et al., 2024). Related, the empirical literature has considered the monetization of consumer search behavior on the platform (H. Choi & Mela, 2019), the effect of sponsored advertising on organic rankings (J. Yang et al., 2024; Moshary, 2025), consumers' perceptions of sponsored products (Sahni & Zhang, 2023; Joo et al., 2024), and network effects (Farronato et al., 2024). Most closely related to our research, Athey and Nekipelov (2011), Lahaie and Pennock (2007), Kim and Pal (2024), and Ghili et al. (2025) address a search engine's problem of setting quality scores. However, these papers abstract from the consideration of commission revenue because search advertising is not transaction-based. Overall, prior research (both theoretical and empirical) has not addressed the issue of improving quality scores in retail media position auctions.

The remainder of this paper proceeds as follows. Section 2 describes the institutional setting and data. Section 3 conducts descriptive analyses to link quality scores to consumer behavior, advertiser behavior, and retailer's revenue. Section 4 develops a structural model of advertiser bidding and uses machine learning to predict consumers' click and purchase decisions. Section 5 conducts policy simulations and quantifies the retailer's profit and advertiser surplus under alternative quality score mechanisms. This section shows that simple quality score enhancements can improve retailer profit and advertiser surplus. Section 6 concludes the paper.

## 2 Empirical Setting and Data Structure

This section outlines the empirical application to illustrate the general institutional details behind quality score auctions and to provide a general overview of the specific context used in our analysis. Specifically, this section begins by describing the data, the empirical setting, and a quality score experiment. Next, this information is used in a series of descriptive analyses intended to characterize how the experimental variation in quality scores affected advertiser and publisher outcomes. The section concludes by using this analysis to demonstrate the potential for enhanced quality scores to improve the retailer's profit prior to motivating the use of a structural model.

### 2.1 Setting

Data are sourced from a mid-size US based online retail marketplace. Upon a consumer's arrival and search at the site, the retailer renders a product listing page. The product listing page includes both sponsored and organic listings, where sponsored positions are sold via a position auction. As consumers are more likely to observe products that are ranked near the top of a page, winning impressions in a position auction increases clicks and purchases. Sponsored slots correspond to a set of top rows of the product listing page.[5] We abstract from organic listings because i) the information on these products is not available from the retailer, and ii) advertised products rarely appear in the organic listings (meaning the two types of listings are largely independent).[6] Upon seeing the product listing page, consumers can filter items by type (e.g., men, women) or category (e.g., shoes) or sort them (e.g., by price). Each time a consumer visits the site or selects a search filter, a product listing webpage loads, and a position auction is triggered. When consumers use search filters, the set of eligible items for the ad auction is restricted to the filtered type or category. Sponsored listings are removed when consumers sort products; hence, sorting is considered a non-purchase decision. Thus, our unit of analysis is the auction of a set of slots on the product listing page as rendered to consumers.

A generalized second price (GSP) quality score auction is used to allocate vendor products to sponsored positions and determine advertiser payments. Advertisers bid on a pay-per-click basis, and a quality score is applied to that bid. The quality score upweights the bids of preferred vendors, so that those who are more likely to sell or receive clicks appear higher on the page, thus generating more revenue

[5]Given the data-sharing agreement, we can not provide screenshots of the retail website. The retail setting runs an eligibility auction and the maximum number of sponsored listings is 6. If the number of eligible advertisers for an auction is less than 6, the number of sponsored listings rendered in that auction decreases accordingly.

[6]Using the Wayback Machine, we collect daily snapshots of product listings from 90 webpages on the retailer's website corresponding to our data period. We find that only 0.19% of the 4022 products that we collect appeared in both sponsored and organic ranks (see Appendix A).

for the retailer. The product of the advertiser's bid and quality score is defined as *bid rank* and is used to determine the auction's winners (positions). For example, if advertiser 1 bids \$0.50 per click with a quality score of 5, and advertiser 2 bids \$1.00 per click with a quality score of 2, then advertiser 1's bid rank is $\$0.50 \times 5 = \$2.50$ and advertiser 2's bid rank is $\$1.00 \times 2 = \$2.00$. Because the auction participants are ranked by descending bid rank, advertiser 1 is ranked above advertiser 2. When bid ranks are tied, the marketplace breaks the ties randomly.

Payments in a GSP auction are determined by the minimum bid that would have won the current position. In the above example, advertiser 1 needs to bid at least \$0.40 per click to achieve high enough bid rank to win the position over advertiser 2 ($\$0.40 \times 5.0 \geqq \$1.00 \times 2.00$). Thus, advertiser 1, although they bid \$0.50, would only pay \$0.40 per click according to GSP rules. In general, the cost-per-click (CPC) for an advertised product $j$ at position $\ell$ with quality score $q_j$ is $cpc_{j,\ell} = \min_p\{pq_j > b_{\ell+1}q_{\ell+1}\} = \frac{b_{\ell+1}q_{\ell+1}}{q_j}$, where $b_{\ell+1}$ and $q_{\ell+1}$ denote the bid and the quality score of the advertiser at position $(\ell + 1)$.

The retailer earns revenues from position auctions in two ways. First, any click received on sponsored products generates a cost-per-click based on the GSP rule just described, and this revenue is referred to as *advertising revenue*. Second, the retailer earns a commission on a purchase (sale), and henceforth this revenue is referred to as *commission revenue*. Commissions from sales are accrued as a percentage of the selling price, which is determined through a negotiated revenue-sharing agreement between the marketplace and each seller. A uniform commission fee, $f$, is assumed for all products in our empirical setting. This fee has been disclosed to the authors, but cannot be shared due to confidentiality arrangements.

Advertisers use an ad buying tool provided by the platform to launch ad campaigns. Advertiser campaign decisions include, for example, campaign start date, the set of products to advertise, and bids.[7] In our empirical setting, an ad campaign can contain multiple sponsored products promoted by the same advertiser. Using a multilevel model to decompose the variance in bids across ad campaigns in the marketplace, we find that 17% of the variation in bids is explained between advertisers, and 77% of the variation in bids is explained between campaigns within the same advertiser. Thus, although it is possible to set different bids across products in a campaign, advertisers typically choose campaign-specific bids.

## 2.2 Data Structure

Auction data are collected from the retailer's adtech partner firm that conducts advertising auctions on the retail marketplace, and these data are coupled with product data collected from the retail partner. The details of data organization are as follows.

[7]The use of advertising buying tools is common in industry, for example, see https://developers.google.com/google-ads/scripts/docs/start.

1. **Auction Data File.** A position auction is triggered for each consumer impression (site page visit). We obtain a sample of auction data that spans 45 days, from November 7 to December 21, 2022. This sample comprises 27.6M rows, with each observation representing a winning advertiser (sponsored slot) in the 11.37M position auctions conducted over 45 days.[8] For each auction and slot (row of observation), the fields include the auction id, advertiser id, campaign id, product id, winning bid, quality score, bid rank ($bid \times QS$) and per click payments. Additional characteristics are also available, including search filters used by the consumer for that ad auction, and the date and hour of the auction. The ranks of winning advertisers (sponsored positions resulting from winning the auctions) are imputed by sorting products in descending order of bid rank in each position auction. The data contains 28 advertisers running 74 campaigns, with a total of 7.15k unique products winning an auction over 45 days. The data also contains the binary consumer click decision for every row (auction-advertiser). In total, consumers generate 102k clicks over 5354 unique products in 11.37M auctions.

2. **Conversion Data File.** The conversion data file records purchase events between November 7, 2022, and March 22, 2023. The marketplace uses a 90-day last-touch attribution model to attribute conversions to impressions; thus, data are collected for 45 days spanning the auction data and 90 days after. The fields include advertiser id, campaign id, product id, units purchased, and the purchased item's price. Upon merging with auction data, a total of 1803 purchase events are recorded across 604 unique products that were sold in the 11.37M ad auctions.

3. **Campaign Data File.** This file contains details of ad campaigns that are active between November 7, 2022 and December 21, 2022, comprising 286.6k observations with each row corresponding to an advertiser-campaign-product-date. The fields include advertiser id, campaign id, product id, campaign name, campaign start and end date, and bids (winning and losing), specified at the daily level for the duration of the campaign. Over 45 days, there are 28 advertisers promoting 8,028 unique products, including those products that never win an ad auction.

4. **Product Data File.** This file contains detailed product attributes for 326k products (rows), corresponding to all sponsored products on the retail website for the entirety of 2022. The fields include product id, full product name, product category, brand, product score (average star rating of products), historical product sales, and price.

[8] The number of positions in each auction varies between 1 to 6 in our setting. On average, each position auction contains 2.4 slots ($= 27.6M/11.37M$) that are auctioned off. Our data contain winning advertiser for each slot in each ad auction.

Linking quality scores to advertisers' bids and retailer's profits is enabled by combining auction and conversion data files. The product data file, coupled with consumer outcomes from auction data, is used to estimate the consumer model. As bids are often endogenous, exogenous variation in ranks is required to link quality scores to market outcomes. Next, we outline the variation in the data that enables causal identification of position effects.

## 2.3 Data Variation

We seek to ascertain how quality scores drive consumer and advertiser outcomes, therefore, the data are chosen to ensure that there is exogenous variation in quality scores. During the auction data period, the retail marketplace conducted a quality score experiment. In this experiment, the retailer randomized, at the auction level, whether a given position auction is a rank-by-bid auction (control auctions) or a quality score auction (treated auctions). Specifically, for all products in the treated auctions, the retailer assigned quality scores proportional to their forecast of the click-through rate of the product, while quality scores were uniformly assigned as 1.0 for all products in the control auctions. The auction data contains all advertising auctions during the experimental period, as well as a similar timeframe before and after the experiment. Figure 1 portrays the daily mean and variance in quality scores on dates in our sample. The figure suggests three phases in which the quality score was varied, which we detail below.

- **Pre-experimental period.** From November 7 to November 22, 2022, the platform used a rank-by-bid rule with a quality score of $QS = 1.0$ for all advertised products.[9] This pre-experimental period consists of 3.62M auctions over a period of 16 days.

- **Quality score experiment period.** The experiment was conducted between November 23 and December 8, 2022, comprising $4.65M$ auctions over 16 days. Randomly selected impressions were assigned to the treatment condition, where the retailer assigned quality scores based on click-through-rate (CTR) forecasts. The control impressions were assigned a quality score of 1.

- **Post-experimental period.** Commencing on December 9, 2022, 13 days of post-experimental data (until December 21, 2022) are collected, consisting of $3.1M$ auctions. These data do not contain exogenous variation in quality scores, and are used for model validation and policy simulations.

In sum, the context and the data contain exogenous variation in ranks due to a quality score experiment. Next, we examine how market outcomes are influenced by changes in quality scores.

[9]Figure 1 suggests some variation in quality scores for Nov 14, 16 and 17. We verify that this was caused by $QS \neq 1$ for a very small number of products as a result of one-off production testing incidents.

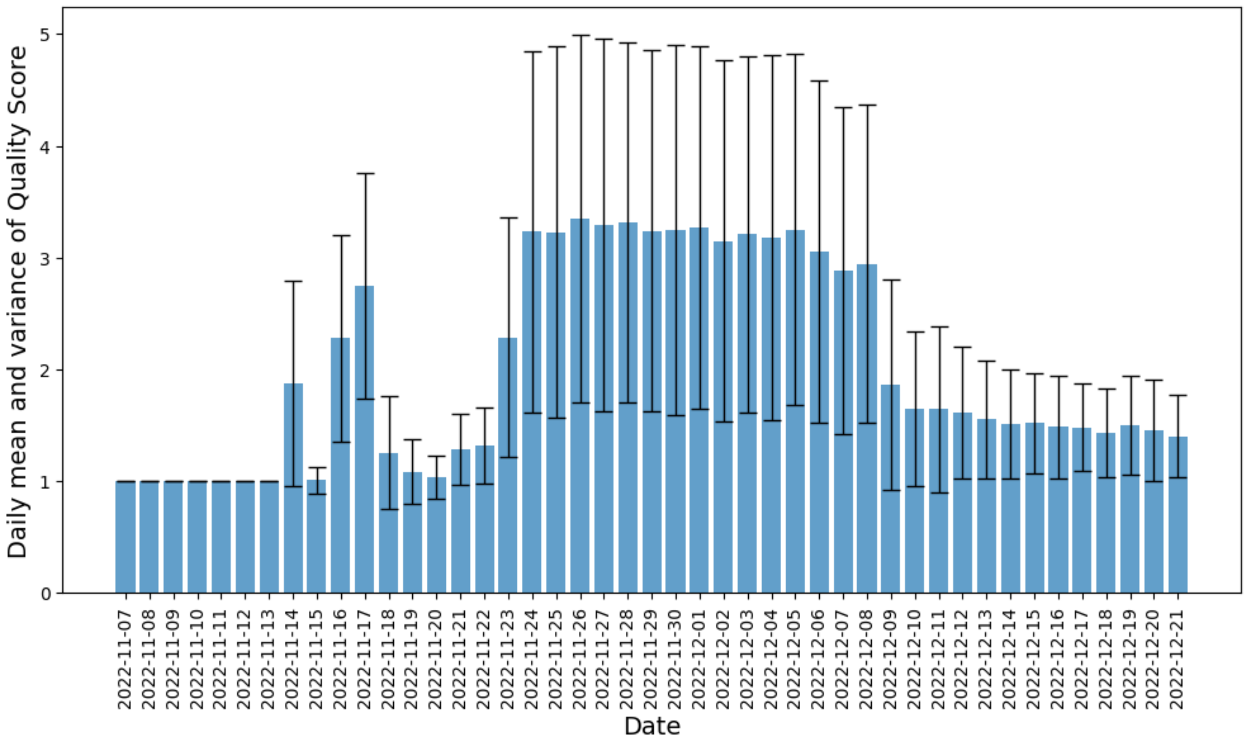


Figure 1: Mean and standard deviation of quality score variation across days. Observations are averaged over all products and auctions, calculated daily ($N = 45$).

# 3 Descriptive Analyses

This section presents descriptive analyses to offer insight into how changes in quality score affect consumer behavior, advertiser behavior, and platform revenues. Section 3.1 considers consumer outcomes. It finds that quality scores drive ranks (impressions), impressions drive clicks, and clicks drive purchases, with each step mediating the preceding step. Next, Section 3.2 shows that advertisers make bidding decisions at the level of a campaign, and bids correlate with impressions won. Finally, Section 3.3 maps quality score variation to advertising revenues (revenue from clicks) and commission revenues (revenue from purchases) earned by the retailer. The data imply that quality scores have a material effect on the retailer's profits.

## 3.1 Effect of Quality Scores on Consumer Outcomes

In this section, we analyze how quality scores affect impressions, clicks, and sales. Our estimates are based on the data corresponding to the quality score experiment (November 23 to December 8, 2022), which contains exogenous rank variation that allows causal estimation of position effects. We first investigate how quality scores influence impressions won by a product.

### 3.1.1 Impressions

This section assesses whether quality scores drive impressions (defined as winning an exposure in one of the top 6 sponsored positions) and whether impressions mediate clicks.

| $\log(I_{jt\tau})$ | (1) | (2) |
|---|---|---|
| Quality score | 0.0704** (0.0337) | × |
| Relative quality score | × | 0.1525** (0.0721) |
| Product Fixed Effects | $Y$ | $Y$ |
| Hour Fixed Effects | $Y$ | $Y$ |
| Date Fixed Effects | $Y$ | $Y$ |
| Observations | 723313 | 723313 |
| $R^2$ (Adjusted) | 0.489 | 0.481 |

Table 1: Effect of quality score on impressions won by a product. Observations are at the product-hour-treatment level. $I_{jt\tau}$ denotes the total (summed) impressions won by product $j$ in hour-date $t$ for treatment cell $\tau \in \{Treated, Control\}$. Standard errors are clustered by advertisers (significance increases when clustering by campaigns or products). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The first column in table 1 reports the results of a regression of the log-transformed number of impressions on quality scores, controlling for product and time fixed effects.[10] The second column reports results of a regression where the independent variable is changed to the quality score of a product relative to the average quality score of all other products in treated auctions at the same time. This column captures the effect of changes in quality scores relative to competing products on the platform.[11] In both columns, the effect of quality scores on impressions remains positive and significant, suggesting that increasing the quality score for a product, all else equal, significantly increases impressions won by the product.

Next, columns 1 and 2 of Table 2 report the results of a regression of quality scores on clicks, controlling for product and time fixed effects. The first column considers the absolute value of quality scores, while the second considers quality scores relative to other advertised products. Both columns show that an increase in quality score significantly increases the clicks received by a product. As clicks cannot occur without impressions, the effect of quality scores is mediated by impressions.

[10] The log-regression provides a percentage change interpretation to the independent variable. The median treated product exhibits two unique quality score values per day, so there is little variation in quality scores within an hour for a product in treated auctions. Hence, the outcome variable (impressions won) is counted at the product-hour-treatment level.

[11] Quality score is a relative score between products. For instance, if the quality score of all products increases by a fixed multiplicative factor, then all consumer outcomes remain unchanged. Thus, the relative quality score of a product captures how high the quality score for a product is compared to other products at the same time (hour-date) for treated cells. In control cells, the relative quality score is 1.0 for all products.

| $C_{jt\tau}$ | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Quality score | 0.0114*** | × | 0.0035** | × |
| | (0.0027) | | (0.0014) | |
| Relative quality score | × | 0.0070* | × | −0.0001 |
| | | (0.0037) | | (0.0025) |
| Impressions won ($1k$) | × | × | 2.8847*** | 2.9178*** |
| | | | (0.5663) | (0.5790) |
| Product Fixed Effects | $Y$ | $Y$ | $Y$ | $Y$ |
| Hour Fixed Effects | $Y$ | $Y$ | $Y$ | $Y$ |
| Date Fixed Effects | $Y$ | $Y$ | $Y$ | $Y$ |
| Observations | 723313 | 723313 | 723313 | 723313 |
| $R^2$ (Adjusted) | 0.102 | 0.097 | 0.186 | 0.185 |

Table 2: Effect of quality scores on clicks received by a product. Observations are at the product-hour-treatment level. $C_{jt\tau}$ denotes the total (summed) number of clicks on product j on hour-date t for treatment cell $\tau \in \{Treated, Control\}$. Standard errors are clustered by advertiser. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

The last two columns add a term for impressions to the regression, and the effect of absolute quality score reduces in magnitude (column 3), while that of relative quality score becomes insignificant (column 4). This is consistent with impressions mediating the effect of quality score on clicks.

#### 3.1.2 Clicks

**Identification of Position Effects.** This section aims to infer the causal impact of positions on clicks received by a product. Recall that the retail platform conducted a quality score experiment, in which each auction ranked products either by descending *bid* (control) or by descending $bid \times QS$ (treatment). Exogenous assignment of impressions to each auction condition leads to exogenous variation in ranks within the product, allowing us to causally estimate the effect of quality score on clicks and purchases.

Although there is exogenous variation in ranks within the same product, our data suffer from selection on products. This is because products with high quality scores are promoted in treated auctions, while products with low quality scores are demoted in the same auction, but not vice versa. Consequently, in the estimation of rank effects, the identifying assumption is of separability between rank and product effects. That is, any product's click probability, at any rank, is separable into a product effect and a position effect. This assumption allows causal estimation of position effects by leveraging exogenous variation in ranks within a product but across experimental cells.

**Estimates.** Table 3 reports the results of a linear probability model estimating the effect of rank, determined by sorting products in descending bid rank ($bid \times QS$), on clicks, controlling for product and time fixed effects. In the first column, the coefficient of quality score is insignificant, suggesting that

| Click(0/1) | (OLS) | (IV) |
|---|---|---|
| *Rank* | $-0.00073^{***}$ | $-0.00072^{*}$ |
| | (0.00019) | (0.00041) |
| Quality Score | 0.00000 | $\times$ |
| | (0.00001) | |
| Product Fixed Effects | $Y$ | $Y$ |
| Date Fixed Effects | $Y$ | $Y$ |
| Hour Fixed Effects | $Y$ | $Y$ |
| Observations | 11.6M | 11.6M |
| $R^2$ (Adjusted) | 0.0046 | 0.0046 |

Table 3: A linear probability model for the effect of rank on clicks. We infer ranks by sorting products in descending bid rank. In case of tied bid rank, rank variable assumes the the minimum possible rank for all tied products. Standard errors are clustered by advertisers. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

ranks mediate the effect of quality score on clicks.

In the second column, we adopt an instrumental variable (IV) strategy with quality scores as instruments for ranks. The quality score is a relevant instrument as differences in ranks for the same product across experimental cells is attributable to exogenous differences in quality score between treated and control auctions. Furthermore, as ranks mediate the effect of quality score on clicks (see first column), then conditioning on ranks, quality scores are assumed to be excludable from clicks. Both models indicate that ranks significantly affect clicks and that a unit increase in ranks (causally) decreases the probability of receiving a click by 0.07%.[12] Hence, conditional on winning an impression, advertisers have an incentive to appear higher in rankings.

### 3.1.3 Purchases

Next, we consider whether quality scores can drive purchases. Because the retail marketplace monetizes clicks and purchases, both outcomes are central to enhancing quality scores.

Table 4 reports the effect of quality scores, ranks, and clicks, on purchases. The first column regresses both rank and quality scores as separate explanatory variables, while the second column uses quality scores as instrumental variables for ranks. In both columns, the effect of rank is insignificant, while that of click is significant, meaning that clicks mediate the effect of ranks on purchases. Receiving a click increases the probability of a purchase by 2.4%. The finding that ranks do not improve purchases (sales), conditional on clicks, is in line with previous literature (e.g., H. Choi & Mela, 2019; Ursu, 2018).

Overall, our results suggest a serial mediation model, with quality scores driving impressions (ranks),

[12] The F-stat for the first-stage regression in the instrumental variable model is 2002, suggesting that quality scores are strong instruments for ranks.

| Purchase(0/1) | (OLS) | (IV) |
|---|---|---|
| *Rank* | 0.0000063 | 0.0000186 |
| | (0.0000056) | (0.0000276) |
| Quality Score | 0.0000003 | × |
| | (0.0000006) | |
| Click(0/1) | $0.023930^{***}$ | $0.023930^{***}$ |
| | (0.005936) | (0.005936) |
| Product Fixed Effects | $Y$ | $Y$ |
| Date Fixed Effects | $Y$ | $Y$ |
| Hour Fixed Effects | $Y$ | $Y$ |
| Observations | 11.6M | 11.6M |
| $R^2$ (Adjusted) | 0.0185 | 0.0185 |

Table 4: Effect of rank (and, clicks) on purchases. We infer ranks by sorting products in descending bid rank. In case of tied bid rank, the rank variable assumes the minimum possible rank for all tied products. Standard errors are clustered by advertisers. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

which in turn generate clicks, ultimately driving purchases.

### 3.2 Advertiser Behavior

In this section, we examine the behavior of advertisers in the marketplace. For confidentiality reasons, bids and prices are scaled by a constant positive factor throughout the paper. Table 5 reports that the average bid for a campaign is around \$0.81, and the average number of products per campaign is 219.

| Campaign characteristics | Mean | Median | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Bid (\$) | 0.814 | 0.774 | 0.060 | 0.585 | 0.99 |
| #Products | 218.6 | 86 | 557.2 | 1 | 4809 |

Table 5: Summary statistics of campaign characteristics. Observations are at campaign level ($N = 76$). All bids are scaled by a constant positive factor.

In our empirical setting, an advertiser could run multiple ad campaigns and promote multiple products in those campaigns simultaneously. Figure 2 reports a scatter plot of bids by advertisers. The median bid is \$0.774, with bids concentrated between \$0.59 and \$0.99. Different bid levels within an advertiser correspond to different campaigns, and the area of the circle is proportional to the number of products in the campaign. Overall, the data exhibit some variation in bids over campaigns; however, there seems to be limited variation between products within a campaign, or across time. This is in line with evidence from our multilevel analysis, which found that 94% of the variation in bids is explained between campaigns (see Section 2). As advertisers choose bids at the campaign level and do not immediately react to quality score changes, the data suggest that quality scores do not influence advertiser behavior within a campaign.

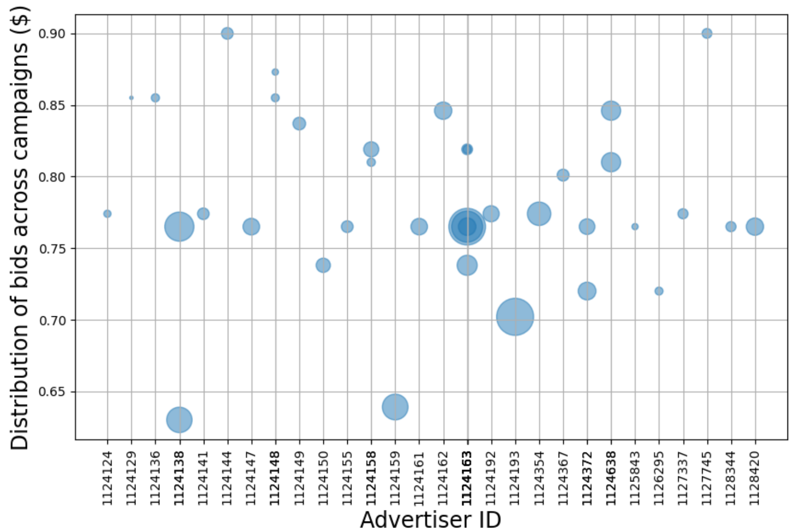


Figure 2: Scatter plot of bids across advertisers ($N = 28$). Different bid levels for the same advertiser denote different campaigns, and the size of the circle indicates the number of products in the campaign. All bids are scaled by a constant positive factor.

### 3.2.1 Bids and Impressions

Next, we explore whether bids drive impressions, because impressions drive clicks and purchases (see Section 3.1). Keeping quality scores fixed, higher bids lead to higher bid ranks, which should lead to higher impressions. Table 6 reports a regression of bids on the daily number of impressions won by a campaign. The table shows that bids exhibit a significant positive correlation with impressions won by a campaign, whether or not controlling for quality scores. Given that impressions drive clicks and purchases, the positive correlation indicates that higher bids lead to increased clicks and purchases at the campaign level. Next, we consider whether consumer outcomes drive revenue for the marketplace.

| $I_{ct}/1000$ | (1) | (2) |
|---|---|---|
| Bid (1$c$) | 1.124** | 1.209** |
| | (0.524) | (0.525) |
| Quality score | × | 3.657 |
| | | (2.492) |
| Date Fixed Effects | $Y$ | $Y$ |
| Observations | 2001 | 2001 |
| $R^2$ | 0.088 | 0.107 |

Table 6: Correlation between bids and impressions won by a campaign. Observations are at the campaign-date level. $I_{ct}$ is the total number of impressions (exposures) won by campaign c on date t. Standard errors are clustered by advertiser. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.


Figure 3: Scatter plot of log(1 + *commission revenue*) versus log(1 + *click revenue*) across sponsored products in retail website. Observations include those products that earn non-zero revenue ($N = 5354$).

### 3.3 Effect of Quality Scores on the Retailer's Revenue Outcomes

In this section, we link consumer outcomes to the retailer's revenue from clicks (ad revenue) and from commissions on purchases (commission revenue). Table 7 shows that ad revenues and commission revenues earned by the retailer have different distributions, at the product level, when aggregating over the 45 days corresponding to position auctions in our data. The lack of a perfect correlation suggests that each component is informative about revenue and could have different implications for choosing quality scores.

| Variables | Mean | Median | Std. Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Ad Revenue ($) | 10.92 | 2.30 | 33.37 | 0.0 | 765.8 |
| Commission Revenue ($) | 3.32 | 0.0 | 22.48 | 0.0 | 658.1 |

Table 7: Summary statistics of total advertising revenue and commission revenue accrued by the retailer. Observations are at the product level, provided the product won at least one auction ($N = 7142$).

Figure 3 shows a scatter plot of commission revenue transformed in logarithmic terms compared to ad revenue converted in logarithmic terms accrued from products. The figure suggests that clicks are more common than purchases. Specifically, of the 7142 products that win an exposure, 5354 earn ad revenue on account of clicks, while only 604 products earn commission revenue on account of purchases. A mild positive correlation ($\rho = 0.47$) between the revenue streams arises because a purchase can only follow a click. The tradeoff between the two sources of revenue suggests that a quality scoring rule based on click-through rates can be further improved. This is a hallmark difference from search advertising, where revenues are based solely on clicks. Next, we examine how quality scores affect the retailer's revenue from clicks and sales.

**Ad Revenue.** Table 8 reports the result of an instrumental variable (IV) regression that estimates the effect of ranks on click payments earned from products. As before, quality scores are used as instruments for ranks, controlling for product and time fixed effects.[13] The first column shows that a unit increase in rank (decrease in position) reduces click payments by 0.076 cents. The second column controls for clicks, and the coefficient of rank becomes insignificant. The coefficient of click suggests that a click event accrues an ad revenue of \$0.75 on average, and that clicks mediate the effect of ranks on ad revenues.

| Click Payment(\$) | (IV1) | (IV2) |
|---|---|---|
| *Rank* | $-0.00076^{**}$ | $-0.00021$ |
| | (0.00028) | (0.00016) |
| Click(0/1) | × | $0.75477^{***}$ |
| | | (0.04964) |
| Product Fixed Effects | *Y* | *Y* |
| Date Fixed Effects | *Y* | *Y* |
| Hour Fixed Effects | *Y* | *Y* |
| Observations | 11.6M | 11.6M |
| $R^2$ (Adjusted) | 0.0034 | 0.9495 |

Table 8: Effect of ranks and clicks on click payments (advertising revenue). Standard errors are clustered by advertisers. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

**Commission Revenue.** Next, we investigate how quality scores and ranks influence commission revenue. Table 9 reports the results of an instrumental variable (IV) approach, estimating the effect of ranks on commission revenues accrued from a product. As before, we use quality scores as instruments for ranks, and use fixed effects to control for product and time. The first column shows that clicks mediate the effect of ranks on commission revenue, which is intuitive since a purchase cannot happen without a click. The second column controls for purchase events and finds that purchases mediate the effect of clicks on commission revenues. Further, a unit purchase event leads to a commission revenue of \$14.24 on average.

Comparing estimates from tables 8 and 9, a unit purchase event leads to a payoff of \$14.24 while a unit click event leads to a payoff of \$0.76. However, purchase events are rare, while click events are more common. This suggests that focusing on click revenues need not guarantee commission revenues, while focusing on commission revenues could decrease click revenues. Thus, it stands to reason that quality score rules that augment clicks and commissions could improve revenue outcomes for the retailer.

[13] Conditional on ranks, quality scores are assumed excludable from clicks (see Section 3.1). Furthermore, quality scores are strong instruments for ranks, as the first-stage of the IV regression leads to an Fstat of 2002.

| Commission Revenue($) | (IV1) | (IV2) |
|---|---|---|
| *Rank* | 0.00061 | 0.00034 |
| | (0.00053) | (0.00023) |
| Click(0/1) | 0.3408*** | −0.00005 |
| | (0.1037) | (0.00012) |
| Purchase(0/1) | × | 14.24*** |
| | | (1.48) |
| Product Fixed Effects | *Y* | *Y* |
| Date Fixed Effects | *Y* | *Y* |
| Hour Fixed Effects | *Y* | *Y* |
| Observations | 11.6M | 11.6M |
| $R^2$ (Adjusted) | 0.0177 | 0.6485 |

Table 9: Effect of ranks, clicks, and purchases on commission revenue. We infer ranks by sorting products in descending bid rank, and calculate the commission revenue assuming an uniform commission fee ($f$) for all products. Standard errors are clustered by advertisers. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

### 3.4 Summary of Descriptive Analyses

This section uses exogenous variation in quality scores to estimate the causal effect of ranks on clicks and sales. All else equal, quality scores influence ranks, ranks drive clicks, and clicks drive sales, with each step mediating the previous one. Next, it is shown that clicks lead to ad revenue while sales lead to commission revenue. As both revenue components are substantively distinct, it is possible that an augmented quality score approach could improve retailer's total profit from position auctions.

## 4 Model

### 4.1 Model Overview

The modeling section characterizes the game played by the retailer and advertisers and organizes its discussion by the steps in this game. Players in this game include: i) the retailer, ii) the advertisers, and iii) the consumers, whose behavior is taken as given. Below, we characterize the timing of the game and each player's objective function, information set, and decision variables.

1. **Retailer.** The retailer moves first and assigns a quality score, $q_j$, to each advertiser $j$. The retailer's objective is to maximize its expected revenue from clicks and commissions. We assume that the retailer knows the click-through rate of each advertiser $j$ at every position $\ell$ (that is, $\alpha_{j,\ell}$), the probability of conversion given a click ($\gamma_j$), and the price of each sponsored product ($p_j$). The retailer and all advertisers are assumed to be risk neutral. The retailer-set commission fee ($f$) and

reserve price ($r_p$) are assumed to be fixed apriori, and considered exogenous to the model.[14]

2. **Advertisers.** Advertisers move second. Each advertiser $j$ decides their bid, $b_j$, based on their per-click value $v_j^c$, their quality score $q_j$, their costs (click and commission payments), and the valuation and quality score of competing advertisers. Advertisers' objectives are to maximize their click value net of advertising costs. Advertisers know their click-through rate at each position ($\alpha_{j,\ell}$) as well as their conversion rate conditional on click ($\gamma_j$). Regarding competition, we follow prior literature and assume a complete information setting (e.g., Varian, 2007; Edelman et al., 2007; Athey & Nekipelov, 2011; Lahaie & Pennock, 2007), in which advertisers know all competing valuations and quality scores.

3. **Consumers.** Consumers move last. We model consumer search arrivals as an exogenous process. Each consumer's probability of clicking on advertiser $j$ in position $\ell$ is given by the probability $\alpha_{j,\ell}$. Similarly, $\gamma_j$ is the average probability that a consumer purchases from advertiser $j$, after clicking on its product.

Following a backward induction order, we first outline the consumer model (Section 4.2). Conditional on consumer response, we outline the advertiser model (Section 4.3). Conditional on advertiser and consumer response, the retailer sets the quality score rule. The discussion of the retailer's problem is relegated to policy simulations (Section 5.1).

## 4.2 Consumer Model

**Approach.** Machine learning is used to project product observables to consumer click and purchase decisions as a function of product ranks. The consumer predictive model is assumed to be invariant to changes in quality scores as the effect of quality scores is mediated by rank. The assumption seems reasonable because consumers are unaware of quality scores, but do observe ranks.

**Separability of Click Rates.** In line with most prior literature on position auctions, we assume separability of click-through rates (e.g., Athey & Nekipelov, 2011; Edelman et al., 2007; Varian, 2007; Lahaie & Pennock, 2007). Further, better positions are assumed to lead to more clicks, all else equal. The latter assumption is supported by Section 3.1.2 which shows that the probability of a click decreases with rank, controlling for products. To formalize these assumptions, let $\mathcal{J}$ denote the set of advertisers (products) on the platform, with $|\mathcal{J}| = J$, and let $L$ represent the maximum number of ranks in an auction. Then, the assumption on click-through rates (henceforth, CTR) can be summarized as follows.

[14] A uniform commission fee of $f$, supplied by the retailer, is applied to all products in the empirical setting. The reserve price for all position auctions is fixed at $r_p = \$0.50$.

**Assumption 1.** *Click through rate* $\alpha_{j,\ell}$ *is multiplicatively separable into an advertiser component* ($\alpha_j$) *and a position component* ($\lambda_\ell$)*, that is* $\alpha_{j,\ell} = \alpha_j \lambda_\ell$ *for all* $j \in \mathcal{J}$*, and* $1 \leq \ell \leq L$*. Moreover, position effects* ($\lambda_\ell$) *decrease with positions, that is,* $\lambda_\ell > \lambda_{\ell+1}$ *for all* $1 \leq \ell \leq L$*.*

Because unranked products do not receive clicks, $\lambda_\ell = 0$ for all $\ell > L$. Moreover, because position effects are relative, the first position effect is normalized to $\lambda_1 = 1$. Thus, $\alpha_j$ is the product's CTR when ranked in the first position.

**Attributes.** The following product attributes are used as explanatory variables in the prediction task: {*Price, Historical Sales, Product-score, Brand, Gender-age-group (ga), Others*}.[15] Details regarding variable pre-processing and feature selection are presented in Appendix B.

**Training, Validation, and Simulation Data.** Auctions before and during the quality score experiment (November 7, 2022 to December 8, 2022), are used for model training, while auctions after the quality score experiment (December 9 to December 21, 2022) are used for model validation. Splitting data in this manner ensures that the training data contain exogenous variation in ranks due to the quality score experiment. The validation data are used for parameter tuning and probability calibration. A subset of the validation data is used for policy simulations (see Section 5.2).

**CTR Prediction.** A *deep learning model* is estimated to forecast the CTRs of products. Model training includes product attributes and ranks. To predict product-specific CTR we set $Rank = 1$, a widely adopted practice for removing off-policy training bias from positions in recommendation systems (e.g., Wang et al., 2024). Specifically, for each product $j$, we estimate $\alpha_j = P(Click = 1|X_j, Rank = 1)$, where $X_j$ are product attributes. Our training data consists of $19.5M$ rows ($8.27M$ position auctions) accruing $77k$ clicks. Consequently, the average CTR in the training set is 0.40%, meaning that the data are imbalanced toward non-events. To increase the reliability of the imbalanced data predictions: (1) rarer observations (clicks) are up-weighted when calculating empirical loss, and (2) area under the receiver operating curve (AUC-ROC) rather than prediction accuracy is used as a measure of model performance. Unlike prediction accuracy, AUC-ROC remains informative even when classes are severely imbalanced, thereby guiding the model toward a better generalization for rare events (Haibo He & Garcia, 2009).

*Network Architecture and Implementation.* A bottleneck network architecture is implemented with a depth of 5 layers ($X_j \rightarrow 32 \rightarrow 128 \rightarrow 32 \rightarrow o/p$), as it enables efficient learning of representations from high-dimensional data without requiring too many hidden layers (e.g., Ba & Caruana, 2013; He et al., 2015). Bottleneck structures constrain model capacity at the first hidden layer, forcing the neural

[15] Price, historical sales, and brand are self-explanatory variables. Product-score denotes the average star rating of products, while gender-age-group denotes the intended product user (e.g., men, women, kids). Others refer to a few other attributes that are used for prediction, but are disguised to protect data confidentiality.

network to learn concise representations of the input features. Subsequent expansions in dimensions (e.g., 32 $\to$ 128) allow the model to capture rich, nonlinear interactions between product attributes, and reducing dimensionality afterward (e.g., 128 $\to$ 32) compels the network to retain only the most salient feature interactions, critical for stable prediction. Hyperparameters such as the number of nodes in the first hidden layer (e.g., 32), class weights for rarer classes, and other regularization parameters (e.g., dropout), are chosen by maximizing AUC-ROC in the validation set. The best choice of hyperparameters achieves a validation AUC-ROC of around 0.66. Further details are relegated to Appendix B.

**CVR Prediction.** The conversion rate (CVR), conditional on click, for each product $j$, is defined as $\gamma_j = P(Purchase = 1|X_j, Click = 1)$. The purchase model is estimated conditional on click, thus the training data consists of $77k$ rows (clicks) and 1,237 purchases, resulting in an average CVR of around 1.6%. Ranks are not included as an explanatory feature because they are insignificant for predicting purchases given clicks (see Section 3.1.3). In other words, once clicked, a product enters the consideration set. The smaller CVR dataset size relative to the CTR data makes deep learning approaches prone to overfitting. Consequently, *gradient-boosted decision trees* (GBDTs) are used to predict $\gamma_j$.

The GBDT algorithm is an ensemble learning method that iteratively builds shallow decision trees (weak learners) and predicts outcomes based on a weighted combination of these learners. GBDTs mitigate overfitting by iteratively optimizing weaker learners, and naturally handle class imbalance by up-weighting misclassified instances in subsequent iterations (e.g., Chen & Guestrin, 2016; Ke et al., 2017). Specifically, at every iteration, the training samples where the algorithm performs poorly are upweighted, and those where it performs well are downweighted, making it an appropriate method for handling class imbalance. Hyperparameters such as class weights, number of leaves, maximum tree depth, and regularization parameters are chosen by maximizing AUC-ROC in the validation set. The best choice of parameters achieves a validation AUC-ROC of around 0.65 (details in Appendix B).

**Probability Calibration.** As the data are highly imbalanced, the implied probability scores of the model tend to overestimate observed click and purchase probabilities for most products. Consequently, raw probability scores from machine learning predictions (that is, CTR and CVR estimates) are calibrated using observed frequencies in the validation data, ensuring that predicted probabilities more accurately reflect actual outcomes. The calibration is performed using the validation set to avoid overfitting.

The procedure is as follows. Given a raw probability score of $\hat{y}$, Platt scaling is used to apply the transformation $P(\hat{y}) = \frac{1}{1+e^{A\hat{y}+b}}$ to fit the true probabilities in the validation data. Platt scaling is widely used to transform model predictions into true posterior probabilities (e.g., Niculescu-Mizil & Caruana, 2005; Platt, 2000). Note that this transformation is monotonic; thus it does not affect the validation AUC-

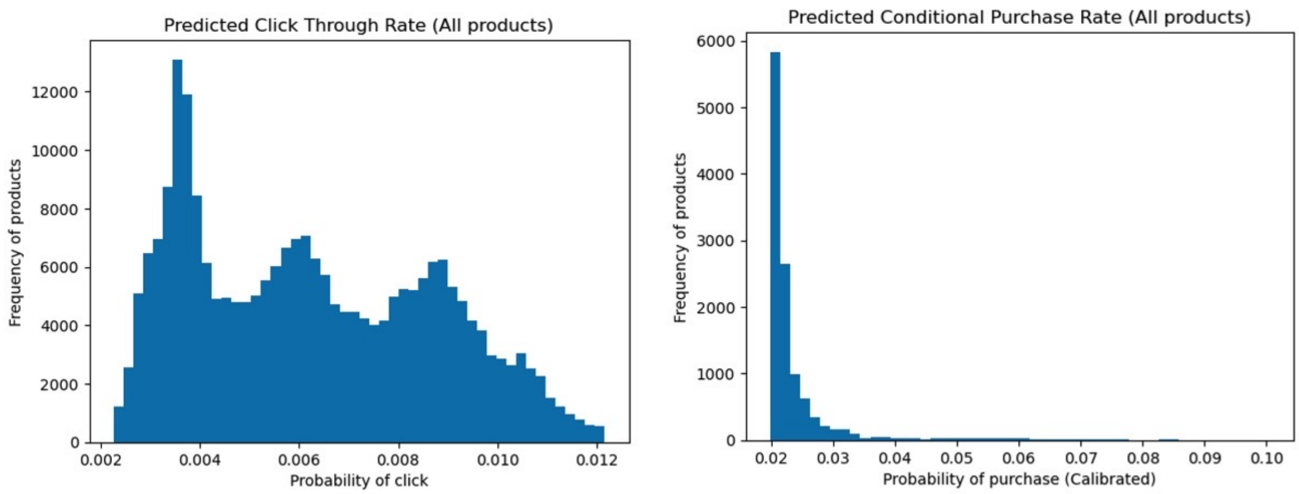


Figure 4: Distribution of (a) predicted click-through rate, and (b) predicted conversion rate, over all sponsored products on retail website ($N = 326k$ products).

ROC.[16] Combining outputs from prediction models (that is, deep learning, GBDT) with the respective probability calibrations, CTR and CVR can be predicted for any product on the retail site.

**Prediction Results.** Figure 4 shows the distribution of the probability of click (CTR) and the probability of purchase (CVR) across all sponsored products on the retail website in our data sample, including products that never won an auction. These predictions are used to compute counterfactual quality scores and perform policy simulations.

### 4.3 Advertiser Model

In this section, we develop a structural model of advertisers' bidding behavior as a function of their valuation, the retailer-set quality score, and competition.

**Click Valuations.** In a generalized second price (GSP) auction, an advertiser is required to submit one bid per auction, although the retailer sells multiple objects (slots). Consequently, bidder valuations can be represented by one-dimensional types, and advertisers are assumed to assign the same value to a click regardless of their position or the consumer viewing the ad (e.g., Edelman et al., 2007). Specifically, each advertiser $j$ assigns a per-click value $v_j^c$ to a click on its product. This per-click value reflects advertisers' expected gains from a click and can reflect sales revenue, margins, and other outcomes. We assume that each advertiser's click valuation is revealed to the retailer (and the econometrician) through their bids. Following prior literature, per-click valuations are assumed invariant to changes in quality score (e.g., Edelman et al., 2007; Varian, 2007; Athey & Nekipelov, 2011; Lahaie & Pennock, 2007).

[16] AUC-ROC is invariant to monotonic transformations of predicted scores, since it depends only on the relative ordering of predictions rather than their absolute values.

**Advertiser Utility.** Advertisers are assumed to have quasi-linear utility. The net utility when advertiser $j$ occupies rank $\ell$ is:

$$U_{j,\ell} = \alpha_j \lambda_\ell (v_j^c - cpc_{j,\ell} - \gamma_j p_j f), \tag{1}$$

where the leading term captures the click-through rate, while the term inside brackets is the advertisers' value net of their costs from a click. Specifically, $cpc_{j,\ell}$ is the per-click price charged to advertiser $j$ at position $\ell$, while $\gamma_j p_j f$ is their expected commission payment. According to the rules of a GSP auction, $cpc_{j,\ell}$ is the minimum price required for the advertiser $j$ to hold position $\ell$, and is written as $cpc_{j,\ell} = \frac{q_{\ell+1} b_{\ell+1}}{q_j}$, where $b_{\ell+1} q_{\ell+1}$ denotes the bid rank of the advertiser at position $(\ell+1)$. Next, $\gamma_j p_j f$ is the expected commission payment if a consumer purchases a clicked product, where $\gamma_j$ is the conversion rate, $p_j$ is the price, and $f$ is the commission fee.

**Complete Information Equilibria.** Following the literature in generalized second price (GSP) position auctions, we assume a complete information setting and consider the pure-strategy Nash equilibrium of the auction game (e.g., Edelman et al., 2007; Varian, 2007; Athey & Nekipelov, 2011; Lahaie & Pennock, 2007; Lahaie, 2006; Katona & Zhu, 2017; Yoon, 2010). This literature notes that the assumption of complete information is a reasonable approximation because all relevant information about bidders is likely to be inferred over time. In the complete information setting, a set of bids characterizes the Nash equilibrium if no advertiser $j$ has an incentive to deviate from their equilibrium position $\ell$. This condition can be written as $U_{j,\ell} \geq U_{j,p}$ for all $p \neq \ell$, or:

$$\alpha_j \lambda_\ell \Big( v_j^c - \frac{q_{\ell+1} b_{\ell+1}}{q_j} - \gamma_j p_j f \Big) \geq \alpha_j \lambda_p \Big( v_j^c - \frac{q_{p+1} b_{p+1}}{q_j} - \gamma_j p_j f \Big), \text{ for all } p > \ell, \text{ and,} \tag{2}$$

$$\alpha_j \lambda_\ell \Big( v_j^c - \frac{q_{\ell+1} b_{\ell+1}}{q_j} - \gamma_j p_j f \Big) \geq \alpha_j \lambda_p \Big( v_j^c - \frac{q_p b_p}{q_j} - \gamma_j p_j f \Big), \text{ for all } p < \ell. \tag{3}$$

To obtain equation (2) for any $p > \ell$, one can start with equation (1), replacing $cpc_{j,\ell}$ with $q_{\ell+1} b_{\ell+1}/q_j$ and $cpc_{j,p}$ by $q_{p+1} b_{p+1}/q_j$, according to the rules of the GSP auction. Conversely, for any $p < \ell$, the advertiser $\ell$ moving up to position $p$ implies that the advertiser previously at position $p$ would now occupy the next position, thus equation (3) replaces $cpc_{j,p}$ by $q_p b_p/q_j$ in this scenario.

For ease of notation, we henceforth refer to $v_j = v_j^c - \gamma_j p_j f$ as an advertiser's *effective valuation*, which is their click valuation after netting out their expected commission payments. To simplify the equilibrium analysis, Edelman et al. (2007) and Varian (2007) propose a refinement of Nash equilibria, which they call the 'envy-free' criterion. In an envy-free equilibrium, the inequality (2) holds for all $p \neq \ell$ rather than only for $p > \ell$. Noting that the advertiser effect ($\alpha_j$) cancels out from both sides of inequality (2), the envy-free equilibrium criterion for any advertiser $j$ can be written as:

$$\lambda_\ell\Big(v_j - \frac{q_{\ell+1}b_{\ell+1}}{q_j}\Big) \geq \lambda_p\Big(v_j - \frac{q_{p+1}b_{p+1}}{q_j}\Big), \ \textit{for all } p \neq \ell. \tag{4}$$

As bids and quality scores are continuous while ranks are discrete, the above inequalities can be satisfied by multiple sets of bids; thus, envy-free Nash equilibria are not unique. Consequently, the equilibrium selection criteria are outlined next.

**Equilibrium Selection.** For the envy-free equilibrium refinement, Varian (2007) demonstrated that it is sufficient to ensure that advertisers do not prefer local position exchanges. That is, an advertiser at position $\ell$ does not prefer to occupy either position $p = \ell - 1$ or position $p = \ell + 1$. Given the generalized second price mechanism, advertiser at position $\ell$ can vary their bid without changing their payments, until they switch to an immediately higher position ($\ell - 1$), or to an immediately lower position ($\ell + 1$).[17] Accordingly, we adopt either of these criteria as natural choices for equilibrium selection, in line with prior literature (e.g., Athey & Nekipelov, 2011; Katona & Zhu, 2017).

Specifically, the equilibrium selection criterion where each advertiser is indifferent between occupying position $\ell$ and the immediately higher position ($\ell - 1$) is referred to as the 'envy-free upper bound', or the *EFUB equilibrium*. Similarly, the equilibrium selection criterion where each advertiser is indifferent between occupying position $\ell$ and an immediately lower position ($\ell + 1$) is referred to as the 'envy-free lower bound', or *EFLB equilibrium*. Because these equilibria correspond to the upper and lower bounds of all envy-free equilibria, we estimate advertisers' valuations and conduct policy simulations at both of these extremes. If our policy insights are robust to the choice of either equilibrium selection, it is likely that our qualitative insights could generalize to any envy-free equilibrium, of which there are infinitely many. Next, we derive the bidding functions for the respective equilibria. Because each advertiser self-selects into an equilibrium rank, each advertiser is indexed by their rank $\ell$ to facilitate exposition.

- **EFUB equilibrium.** The EFUB equilibrium arises when each advertiser $j$ is indifferent between occupying their rank ($\ell$) and the one immediately above ($\ell - 1$). Referring to advertiser $j$ by their equilibrium rank $\ell$ and substituting $p$ by $\ell - 1$, equation (4) can be written as:

$$\lambda_\ell(v_\ell q_\ell - q_{\ell+1}b_{\ell+1}) \geq \lambda_{\ell-1}(v_\ell q_\ell - q_\ell b_\ell). \tag{5}$$

Because the EFUB bid corresponds to the advertiser's indifference between occupying positions $\ell$ and $\ell + 1$, the above inequality holds with equality. Consequently, the EFUB equilibrium bids satisfy the following recursive relationship:

[17] For advertiser in position $\ell$, per-click GSP payment is $\frac{q_{\ell+1}b_{\ell+1}}{q_\ell}$, which is independent of own bid, $b_\ell$, given rank $\ell$.

$$\lambda_{\ell-1} q_\ell b_\ell^{EFUB} = (\lambda_{\ell-1} - \lambda_\ell) v_\ell q_\ell + \lambda_\ell q_{\ell+1} b_{\ell+1}^{EFUB}, \tag{6}$$

where $v_\ell$ denotes the effective valuation (click valuation net commission payments) for the advertiser at position $\ell$. The recursion starts at $\ell = L+1$, with $\lambda_{L+1} = 0$ by assumption 1, giving $b_{L+1}^{EFUB} = v_{L+1}$. That is, it is optimal for the first excluded bidder to bid their effective value (click value net commission payments). Solving the recursion for advertisers that are ranked ($\ell \leq L$), EFUB equilibrium bids for advertiser at rank $\ell$ can be written as the following:

$$b_\ell^{EFUB} q_\ell \lambda_{\ell-1} = \sum_{p=\ell}^{L} (\lambda_{p-1} - \lambda_p) v_p q_p, \ \textit{for all } 2 \leq \ell \leq L. \tag{7}$$

Edelman et al. (2007) and Varian (2007) argue that this equilibrium is most likely to be realized in practice, and show that the click revenues accrued coincide with Vickrey payments.

- **EFLB equilibrium.** Next, the EFLB equilibrium selection arises when each advertiser $j$ is indifferent between occupying their position ($\ell$) and the one below ($\ell + 1$). Referring to advertiser $j$ by their equilibrium rank $\ell$ and substituting $p$ by $\ell + 1$, equation (4) can be written as:

$$\lambda_\ell (v_\ell q_\ell - q_{\ell+1} b_{\ell+1}) \geq \lambda_{\ell-1} (v_\ell q_\ell - q_{\ell+2} b_{\ell+2}) \tag{8}$$

Because the EFLB bid corresponds to the advertiser's indifference between occupying positions $\ell$ and $\ell + 1$, the above inequality holds with equality. Thus, EFLB equilibrium bids satisfy the following recursive relationship:

$$\lambda_\ell q_{\ell+1} b_{\ell+1}^{EFLB} = (\lambda_\ell - \lambda_{\ell+1}) v_\ell q_\ell + \lambda_{\ell+1} q_{\ell+2} b_{\ell+2}^{EFLB}, \tag{9}$$

where $v_\ell$ denotes advertiser's effective value (click valuation net commission payments). The recursion starts at $\ell = L$, giving $q_{L+1} b_{L+1}^{EFLB} = v_L q_L$ (as $\lambda_{L+1} = 0$), where $v_L$ is the effective value for the advertiser at rank $L$. Re-writing equation (9) with focal advertiser at rank $\ell - 1$ (instead of $\ell$) and solving the recursion, EFLB equilibrium bids for advertiser $\ell$ can be written as the following:

$$b_\ell^{EFLB} q_\ell \lambda_{\ell-1} = \sum_{p=\ell}^{L} (\lambda_{p-1} - \lambda_p) v_{p-1} q_{p-1}, \ \textit{for all } 2 \leq \ell \leq L. \tag{10}$$

All envy-free equilibria are individually rational for each advertiser. To see this, set $p = L+1$ in equation (4), which gives us that $v_j^c - \gamma_j p_j f - cpc_{j,\ell} \geq 0$ as $\lambda_{L+1} = 0$. Because the derived EFUB and EFLB equilibria are a refinement of the envy-free criterion, they satisfy advertisers' participation constraints;

thus, advertisers' expected payments (from clicks and commissions) are lower than per-click valuations.

### 4.3.1 Identification and Estimation of Advertiser Valuations

**Approach.** In the previous section, advertisers' bids were derived as a function of their valuations, quality scores, and competition. However, note that bids and quality scores are observed, but valuations are not. Consequently, observed bids are inverted to uncover each advertiser's per-click valuation. The subsequent discussion outlines how these primitives are uncovered.

As the envy-free Nash equilibrium treats each auction as separate, one could consider identification of valuations for each position auction. Let $j_\ell$ be any advertiser $j$ that occupies position $\ell$, $v_{j_\ell}$ be the implied effective valuation (click value net commission payments) for the advertiser, and $q_{j_\ell}$ be the quality score. Then, the envy-free criterion of equation (4), for any positions $\ell$ and $p$, can be written as follows:

$$(\lambda_\ell - \lambda_p)v_{j_\ell}q_{j_\ell} \geq \lambda_\ell q_{j_{\ell+1}} b_{j_{\ell+1}} - \lambda_p q_{j_{p+1}} b_{j_{p+1}}. \tag{11}$$

For $p > \ell$, substituting $p$ by $\ell + 1$ gives the sharpest lower bound for advertisers' valuations at every position $\ell$. Conversely, for $p < \ell$, substituting $p$ by $\ell - 1$ and reversing the inequality (as $\lambda_{\ell-1} > \lambda_\ell$), gives the sharpest upper bound for advertisers' valuations at every position $\ell$. Then, the sequence of inequalities for all positions can be summarized using the following equation:

$$q_{j_1}v_{j_1} \geq \ldots \geq \underbrace{\frac{q_{j_\ell}b_{j_\ell}\lambda_{\ell-1} - q_{j_{\ell+1}}b_{j_{\ell+1}}\lambda_\ell}{\lambda_{\ell-1} - \lambda_\ell}}_{ICC_{\ell-1,\ell}} \geq q_{j_\ell}v_{j_\ell} \geq \underbrace{\frac{q_{j_{\ell+1}}b_{j_{\ell+1}}\lambda_\ell - q_{j_{\ell+2}}b_{j_{\ell+2}}\lambda_{\ell+1}}{\lambda_\ell - \lambda_{\ell+1}}}_{ICC_{\ell,\ell+1}} \geq \ldots \geq q_{j_L}v_{j_L}, \tag{12}$$

where $ICC_{\ell-1,\ell}$ (respectively, $ICC_{\ell,\ell+1}$) denote the incremental cost of click between positions $\ell - 1$ and $\ell$ (respectively, $\ell$ and $\ell + 1$).[18] Equation (12) is used for identification and estimation of advertisers' valuations, following the steps in Athey and Nekipelov (2011), henceforth AN.

**Identification.** Recall that given quality scores, a range of bids can achieve the same position. Thus, given observed positions and bids, valuations can only be identified over a range. This means that advertisers' per-click valuations are set-identified but not point-identified. The subsequent discussion seeks to infer valuations at the upper and lower bounds of this set, as they correspond (respectively) to the EFUB and EFLB equilibria derived in the last section.

As there are multiple envy-free equilibria, equation (12) provides bounds on each advertiser's valuation (at the impression level), for any position $\ell$, which can be written as below:

[18] Note that $ICC_{\ell-1,\ell}$ (and $ICC_{\ell,\ell+1}$) are defined at the auction-position level, and can be calculated from observed bids, quality scores, and rank effect estimates ($\lambda_\ell$).

$$q_{j_\ell} v_{j_\ell} \in [ICC_{\ell,\ell+1}, ICC_{\ell-1,\ell}] \text{ for all } \ell = 2, \ldots, L, \text{ and, } q_{j_1} v_{j_1} \in [ICC_{1,2}, \infty). \tag{13}$$

The lower bound of equation (13) corresponds to valuations implied by the EFLB equilibrium (hence, 'envy-free lower bound' nomenclature), while the upper bound corresponds to valuations implied by the EFUB equilibrium (hence, 'envy-free upper bound' nomenclature). Because the valuations implied by EFLB (respectively, EFUB) equilibrium are the lowest (respectively, highest) among all envy-free criteria, estimating valuations and conducting policy simulations at either extrema helps to quantify the lower bound (respectively, upper bound) of the retailer's revenue across all envy-free equilibria.

In summary, the implied effective valuation (click value net commission payment) for each impression corresponding to the respective equilibria is estimated as follows:

$$v_{j_\ell}^{EFLB} = \frac{ICC_{\ell,\ell+1}}{q_{j_\ell}}, \text{ for } \ell \geq 1, \text{ and, } v_{j_\ell}^{EFUB} = \frac{ICC_{\ell-1,\ell}}{q_{j_\ell}} \text{ for } \ell \geq 2. \tag{14}$$

The collection of all impression level values ($v_{j_\ell}$) for each advertiser $j$, across all impressions (and products) where the advertiser won an auction, provides the distribution of each advertiser's effective value.

**Estimation.** Note that advertisers' valuations are unknown to the retail platform (and the econometrician). Consequently, observed bids and quality scores are used to estimate each advertiser's effective valuation, that is, click valuation net commission payments, using equation (14). The estimation approach consists of three steps: (1) estimation of $ICC_{\ell,\ell+1}$ for each rank $\ell$ corresponding to every auction, (2) given ICC curves, estimation of implied effective valuation for the EFLB and EFUB equilibria, using equation (14), and (3) given estimated impression values, construction of advertiser-specific valuation distribution over all impressions and products.

The first step aims to estimate the ICC curves. Note that observed bids and quality scores are consistent with envy-free equilibria if and only if $ICC_{\ell,\ell+1}$ is non-increasing in $\ell$ for each auction. Otherwise, the lower bound of equation (13) could surpass its upper bound, which violates the modeling assumption. As this condition could be empirically violated in the data, approximate perturbations of the ICC curves are performed to correct any monotonicity violation, as suggested in AN. Approximate perturbations aim to find the minimal perturbation vector, $d$, which ensures that $ICC^d_{\ell-1,\ell} \geq ICC^d_{\ell,\ell+1}$ for each rank $\ell$, where $ICC^d_{\ell-1,\ell}$ (respectively, $ICC^d_{\ell,\ell+1}$) are the approximate perturbations of $ICC_{\ell-1,\ell}$ (respectively, $ICC_{\ell,\ell+1}$). This corresponds to solving the following quadratic program for each auction:

$$\min_d \sum_{\ell=1}^{L} (1 - d_\ell)^2, \text{ s.t. } ICC^d_{\ell-1,\ell} \geq ICC^d_{\ell,\ell+1}, \tag{15}$$

where $ICC^d_{\ell-1,\ell} = \frac{q_{j_\ell} b_{j_\ell} \lambda_{\ell-1} d_{\ell-1} - q_{j_{\ell+1}} b_{j_{\ell+1}} \lambda_\ell d_\ell}{\lambda_{\ell-1} - \lambda_\ell}$ and $ICC^d_{\ell,\ell+1} = \frac{q_{j_{\ell+1}} b_{j_{\ell+1}} \lambda_\ell d_\ell - q_{j_{\ell+2}} b_{j_{\ell+2}} \lambda_{\ell+1} d_{\ell+1}}{\lambda_\ell - \lambda_{\ell+1}}$. For ICC curves where violations do not occur, $d = 1$ by construction. When violations do occur, equation (15) tries to find the minimal perturbation vector, $d \neq 1$, that corrects for these violations. The structural interpretation of the vector of perturbation weights, $d$, is the uncertainty in competing advertisers' bids and quality scores (see Varian, 2007; Athey & Nekipelov, 2011).

Let $d^*$ be the $L$ dimensional perturbation vector, estimated at the auction level, using the equation (15). In the second step, impression-level valuations are recovered as per the following equation:

$$\hat{v}^{EFLB}_{j_\ell} = \frac{ICC^{d^*}_{\ell,\ell+1}}{q_{j_\ell}},\ for\ \ell \geq 1,\ and,\ \hat{v}^{EFUB}_{j_\ell} = \frac{ICC^{d^*}_{\ell-1,\ell}}{q_{j_\ell}}\ for\ \ell \geq 2. \tag{16}$$

Note from equation (16) that the effective valuation for the highest-ranked advertiser is not identified in the EFUB equilibrium. Following AN, we set $\hat{v}^{EFUB}_{j_1} = \bar{b}$ (highest bid in our data).

In the final step, all impression level values for a given campaign are collected to plot the distribution of effective valuations for each campaign.[19] As an institutional detail, the retail setting conducts eligibility auctions, where products that match the consumer's search filters (e.g., hats, men) are eligible to participate in the position auction. The estimation procedure assumes that these eligibility criteria are exogenous to advertisers' bidding decisions.[20] Consequently, estimated valuations should be interpreted as valuations conditioned on being eligible to win a given auction.

**Results.** Recall that the data contains three phases in which the quality score is varied (see Section 2.3). A subset of data before and during the quality score experiment (November 7 to December 8, 2022), consisting of rank by bid auctions ($q_{j_\ell} = 1$ for all advertisers), is used for the estimation of the advertiser model. This choice mirrors the assumption in equilibrium analysis, as observed bids in rank-by-bid position auctions reflect the rest point of bids in the data, hence are assumed to reflect the complete information equilibrium. As advertisers are not informed of the quality score experiment, it is found that position auctions where $q_{j_\ell} \neq 1$ do not see changes in bids (see Section 3.2). Consequently, these auctions are excluded from the estimation sample because the estimated valuations from these auctions do not reflect true valuations.

As all items in an ad campaign have the same bid, we presume that per-click valuations are the same for any product in a campaign. The collection of all estimated $\hat{v}^{EFUB}_{j_\ell}$ (respectively, $\hat{v}^{EFLB}_{j_\ell}$) for a given campaign $j$, across all impressions and products where the advertiser won a sponsored slot, gives

[19] Similar to AN, we assume that if an advertiser runs multiple campaigns, then they separately optimize bids for each of their campaigns. As the retail setting ensures that two campaigns from the same advertiser do not win the same auction, the assumption that advertisers' objective function is separable in bids across campaigns is a feasible one.

[20] This assumption is based on the findings in Section 3.2, where bidding is at the campaign level, and does not vary by consumer filters. In policy simulations, we simulate the respective eligibility criteria for each simulated auction (Section 5).

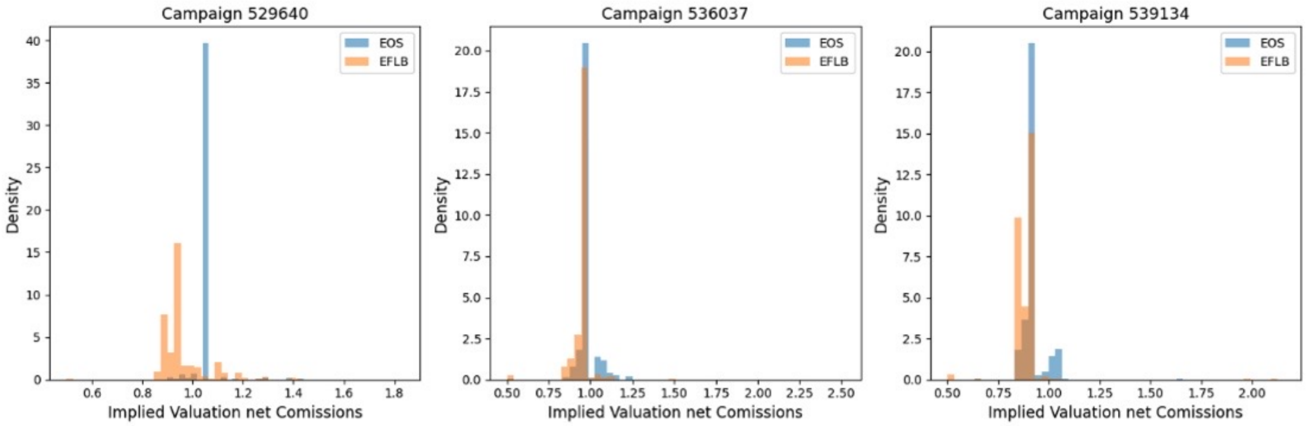


Figure 5: Distribution of implied effective valuations for the EFLB equilibrium ($\hat{v}^{EFLB}$) and the EFUB equilibrium ($\hat{v}^{EFUB}$) in the top 3 campaigns in data (in terms of volume of auctions won).

the distribution of each campaign's implied effective valuation. Figure 5 presents a histogram plot of estimated valuations for the 3 campaigns that were the highest winning in our data (highest in terms of the number of impressions won in our sample). Consistent with the monotonicity of the ICC curves (see equation (13)), the distribution of valuations implied by EFUB equilibrium first-order stochastically dominates those implied by the EFLB equilibrium. That is, the EFUB distribution places a greater mass on higher valuations compared to the EFLB distribution. Following AN, we select the median value across impressions as the campaign-specific implied effective value for policy simulations.[21]

# 5 Policy Simulations

This section details the objective function of the retailer and then performs policy simulations to investigate retailer and advertiser outcomes under counterfactual quality scoring mechanisms. The structural model developed in Section 4.3 is used to predict equilibrium bids in response to the quality score rule. Given quality scores and bids, we predict counterfactual ranks in the induced equilibrium, and use machine learning to predict consumer response to winning advertisers (Section 4.2). Based on these model predictions, we quantify retailer profits and advertiser surplus in alternative quality score rules.

The first counterfactual simulation illustrates that augmenting click-based quality scores with commission revenues can increase the retailer's profit and the advertisers' surplus. Guided by this insight, the second simulation exercise considers a family of quality scoring mechanisms intended to illustrate the trade-offs between click-based and commission-based quality scores. In this class of mechanisms, we

[21] As explained in AN, the median valuation is chosen rather than the mean to reduce sensitivity to the first-position advertiser in the EFUB model, whose valuation is not identified and set to be the upper bound of all bids.

quantify the profit-maximizing and welfare-maximizing quality score as a weighted function of clicks and commissions and show that quality scores that balance both can benefit the retailer and all advertisers. A final set of simulations investigates the revenue-efficiency trade-off in retail media position auctions and quantifies a retailer's incentive to improve its own profit at the cost of advertiser surplus.

### 5.1 Retailer's Objectives and Simulation Procedure

This section first describes the objective functions influenced by the quality scores that we consider in the simulation exercise: retail profit and social welfare. The section subsequently overviews how these objectives are simulated.

**Retailer and Social Welfare Objective Functions.** The retailer's profit from a position auction is the sum of their revenue from clicks and sales received on all winning advertisers (sponsored slots) in the auction. Upon receiving a click on an advertiser $j$ located at rank $\ell$, the retailer earns a click revenue of $cpc_{j,\ell} = \frac{q_{\ell+1}b_{\ell+1}}{q_j}$ per the rules of the generalized second price (GSP) auction. If the click leads to a sale, then the retailer also earns revenue from commissions, given by $\gamma_j p_j f$, where $\gamma_j$ is the purchase rate, $p_j$ is the price, and $f$ denotes the commission fee. Thus, the total expected profit earned by the retailer from a position auction is:

$$R(\mathbf{q}) = \sum_{j=1}^{J}\sum_{\ell=1}^{L} \alpha_j \lambda_\ell \Big(\frac{q_{\ell+1}b_{\ell+1}}{q_j} + \gamma_j p_j f\Big) \mathbf{1}(j \to \ell)(\mathbf{q}), \tag{17}$$

where $\mathbf{1}(j \to \ell)(\mathbf{q})$ is the indicator function that advertiser $j$ occupies rank $\ell$, $\alpha_j \lambda_\ell$ is the probability of receiving a click, and the summation is over all advertisers and auctioned ranks. The above function is non-concave in quality scores, and thus it cannot be optimized analytically. As a result, policy simulations are used to enhance quality scores.

The retail platform could also choose to improve social welfare in the market. Social welfare is the total surplus earned by the retailer and all advertisers in the market. Recall that equation (1) outlines the expected utility earned by the advertiser $j$ when occupying position $\ell$, which is written as $U_{j,\ell} = \alpha_j \lambda_\ell (v_j^c - cpc_{j,\ell} - \gamma_j p_j f)$, where $cpc_{j,\ell}$ and $\gamma_j p_j f$ are defined earlier. Summing this expression over all advertisers who win a slot and adding the retailer's profit function from equation (17), the social welfare in the market is the following:

$$SW(\mathbf{q}) = \sum_{j=1}^{J}\sum_{\ell=1}^{L} \alpha_j \lambda_\ell v_j^c \times \mathbf{1}(j \to \ell)(\mathbf{q}). \tag{18}$$

Click and commission payments do not contribute to social welfare because they are just payment (welfare) transfers from the advertiser to the retailer.

**Simulating Advertiser Position Allocations and Payments.** Quality scores affect advertisers' positions and their per-click payments. These outcomes are central to computing the retailer's profit and social welfare objective functions. Thus, to predict how advertiser and retailer outcomes change with quality scores, counterfactual simulations proceed in two steps: 1) determining the allocations of advertisers to positions, and 2) determining advertiser payments for clicks and commissions. Specifically, the counterfactual simulation proceeds as follows:

1. *Determining Allocations.* Allocations in the envy-free equilibria are monotonically decreasing in $v_j q_j$, where $v_j = v_j^c - \gamma_j p_j f$ is advertiser $j$'s click valuation net their commission payments. To see this, note that in an envy-free equilibrium, any advertiser at position $\ell$ does not prefer to occupy position $p \neq \ell$, and similarly any advertiser at position $p$ does not prefer to occupy position $\ell \neq p$. Consequently, equation (4) can be written from the perspective of advertisers at rank $\ell$ and $p$, respectively, as follows:

$$(\lambda_\ell - \lambda_p) v_\ell q_\ell \geq \lambda_\ell q_{\ell+1} b_{\ell+1} - \lambda_p q_{p+1} b_{p+1}, \textit{ and,} \tag{19}$$

$$(\lambda_p - \lambda_\ell) v_p q_p \geq \lambda_p q_{p+1} b_{p+1} - \lambda_\ell q_{\ell+1} b_{\ell+1}. \tag{20}$$

   Adding both equations gives that $(\lambda_\ell - \lambda_p)(v_\ell q_\ell - v_p q_p) \geq 0$, which means that both terms should have the same sign. Consequently, because $\lambda_\ell$'s are monotonically decreasing in $\ell$ (Assumption 1), it follows that $v_\ell q_\ell$'s are also monotonically decreasing in the equilibrium rank, $\ell$. Thus, equilibrium ranks in the EFUB and EFLB equilibria follow the descending order of $vq$. Given this property, the $\ell$th highest value of $v_j q_j$ is assigned rank $\ell$, for each simulated auction.[22]

2. *Determining Payments.* Recall from equation (17) that retailer's profits are accrued from advertisers' payments from clicks and commissions. The below discussion elaborates on how both revenue components are calculated, and combined, to calculate the outcomes of interest.

   (a) *Click Revenues.* Advertisers' per-impression payments are the product of the costs per click and likelihood of clicks (CTRs). We discuss the computation of each in turn.

   First, regarding costs per-click, for any rank $\ell < L$, where $L$ is the lowest rank, per-click payments are recursively calculated, using equation (6) for the EFUB equilibrium, and equation (9) for the EFLB equilibrium. Payments at rank $L$ form the base case of the recursion, and

[22] An institutional detail is that the retail setting runs eligibility auctions. That is, given consumers' choice of product filters (e.g., hats, men), sponsored products eligible for the filter are eligible to bid for the auction. Consequently, for each auction, we simulate the eligibility criterion used by the retail platform, and determine allocations conditional on eligibility.

are computed by (1) setting $\lambda_{L+1} = 0$ in equation (6) or (9), if at least $L + 1$ advertisers are eligible for the auction, or (2) setting payment equal to the reserve price, $r_p = \$0.5$, if only $L$ advertisers are eligible for the auction.

Second, regarding click-through rates (CTRs) by position, an advertiser $j's$ expected CTR at the first rank ($\ell = 1$) is $\alpha_j$, is estimated using machine learning as described in Section 4.2. For $\ell \geq 2$, the expected CTR is appropriately reduced by the causal rank effect estimate from Table 3 to predict advertiser's effective CTR (that is, $\alpha_j \lambda_\ell$) at position $\ell$.

Together, these quantities capture the first term of equation (17), that is $\alpha_j \lambda_\ell \times cpc_{j,\ell}$.

(b) *Commission Revenues.* Commission revenues at any rank $\ell$ are obtained by multiplying advertiser's effective CTR ($\alpha_j \lambda_\ell$) with the expected commission revenues given a click, that is $\gamma_j p_j f$. These quantities capture the second term of equation (17), that is $\alpha_j \lambda_\ell \times \gamma_j p_j f$.

(c) *Retailer profits, advertiser surplus, and social welfare.* Retailer's profit can be calculated by adding revenues from clicks ($\alpha_j \lambda_\ell \times cpc_{j,\ell}$) with those from commissions ($\alpha_j \lambda_\ell \times \gamma_j p_j f$), and summing them over all ranks $\ell$, per equation (17).

Next, given estimates of each advertiser's per-click valuation ($v_j^c$) and their effective CTR ($\alpha_j \lambda_\ell$) calculated above, social welfare is calculated per equation (18). The difference between social welfare and retailer profits quantifies total advertiser surplus in the market.

### 5.2 Simulation Results

Policy simulations are conducted in a subset of the validation data, that is, rank-by-bid auction data after the quality score experiment (December 9 to December 21, 2022), leading to $654k$ position auctions and $1.55M$ rows. The results are summarized below.

#### 5.2.1 Counterfactual 1: Simulating Simple Quality Score Rules

First, policy simulations are performed to investigate how simple quality score rules could affect retailer profits and advertiser surplus in equilibrium. The simulated quality score rules include: (1) $q_j^0 = 1.0$ or a rank by bid benchmark, (2) $q_j^1 = \alpha_j$ or quality scores based on advertisers' click-through rate (CTR) predictions, (3) $q_j^2 = \alpha_j \gamma_j \overline{p}_j f$ or quality scores based on advertisers' expected normalized commission payments to the retail website, where $\overline{p}_j$ is the price divided by a dollar to make it unitless, producing normalized commissions, and (4) $q_j^3 = \alpha_j(1+\gamma_j \overline{p}_j f)$ or quality scores that augment advertisers' CTR with normalized commission payments.[23] These rules are chosen because $q_j^0$ serves as a benchmark without

[23] Prices are normalized since quality scores are unitless weights that are applied to advertiser bids.

quality scores, $q_j^1$ reflects a purely click-based rule, $q_j^2$ represents a commission-based rule, and $q_j^3$ applies a combination of the latter two extremes. First, results in the EFUB equilibrium selection are discussed, followed by results in the EFLB equilibrium selection.

**EFUB Equilibrium Results.** The valuations from the EFUB equilibrium are the upper bound of valuations from all envy-free equilibria (see Section 4.3.1), consequently these predictions provide an upper bound on the retailer's profits from the simulated quality score rules.

Table 10 provides estimates of click revenues (and CTRs), commission revenues (and conversion rate), as well as total retailer profits from different simulated quality score mechanisms. The rank-by-bid benchmark ($q_j^0$) fares the worst in terms of retailer profits, as well as click and commission revenue. A CTR-based quality score ($q_j^1$) accrues the highest click revenues (and volume of clicks), but not the highest commissions. Conversely, a commission-based quality score ($q_j^2$) accrues the highest commission revenue, but not the highest click revenue. Further, total retailer profits from $q_j^1$ exceed those of $q_j^2$, indicating that clicks generate more profits than commissions in our empirical setting. A quality score rule that combines CTR with commissions ($q_j^3$), achieves the highest retailer profit among the simulated rules, even though click revenues (commission revenues) are dominated by $q_j^1$ (respectively, $q_j^2$).

| Quality Score | Click Revenue (CTR) | Commission Revenue (CVR) | Total Retailer Profits |
|---|---|---|---|
| $q_j^0 = 1$ | \$9,818 (0.61%) | \$2,954 (0.014%) | \$12,742 |
| $q_j^1 = \alpha_j$ | \$12,595 (0.92%) | \$3,562 (0.020%) | \$16,157 |
| $q_j^2 = \alpha_j \gamma_j \overline{p}_j$ | \$6,302 (0.60%) | \$7,444 (0.017%) | \$13,746 |
| $q_j^3 = \alpha_j(1 + \gamma_j \overline{p}_j f)$ | \$10,694 (0.85%) | \$6,468 (0.022%) | \$17,162 |

Table 10: Predictions of (1) click revenues (and click through rates), (2) commission revenue (and conversion rates), as well as (3) total retailer profits, from simulated quality score mechanisms in EFUB equilibrium selection.

| Quality score rule | Advertiser Surplus | Social Welfare |
|---|---|---|
| $q_j^0 = 1$ | \$510 | \$13,252 |
| $q_j^1 = \alpha_j$ | \$2,815 | \$18,972 |
| $q_j^2 = \alpha_j \gamma_j \overline{p}_j$ | \$3,628 | \$17,374 |
| $q_j^3 = \alpha_j(1 + \gamma_j \overline{p}_j f)$ | \$3,450 | \$20,612 |

Table 11: Predictions of (1) advertiser surplus and (2) social welfare (sum of retailer profit and advertiser surplus) from simulated quality score mechanisms in EFUB equilibrium selection.

Table 11 characterizes advertiser surplus and social welfare under counterfactual quality score mechanisms. Advertiser surplus is lowest when the retailer implements a rank-by-bid allocation ($q_j^0$). In this case, the match value between consumers and advertisers is poor, leading to a lower volume of clicks, and hurting advertiser surplus. Second, per-click prices are higher if advertisers with higher bids win auctions, further reducing advertiser surplus. Thus, quality scores can generate positive outcomes for both advertisers and retailers. As both retailer profits and advertiser surplus are lowest in rank-by-bid position auctions, the social welfare is also the lowest.

Furthermore, the advertiser surplus is low when using CTR-based scores ($q_j^1$), but improves when augmenting the CTR with commissions ($q_j^2$ or $q_j^3$). This suggests that incorporating commissions into click-based quality scores could improve advertiser surplus, in addition to improving retailer profits (see Table 10). Driven by the increase in advertiser surplus and retailer profit, social welfare is the highest when retailer uses a combination of CTR and commissions ($q_j^3$).

**EFLB Equilibrium Results.** Results from the EFLB equilibrium remain qualitatively similar to those obtained in the EFUB equilibrium. Tables for retailer profit, advertiser surplus, and social welfare are moved to the appendix for brevity (see Tables 15 and 16 in Appendix C).

Overall, this section establishes that click-based quality scores can be improved by augmenting them with commissions. However, it is possible that there could be better ways to combine clicks and commissions than what is explored in this section. To investigate this, we proceed to optimize quality scores over a parameterized class of mechanisms.

### 5.2.2 Counterfactual 2: Optimizing Over a Class of Quality Score Mechanisms

Section 5.2.1 finds that augmenting CTR-based quality scores with normalized commissions ($\gamma_j \overline{p}_j f$) improves both retailer profits and advertiser surplus. Guided by this insight, this section introduces a class of quality scoring mechanisms and estimates the profit-maximizing and welfare-maximizing rules in this family. Specifically, the following class of (exogenous) quality scoring mechanisms is considered:

$$q_j(\theta) = (1-\theta)\alpha_j + \theta \alpha_j \gamma_j \overline{p}_j f, \;\; \textit{where } \theta \in [0,1]. \tag{21}$$

Compared with Section 5.2.1, the approach nests the CTR based quality scores when $\theta = 0$ (as $q_j(0) = \alpha_j$), the commissions-based quality score when $\theta = 1$ (as $q_j(1) = \alpha_j \gamma_j \overline{p}_j f$), and the combined click and commissions rule when $\theta = 0.5$ (as $q_j(0.5) = 0.5\alpha_j(\gamma_j \overline{p}_j f + 1)$).[24] Optimizing over $\theta \in [0,1]$ captures the (empirically) optimal level of weighting of advertisers' CTRs with their expected commission payments, which could enhance outcomes in a retail position auction. The simulations vary $\theta$ on a grid between 0 and 1 at intervals of 0.1. The results of the EFUB equilibrium are discussed next, and the results of the

[24] As quality scores are scale invariant, multiplicative constants such as 0.5 do not change allocations or payments. In other words, outcomes in $q_j = 0.5\alpha_j(\gamma_j \overline{p}_j f + 1)$ are the same as those in $\tilde{q}_j = \alpha_j(\gamma_j \overline{p}_j f + 1)$.

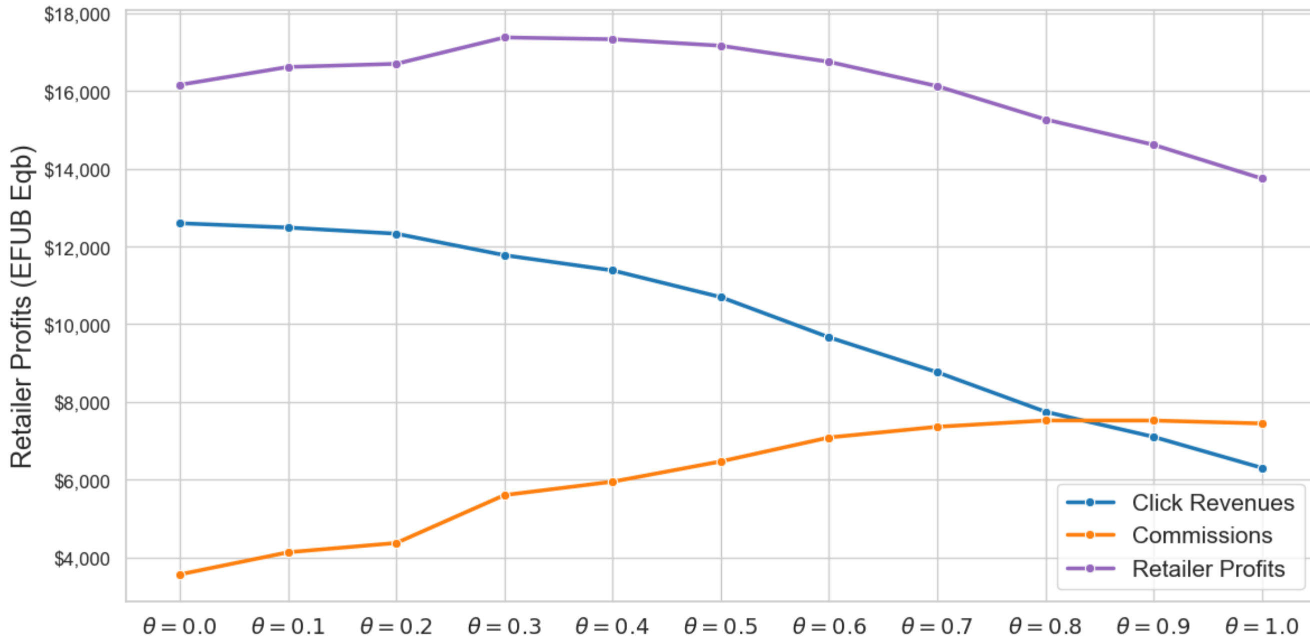


Figure 6: Predictions of (1) click revenues, (2) commission revenues, and (3) total retailer profits, in the EFUB equilibrium selection, for quality scoring rules of the form $q_j(\theta) = \alpha_j\big(\theta\gamma_j\overline{p}_j f + (1-\theta)\big)$, varying $\theta \in [0,1]$ over a grid at intervals of 0.1.

EFLB equilibrium are relegated to Appendix C.

**EFUB Equilibrium Results.** Figure 6 shows predictions of retailer profits and its components for the EFUB equilibrium at different values of $\theta$. Click revenues peak when quality scores are purely click-based ($\theta = 0$) and then decline monotonically as $\theta$ increases. Conversely, commission revenues rise with $\theta$ at first and plateau as $\theta$ approaches 1. Total retailer profits are concave in $\theta$, reaching their maximum at $\theta = 0.3$. Next, we characterize the rationale underpinning this result.

The retailer faces a trade-off between the quantity of clicks and the potential profits earned from a click as well as the commissions that are accrued from a sale. A click-based quality score maximizes the volume of clicks, hence click revenues earned by the retailer are the highest under this rule. Increasing the weight on commissions in the quality score rule affects retailer profits in two ways. First, as advertisers with higher commission payments are assigned higher ranks, commission revenues earned by the retailer start to increase, while the click revenues start to decline. Second, purchases are mediated by clicks in a position auction (see Section 3.1.3). Thus, a higher weight on commissions also decreases the volume of purchases, which, in turn, decreases commission revenues. Consequently, as the quality score rule places greater emphasis on commissions, (1) click revenues decline, as expected, while (2) commission revenues initially rise and then plateau (see $\theta = 0.8$ to $\theta = 1.0$ in figure 6). Driven by these trade-offs, total retailer profits follow a concave curve and are maximized when quality scores slightly favor clicks over

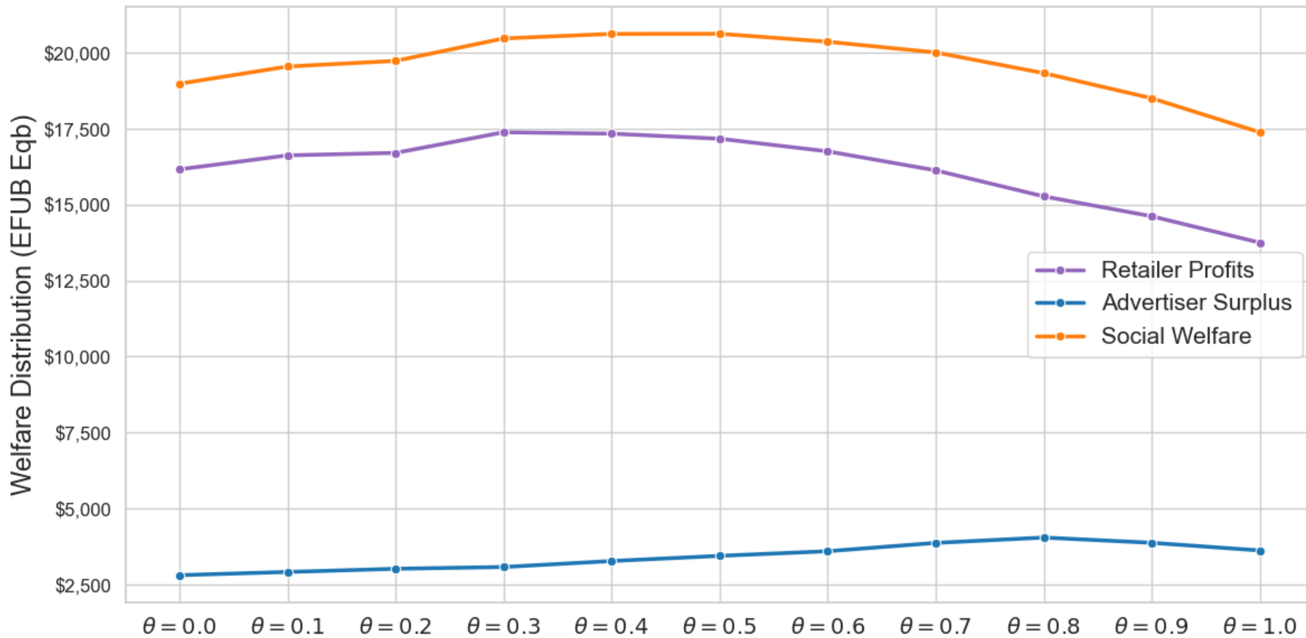


Figure 7: Predictions of (1) click revenues, (2) commission revenues, and (3) total retailer profits, in the EFUB equilibrium selection, for quality scoring rules of the form $q_j(\theta) = \alpha_j\big(\theta\gamma_j\overline{p}_j f + (1-\theta)\big)$, varying $\theta \in [0,1]$ over a grid at intervals of 0.1.

commissions, with the optimal around $\theta = 0.3$.

Next, figure 7 shows the predictions of advertiser surplus and social welfare in the market as a function of $\theta$ in the EFUB equilibrium. Total advertiser surplus is the lowest at $\theta = 0$, is concave in $\theta$, and is maximized at $\theta = 0.8$. The pattern can be explained as follows. Advertisers' per-click valuations are constant, so advertisers trade off the value of clicks with the payments for those clicks. When quality scores are increasingly click-based, the advertisers' per-click payments increase owing to greater competition for slots; as a result, their welfare declines. In contrast, commission payments (unlike clicks) are invariant to the advertiser's rank given that a purchase occurred. Hence, the competition for positions induced by commissions is lower. Thus, an increasing weight on advertisers' commission payments softens competition. Softer competition leads to a decreased per-click payment and, therefore, to an increased advertiser surplus from a click. As a result, advertiser surplus increases from $\theta = 0$ to $\theta = 0.8$. However, advertiser surplus suffers if quality scores become fully commission-based, as advertisers realize no surplus if there is no click. A quality score rule that promotes advertisers based only on commissions only creates a situation where high-commission goods with no clicks (and thus no sales) are favored over low-commission goods with many clicks. As a result, total advertiser surplus decreases for high $\theta$, leading to a concave function for advertiser surplus which is maximized at $\theta = 0.8$.

Finally, as expected, social welfare is maximized at $\theta = 0.5$, which lies between the retailer's profit maximizing rule, $\theta = 0.3$, and the advertiser surplus maximizing rule, $\theta = 0.8$. This suggests that while an augmented quality score approach improves both retailer profits and advertiser surplus, the retailer still faces an efficiency revenue tradeoff in retail media position auctions. In the next section, we discuss this tradeoff.[25]

#### 5.2.3 Counterfactual 3: Quantifying the Efficiency Revenue Tradeoff

A position auction is said to be efficient if the social welfare of the auction is maximized, and a welfare-maximizing quality score typically differs from a profit-maximizing one (see Section 5.2.2). This section quantifies the efficiency-revenue tradeoff in retail media position auctions. Specifically, this section investigates score squashing, which transforms raw quality scores to increase revenue at the cost of efficiency (e.g., Lahaie & Pennock, 2007; Athey & Nekipelov, 2011; Kim & Pal, 2024). Squashed scores are generated as $\tilde{q}_j(s) = q_j^s$, where $q_j$ is the raw score and $s \in [0, 1]$ is the squashing factor. Score squashing can be understood as per the insights of Myerson (1981), who shows that biasing auctions in favor of 'weaker' bidders can potentially increase revenue at the expense of welfare. Regarding quality score squashing, the chief economist at Yahoo! remarks: '*When someone has a really high ad click probability, they're very hard to beat, so it's not a really competitive auction. So that they don't just win every auction, we do squashing. This makes the auction more competitive.*'[26]

We provide intuition regarding quality score squashing with an example. Consider 2 advertisers A and B with the same effective value, say $v_a = v_b = 1$, but different quality scores. Suppose that advertiser A has a raw score of 0.7 while advertiser B has a raw score of 0.3, and the retailer applies a squashing factor of 0.5. Then, the squashed quality scores are $q_a = 0.84$ for advertiser A and $q_b = 0.55$ for advertiser B. Note that the squashed value weights, that is $v_j q_j$, for both advertisers are closer to each other now (difference of 0.29 versus 0.4). This increases competition in the market, increasing equilibrium bids and payments. Furthermore, note that the per-click price for advertiser A also increases, as GSP payments depend on the ratio of scores. Specifically, the 'better' advertiser pays more per click under squashed scores, as $0.55/0.84 = 0.65 > 0.3/0.7 = 0.43$. In general, the revenue impact of squashing depends on the rank order of competing advertisers based on quality scores, weighted bids, and squashed weighted bids (see Kim & Pal, 2024; Lahaie & Pennock, 2007 for more details), and it is a practical tool employed by search engines to improve revenues at the cost of welfare.

Squashing is considered for the rule $q_j = \alpha_j(1 + \gamma_j \overline{p}_j f)$, as it is the welfare-maximizing rule from

[25] The results in the EFLB equilibrium are qualitatively similar to the EFUB results. Figures for retailer profits, advertiser surplus, and social welfare are moved to the Appendix for brevity (see figures 12 and 13 in Appendix C).

[26] See *www.theregister.com*/2010/09/16/*yahoo_does_squashing*/ for details.

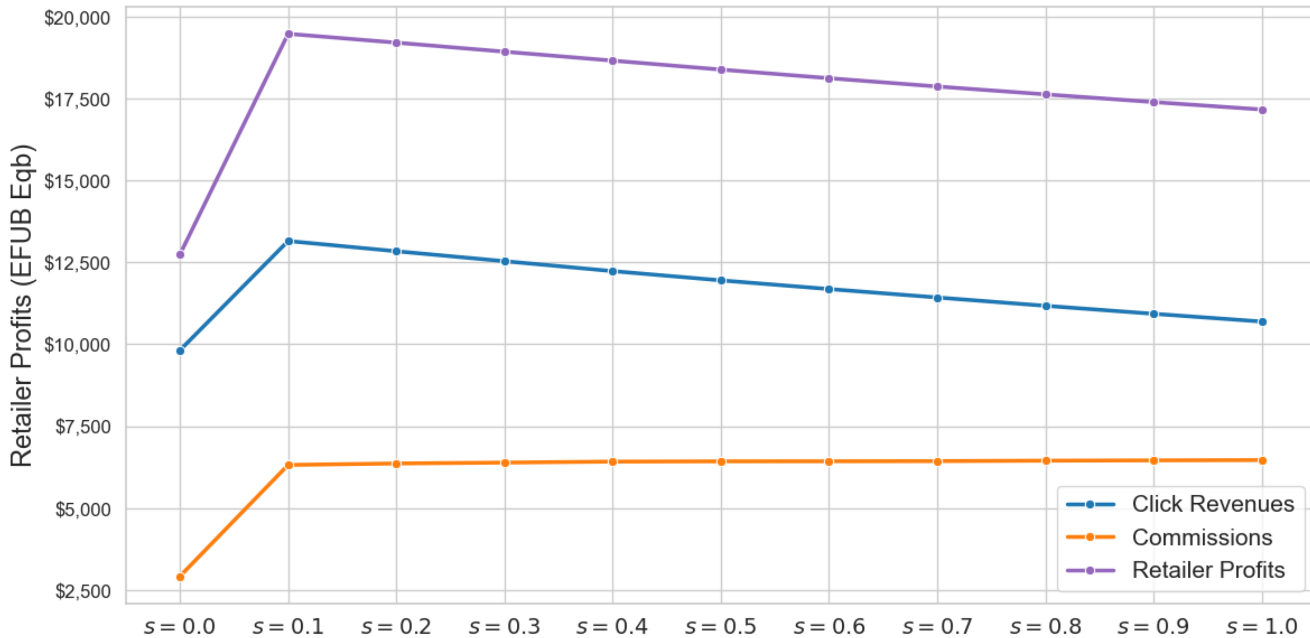


Figure 8: Predictions of (1) click revenues, (2) commission revenues, and (3) total retailer profits, in the EFUB equilibrium selection, for quality scoring rules of the form $q_j(s) = \left(\alpha_j(\theta\gamma_j\overline{p}_j f + (1-\theta)\right)^s$, varying $s \in [0,1]$ at intervals of 0.1.

Section 5.2.2. Thus, the squashed family of considered rules is $\tilde{q}_j(s) = \left(\alpha_j(1+\gamma_j\overline{p}_j f)\right)^s$, where the squashing factor, $s$, is varied between 0 and 1 in intervals of 0.1. The results of the EFUB equilibrium are discussed next, while the EFLB equilibrium results are moved to Appendix C.

**Results.** Figure 8 outlines the retailer's profit and its components under different squashing factors, where $s = 1$ denotes no squashing and $s = 0$ denotes a rank by bid mechanism. Click revenues increase with more squashing (that is, as $s$ decreases), while commission revenues remain flat. Driven by an increase in click revenues, retailer profits increase with squashing. However, in the rank by bid extreme ($s = 0$), allocations are so inefficient that the click and commission revenue components are hurt, reducing total profits.

Figure 9 shows advertiser surplus and social welfare in the market. Advertiser surplus monotonically decreases with squashing, while the social welfare remains flat. Thus, retailer profits increase as advertiser surplus decreases, meaning that a retailer can use squashing to increase its own profit at the expense of advertisers.

Hence, quality score squashing is a surplus extraction tool for the retailer. Squashing increases click revenues by intensifying the competition for clicks, making advertisers pay more per click, and lowering their surplus. In contrast to the enhanced quality scores described in Section 5.2.2, that lead to mutually beneficial outcomes, squashing works by extracting revenue from advertisers at the cost of their welfare.

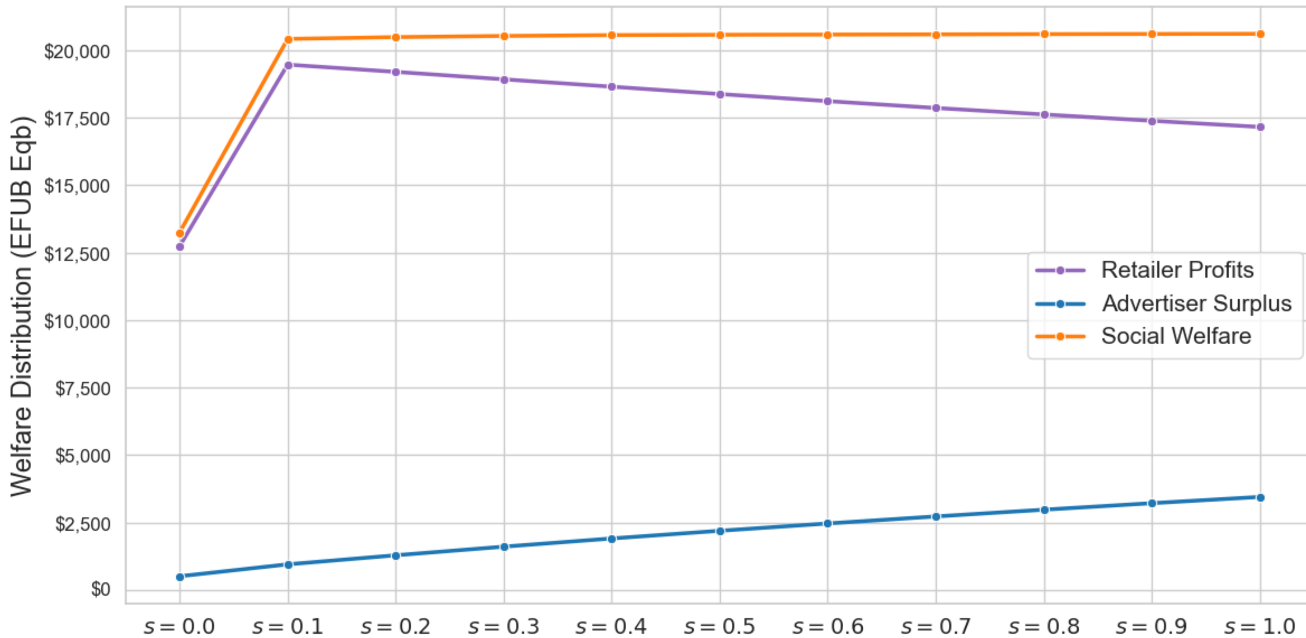


Figure 9: Predictions of (1) click revenues, (2) commission revenues, and (3) total retailer profits, in the EFUB equilibrium selection, for quality scoring rules of the form $q_j(s) = \left(\alpha_j(\theta\gamma_j\overline{p}_j f + (1-\theta)\right)^s$, varying $s \in [0, 1]$ at intervals of 0.1.

This could have the adverse effect of driving advertising dollars to other retail platforms.

## 6 Conclusion

Retail media is one of the fastest growing channels of digital advertising spend, projected to reach $130 billion by 2028, and account for nearly a quarter of all media spend.[27] Retail media use position auctions to monetize advertising. In these auctions, advertisers bid for better positions on the retail website (e.g., Walmart, Kroger). The goal of this research is to suggest improvements to the rank-by-bid and CTR based quality score mechanisms commonly used to allocate and price retail position auctions.

Most prior literature considers the role of quality scores in search engine advertising, but retail media differ because commissions also play a role in retailer outcomes. A key substantive insight of this research is that retailers indeed benefit from appropriately trading off click and commission revenue in order to improve their total profits. From the advertisers' perspective, augmenting commission revenues into quality scores can soften the competition for ranks, reducing per-click prices, and increasing advertiser surplus. At the same time, monetization in position auctions is contingent on a click, hence carefully augmenting click-through rates with commissions can improve retailer and advertiser outcomes.

[27]https://www.emarketer.com/content/whats-shaping-retail-media-now-through-2028-search-endemic-advertisers-financial-media-competition

Our empirical application uses granular auction-advertiser data from an online retailer that conducted a quality score experiment to ensure exogenous within-product variation in ranks. This variation enables causal estimation of position effects. Using a structural model of advertiser bidding and a machine learning model for clicks and purchases, quality score enhancements are shown to improve retailer profits by 7% and advertiser surplus by 42%, creating a win-win scenario. Furthermore, it is shown that the retailer can further improve its profit at the cost of advertiser surplus by using quality score squashing.

**Limitations and directions for future research.** Our research can be extended along a number of dimensions. Building on prior literature on position auctions and to ensure the existence of an equilibrium, we assume a full-information equilibrium for deriving advertisers' equilibrium bids. It might be possible to soften these assumptions by using a mean-field equilibrium approach to relax the strong informational requirements on advertisers in our setting. In addition, to highlight the tradeoffs inherent in retail media position auctions, this paper considers a specific class of quality score mechanisms and a specific tool for surplus extraction (e.g., squashing). Future research could generalize these insights to a wider range of quality scores. Next, following the literature on position auctions (e.g., Athey & Nekipelov, 2011; Lahaie & Pennock, 2007), we take reserve prices as given. This assumption is reasonable inasmuch as quality scores are scale-invariant; thus, they can be appropriately scaled such that reserve prices do not bind. Yet, the potential exists to integrate quality score approaches such as we devise with flexible reserve prices to add more flexibility in pricing ads, suggesting additional means to further enhance retail profits and/or social welfare. As a final extension, we seek to field test the mechanisms developed in this paper.

Given the rapid growth of retail media advertising markets in general, and position auctions in particular, we hope this research is a meaningful step towards enhancing these markets and spurs future research on mechanism design in retail media advertising.

## References


Abhishek, V., Jerath, K., & Zhang, Z. J. (2016, 8). Agency Selling or Reselling? Channel Structures in Electronic Retailing. *Management Science*, *62*(8), 2259–2280. doi: 10.1287/mnsc.2015.2230

Amaldoss, W., Desai, P. S., & Shin, W. (2015, 3). Keyword Search Advertising and First-Page Bid Estimates: A Strategic Analysis. *Management Science*, *61*(3), 507–519. doi: 10.1287/mnsc.2014.2033

Amaldoss, W., Jerath, K., & Sayedi, A. (2016, 3). Keyword Management Costs and "Broad Match" in Sponsored Search Advertising. *Marketing Science*, *35*(2), 259–274. doi: 10.1287/mksc.2015.0919

Athey, S., & Nekipelov, D. (2011). *A Structural Model of Sponsored Search Advertising Auctions.*

Ba, L. J., & Caruana, R. (2013, 12). Do Deep Nets Really Need to be Deep?

Berman, R., & Katona, Z. (2013, 7). The Role of Search Engine Optimization in Search Marketing. *Marketing Science*, *32*(4), 644–651. doi: 10.1287/mksc.2013.0783

Chen, T., & Guestrin, C. (2016, 8). XGBoost. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). New York, NY, USA: ACM. doi: 10.1145/2939672.2939785

Choi, H., & Mela, C. F. (2019, 11). Monetizing Online Marketplaces. *Marketing Science*, mksc.2019.1197. doi: 10.1287/mksc.2019.1197

Choi, W. J., Jerath, K., & Sayedi, A. (2024). *Endemic and Non-Endemic Retail Media: The Role of Brand Response* (Tech. Rep.).

Edelman, B., Ostrovsky, M., & Schwarz, M. (2007, 2). Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords. *American Economic Review*, *97*(1), 242–259. doi: 10.1257/aer.97.1.242

Farronato, C., Fong, J., & Fradkin, A. (2024, 1). Dog Eat Dog: Balancing Network Effects and Differentiation in a Digital Platform Merger. *Management Science*, *70*(1), 464–483. doi: 10.1287/mnsc.2023.4675

Ghili, S., Rafieian, O., & Rashid, M. (2025). *Auctions Meet Bandits: An Empirical Analysis* (Tech. Rep.).

Ghose, A., & Yang, S. (2009, 10). An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *Management Science*, *55*(10), 1605–1622. doi: 10.1287/mnsc.1090.1054

Haibo He, & Garcia, E. (2009, 9). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. doi: 10.1109/TKDE.2008.239

He, K., Zhang, X., Ren, S., & Sun, J. (2015, 12). Deep Residual Learning for Image Recognition.

Joo, M., Shi, J., & Abhishek, V. (2024, 7). Do Sellers Benefit from Sponsored Product Listings? Evidence from an Online Marketplace. *Marketing Science*, *43*(4), 817–839. doi: 10.1287/mksc.2021.0293

Katona, Z., & Zhu, Y. (2017). Quality Score that Makes You Invest. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2954707

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon et al. (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc.

Kim, D., & Pal, P. (2024, 8). *Nonparametric estimation of sponsored search auctions and impact of Ad quality on search revenue.* doi: 10.47004/wp.cem.2024.1624

Lahaie, S. (2006, 6). An analysis of alternative slot auction designs for sponsored search. In *Proceedings of the 7th acm conference on electronic commerce* (pp. 218–227). New York, NY, USA: ACM. doi: 10.1145/1134707.1134731

Lahaie, S., & Pennock, D. M. (2007, 6). Revenue analysis of a family of ranking rules for keyword auctions. In *Proceedings of the 8th acm conference on electronic commerce* (pp. 50–56). New York, NY, USA: ACM. doi: 10.1145/1250910.1250918

Long, F. (2023). Retail Media at Online Platforms as Resellers or Marketplaces. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4507593

Long, F., & Amaldoss, W. (2024, 9). Self-Preferencing in E-Commerce Marketplaces: The Role of Sponsored Advertising and Private Labels. *Marketing Science*, *43*(5), 925–952. doi: 10.1287/mksc.2022.0262

Long, F., Jerath, K., & Sarvary, M. (2022, 1). Designing an Online Retail Marketplace: Leveraging Information from Sponsored Advertising. *Marketing Science*, *41*(1), 115–138. doi: 10.1287/mksc.2021.1307

Lu, S., Zhu, Y., & Dukes, A. (2015, 11). Position Auctions with Budget Constraints: Implications for Advertisers and Publishers. *Marketing Science*, *34*(6), 897–905. doi: 10.1287/mksc.2015.0931

Moshary, S. (2025, 3). Does Sponsored Search Advertising Augment Organic Search? Evidence from an E-Commerce Platform. *Management Science*. doi: 10.1287/mnsc.2024.04792

Myerson, R. B. (1981, 2). Optimal Auction Design. *Mathematics of Operations Research*, *6*(1), 58–73. doi: 10.1287/moor.6.1.58

Narayanan, S., & Kalyanam, K. (2015, 5). Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach. *Marketing Science*, *34*(3), 388–407. doi: 10.1287/mksc.2014.0893

Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on machine learning - icml '05* (pp. 625–632). New York, New York, USA: ACM Press. doi: 10.1145/1102351.1102430

Platt, J. (2000, 6). Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Adv. Large Margin Classif.*, *10*.

Sahni, N. S., & Zhang, C. (2023, 11). Are consumers averse to sponsored messages? The role of search advertising in information discovery. *Quantitative Marketing and Economics*. doi: 10.1007/s11129-023-09270-z

Ursu, R. M. (2018, 8). The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions. *Marketing Science*, *37*(4), 530–552. doi: 10.1287/mksc.2017.1072

Varian, H. R. (2007, 12). Position auctions. *International Journal of Industrial Organization*, *25*(6), 1163–1178. doi: 10.1016/j.ijindorg.2006.10.002

Wang, Y., Tao, L., & Zhang, X. X. (2024, 8). Recommending for a Multi-Sided Marketplace: A Multi-Objective Hierarchical Approach. *Marketing Science*. doi: 10.1287/mksc.2022.0238

Yang, J., Sahni, N. S., Nair, H. S., & Xiong, X. (2024, 1). Advertising as Information for Ranking E-Commerce Search Listings. *Marketing Science*, *43*(2), 360–377. doi: 10.1287/mksc.2021.0292

Yang, S., & Ghose, A. (2010, 7). Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science*, *29*(4), 602–623. doi: 10.1287/mksc.1090.0552

Yao, S., & Mela, C. F. (2011, 5). A Dynamic Model of Sponsored Search Advertising. *Marketing Science*, *30*(3), 447–468. doi: 10.1287/mksc.1100.0626

Yoon, K. (2010, 7). Optimal Quality Scores in Sponsored Search Auctions: Full Extraction of Advertisers' Surplus. *The B.E. Journal of Theoretical Economics*, *10*(1). doi: 10.2202/1935-1704.1649

# A Appendix A: Data Analysis Details

## A.1 Additional Data Collection and Analysis

We collect additional data and conduct analyses to establish certain features of our setting. First, we seek to assess what the counterfactual sales might be for sponsored products were they to not appear in a sponsored rank. Based on discussions with the retailer, we learned that these advertisers rarely appear in the organic search results, so our working assumption is that counterfactual revenues from not advertising are zero. To further verify this, we scrape data from the website.

Collecting historical webpage data using the Wayback Machine, we investigate the overlap between sponsored and organic products in our marketplace. We find that sponsored products never appear in organic ranks on the first product listing page (top 78 ranks) for almost all the web pages we collect. This finding supports the assumption that the counterfactual sales for a sponsored product, if it does not win a sponsored auction slot, are zero.

Second, we seek to assess whether prices are invariant to changes in quality scores, as assumed in the counterfactual analyses. Using web page data collected at the product-date level, we observed no variation in the listed price for almost all products in our setting, regardless of quality score. Based on this finding, we assume price invariance. As an additional note, the retailer does not change its commissions in response to changes in quality scores, so these are treated as fixed as well.

Details of data collection and details of these analyses are reported below.

**Data collection effort.** Using the Wayback Machine, we retrospectively gathered historical web page data from the retailer that span dates from October 1, 2022, to December 31, 2022. These dates are chosen to match the period for the data we obtain from the marketplace (see Section 2). The Wayback Machine stores snapshots of data from webpages, which allows us to scrape these snapshots for the period of interest. The data is collected at the product-date level and contains details such as product identifiers, product names, price, webpage rank, and a tag to determine whether the product appeared in a sponsored or organic rank. In total, we acquire 6,971 product records (representing 4,222 unique products) from over 90 different web pages from the marketplace's website.

The Wayback Machine allows us to capture the price for all products on the first page of items. On average, this amounts to 78 product ranks (26 rows of 3 products each) for every webpage we collect. For each product (rank), we also collect tags to decipher whether the product appears in a sponsored rank or organic rank. Of the 78 product ranks on the first page, a page could constitute up to 6 sponsored products, and there are no sponsored listings beyond the first page.

### A.1.1 Findings

**Advertisers Rarely Appear in the Organic Listings.** First, only 0.19% of the 4,222 products appear in both sponsored and organic ranks. In particular, of the 4,222 unique products we collect, 3,874 products appeared only in organic slots, 340 products appeared only in sponsored slots, and 8 products appear in both sponsored and organic. Thus, the proportion of products that appear in both sponsored and organic ranks is negligible in our setting.

**Advertised Products are Weak Substitutes for Organic Products.** Second, products appearing in sponsored positions are systematically different from products appearing in organic positions, so they are not close substitutes. Of the 73 unique brands for which we collect data, only 36% of the brands have at least one sponsored product. Thus, many brands in the marketplace do not run ad campaigns to promote their products.

**Item Prices do Not Vary.** Third, we find that among the 4,222 products collected, only 21 products exhibit any fluctuation in prices across different dates (standard deviation $> 0$). Among products that do experience a price change, the magnitude of price fluctuation is minimal, with only 3 products (of 21) experiencing a standard deviation of $> \$10$ in their price across all dates.

## A.2 Price Dataset Construction

The data provided by the marketplace contains missing prices for many products. In order to partially address the missingness in prices, we integrate the data from the Wayback Machine with data collected from the marketplace.[28] After this step, the price of the product is still missing for 44% of the products.

Table 12 provides the descriptive statistics for the 181,690 known prices. We address missing data in prices using machine learning methods as discussed below.

| | Mean | Median | Std.Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Price | 78.06 | 44.99 | 211.84 | 0.99 | 39999.99 |

Table 12: Summary of observed prices.

### A.2.1 Gradient Boosting to impute missing prices

In this section, we outline the machine learning (ML) regression model we use to impute missing prices. Our procedure consists of two steps. First, data with known prices are used to train the regression model,

[28] The price data collected from the Wayback Machine is at the product-date level. As there is no temporal variation in prices, the average price of the collected data is merged with the data from the marketplace.

based on product characteristics. Subsequently, the ML model is used to predict missing prices. We opted for LightGBM (Light Gradient Boosting Machine) for our prediction task as it had good predictive outcomes. Gradient boosting methods rely on ensemble learning using weak learners (trees), thus are useful to prevent overfitting and improve out-of-sample performance. Furthermore, boosting techniques are capable of effectively handling categorical variables and solving for missing values in explanatory variables.

*Data pruning.* Note from table 12 that the minimum observed price is close to zero ($p = \$0.99$) and the maximum observed price is very high ($p = \$39,999$). Thus, we first prune our training data to avoid fitting our regression model on outlier prices. The pruned training data correspond to the middle 95% interval of prices, totaling 173,980 observations used for training. The $2.5th$ percentile of price corresponds to $9.99 and the $97.5th$ percentile of price corresponds to $321.99.

*Hyperparameter selection using cross-validation.* Using the pruned data, we train a five-fold cross-validation model to improve out-of-sample performance. Thus, we control overfitting by selecting the hyper parameters that achieve the best performance in the cross-validation procedure. Hyperparameters are selected using grid-search over pre-specified values for different attributes. The attributes considered are: (1) tree depth, (2) number of trees in the ensemble, (3) maximum number of leaves, (4) minimum number of children in each split, and (5) a regularization term. We observed that the ML method underpredicts high price in-sample, thus we minimize the mean absolute error in the cross-validation objective. The best model achieved a root mean square error of 24.39, a mean absolute error of 13.18, and an R squared value of 72.1% for observations in the sample, indicating a good model fit and effective price imputation. Figure 10 shows a scatter plot of actual versus predicted prices in training data ($174k$ samples). Note that the model underpredicts price for higher values, which means that the commission revenue lift that we obtain in our counterfactual simulations are lower bounds of the actual revenue lift from our proposed simulation procedure.

*Final price imputation.* For those products with missing prices, we use the prediction of the ML model as the final price of the product. For products with available prices (including prices lower than $2.5th$ percentile and higher than $97.5th$ percentile), we use the observed price as the final price. In this way, we create a comprehensive price dataset for all products available to us. Table 13 provides the descriptive statistics of the final price for the 326,432 products in our data.

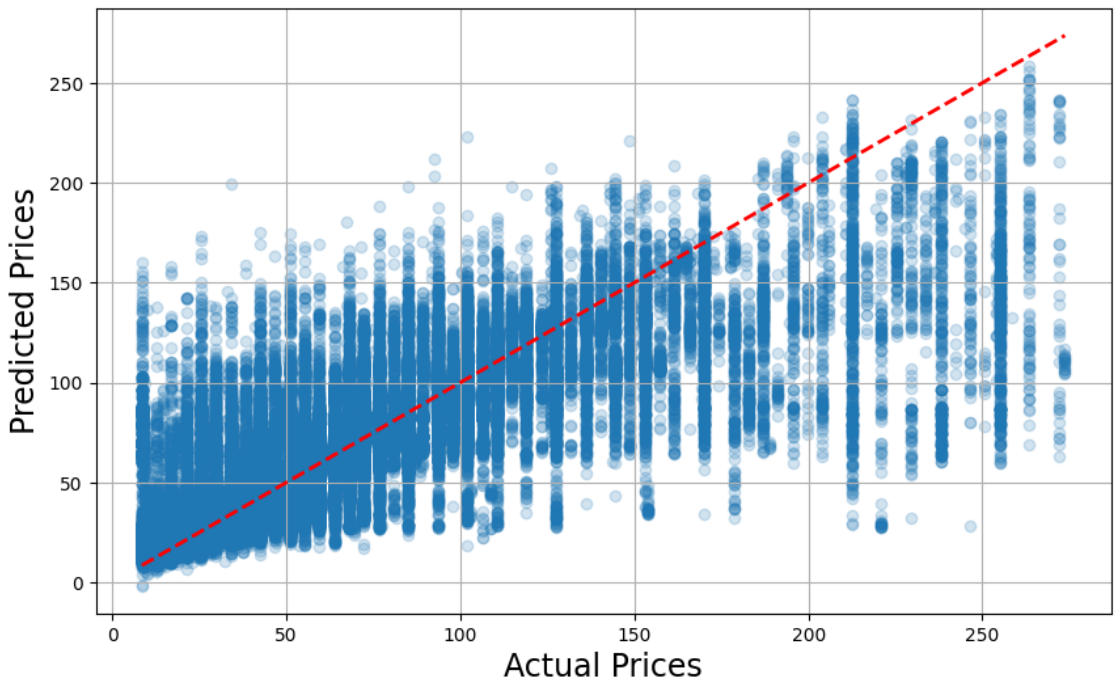


Figure 10: Actual Prices vs Predicted Prices in the training data ($N = 173,980$). All prices are scaled by a constant positive factor.

| | Mean | Median | Std.Dev. | Min. | Max. |
|---|---|---|---|---|---|
| Price | 71.20 | 55.24 | 136.92 | 0.124 | 33999.99 |

Table 13: Summary of prices (observed and imputed) for all sponsored products in the data. Prices are scaled by a constant positive factor.

# B Appendix B: Consumer Model Estimation Details

In this section, we provide further details regarding the consumer model. We first describe feature selection and preprocessing and then discuss details of model performance.

**Feature selection.** We consider all product attributes in our data as potential explanatory variables to predict clicks and conversions. A series of logistic regression models (predicting clicks) incrementally includes more features in the model, and assesses whether the fit of the model improves upon inclusion of the new features.

1. A logistic regression model was fit with all continuous variables: price, annual sales, product score, and rank. All coefficients were significant and the pseudo-$R^2$ of the regression was 0.019.

2. The categorical variable, Brand, was included in the logistic regression, and the pseudo-$R^2$ of the regression increased to 0.0354, indicating that Brand is highly predictive of clicks.

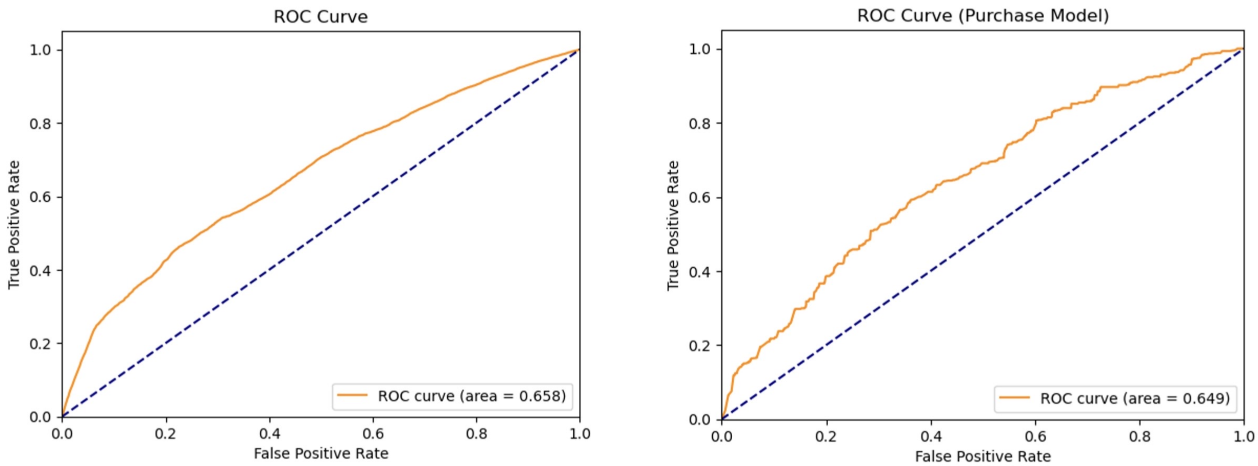


Figure 11: Receiver operating curve in the validation set for $N = 3.1M$ position auctions from Dec 9 to Dec 21, 2022. The left pane contains ROC curve for the deep learning based click forecast model, while the right pane contains ROC curve for the gradient boosting decision tree based purchase forecast model.

3. After Brand, the other potential categorical features in the data are considered, such as gender or age-group, or discrete product attributes to arrive at the final set of features. We retained those attributes that are most predictive of clicks.

**Feature pre-processing.** Both continuous and categorical features were pre-processed before being input into the Machine Learning model. The details are specified below:

- *Handling NaNs.* Missing data for continuous variables (e.g., productscore, annual sales) was handled by imputing the average variable value for the brand. For instance, if annual sales is missing for a given product, the average sales of the Brand is the imputed value. A separate label called "missing" was imputed for categorical variables' missing values.

- *Normalizing Continuous variables.* All continuous variables were scaled and normalized to ensure that their feature importance is not inflated due to high absolute values.

**Model Performance.** Figure 11 portrays the receiver operating curve (ROC) over the validation set for both the click (deep leanrning) and purchases (gradient boosted decision tree) models. The validation data comprises of $N = 3.1M$ position auctions after the quality score experiment, from Dec 9 to Dec 21, 2022. The area under the receiver operating curve (AUC-ROC) is 0.658 for the deep learning based click model, while it is 0.649 for the GBDT based purchase model.

**Model Fit in Policy Simulation Sample.** Policy simulations are conducted on a subset of the validation data, comprising rank-by-bid auctions for Dec 9 to 21, 2022, containing $654k$ position auctions ($1.55M$ rows). A close alignment of observed and predicted clicks and purchases provides evidence that the prediction models capture consumer behavior. We assess this alignment in Table 14, which compares realized and predicted outcomes. The table suggests that aggregate statistics are relatively well aligned.

| | Clicks (Realized/Predicted) | Conversions (Realized/Predicted) |
|---|---|---|
| Event Rate | 0.36%/0.37% | 0.007%/0.008% |
| Occurrences | $5,539/5,653$ | 102/127 |
| Revenues | \$3,417/\$3,605 | \$1,183/\$1,947 |

Table 14: Realized versus predicted outcomes in rank-by-bid auctions used for policy simulations ($N = 654k$ position auctions). The first row compares event rates, that is CTR or CVR, the second row compares occurrences, that is number of clicks or purchases, and the third row compares revenues accrued from clicks or purchases.

# C Appendix C: Policy Simulations Results For EFLB Equilibrium

The advertiser valuations from the EFLB equilibrium are the lower bound of all envy-free equilibria; consequently, this equilibrium provides a lower bound on the retailer's profits from the simulated quality score rules (see Section 4.3.1).

## C.1 Counterfactual 1: Simulating Simple Quality Score Rules

Table 15 provides predictions of retailer profits and its components, while table 16 quantifies advertiser surplus and social welfare in different quality score mechanisms in the EFLB equilibrium. Trends in EFLB equilibrium remain the same as those in the EFUB equilibrium.

| Quality score rule | Ad Revenue (CTR) | Commission Revenue (CVR) | Total Retailer Profits |
|---|---|---|---|
| $q_j^0 = 1$ | \$5,562 (0.42%) | \$4,528 (0.012%) | \$10,091 |
| $q_j^1 = \alpha_j$ | \$8,577 (0.77%) | \$4,592 (0.019%) | \$13,170 |
| $q_j^2 = \alpha_j \gamma_j \bar{p}_j$ | \$4,787 (0.48%) | \$6,973 (0.014%) | \$11,760 |
| $q_j^3 = \alpha_j(1 + \gamma_j \bar{p}_j f)$ | \$7,376 (0.70%) | \$6,687 (0.018%) | \$14,063 |

Table 15: Predictions of (1) ad revenues (click through rates), (2) commission revenue (conversion rates), and (3) total retailer profits, from simulated quality score rules in EFLB equilibrium selection.

## C.2 Counterfactual 2: Optimizing Over a Family of Quality Score Mechanisms

The results in the EFLB equilibrium remain qualitatively similar to those in the EFUB equilibrium, and the intuition is discussed in the main text (see Section 5.2.2).

| Quality score rule | Advertiser Surplus | Social Welfare |
|---|---|---|
| $q_j^0 = 1$ | \$181 | \$10,272 |
| $q_j^1 = \alpha_j$ | \$1,122 | \$14,292 |
| $q_j^2 = \alpha_j \gamma_j \overline{p}_j$ | \$1,396 | \$13,156 |
| $q_j^3 = \alpha_j (1 + \gamma_j \overline{p}_j f)$ | \$1,549 | \$15,612 |

Table 16: Predictions of (1) advertiser surplus and (2) social welfare (sum of retailer profit and advertiser surplus) from simulated quality score rules in EFLB equilibrium selection.

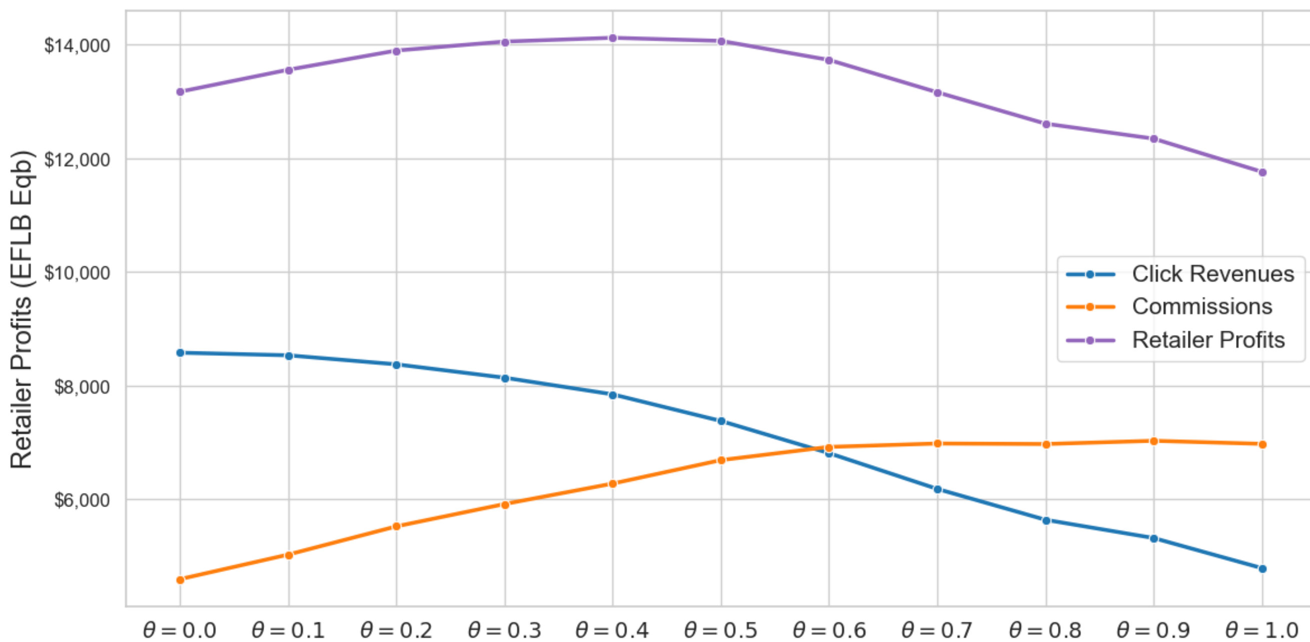


Figure 12: Predictions of (1) click revenues, (2) commission revenues, and (3) total retailer profits, in the EFLB equilibrium selection, for quality scoring rules of the form $q_j(\theta) = \alpha_j\big(\theta\gamma_j\overline{p}_j f + (1-\theta)\overline{b}\big)$, varying $\theta \in [0, 1]$ over a grid at intervals of 0.1.

Figure 12 shows retailer profits and its components in the EFLB equilibrium as a function of $\theta$, under the family of quality scoring mechanisms given by $q_j(\theta) = \alpha_j\big(\theta\gamma_j\overline{p}_j f + (1-\theta)\big)$, where $\theta \in [0, 1]$. As in the EFUB equilibrium, click revenues decrease with $\theta$, while commission revenues increase. The retailer continues to trade off the quantity of clicks with the profits from a click. Thus, total profits remain a concave function of $\theta$, and are maximized at $\theta = 0.4$.

Next, figure 13 shows advertiser surplus and social welfare in the EFLB equilibrium as a function of $\theta$, for the same set of quality scoring mechanisms. Advertiser surplus is maximized at $\theta = 0.6$ while the social welfare (sum of retailer profit and advertiser surplus) is maximized at $\theta = 0.5$.

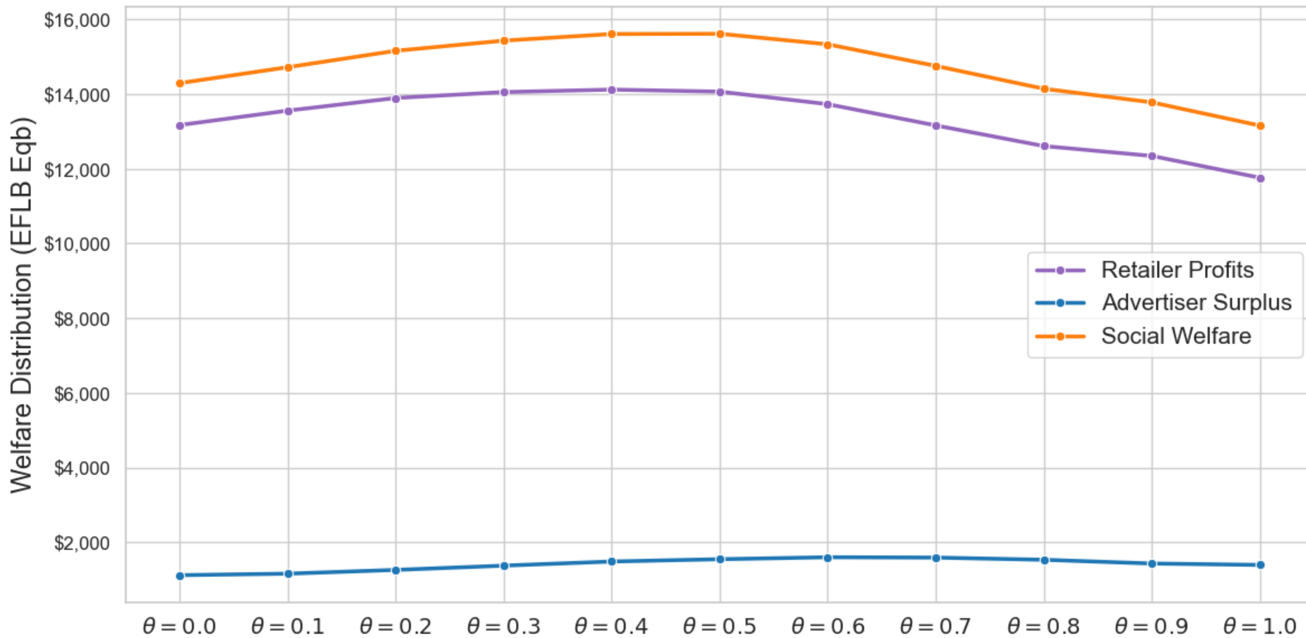


Figure 13: Predictions of (1) click revenues, (2) commission revenues, and (3) total retailer profits, in the EFLB equilibrium selection, for quality scoring rules of the form $q_j(\theta) = \alpha_j\big(\theta\gamma_j\overline{p}_j f + (1-\theta)\overline{b}\big)$, varying $\theta \in [0, 1]$ over a grid at intervals of 0.1.

### C.3 Counterfactual 3: Quantifying the Efficiency Revenue Tradeoff

Figure 14 outlines retailer profits and its components in the EFLB equilibrium as a function of $s$, under the quality score squashing mechanisms given by $\tilde{q}_j(s) = \big(\alpha_j(\gamma_j\overline{p}_j f + 1)\big)^s$, where $s \in [0, 1]$. Next, figure 15 describes advertiser surplus and social welfare in the EFLB equilibrium as a function of $s$, that is, for the same set of squashed quality scores.

As in the EFUB equilibrium, advertiser surplus decreases with increased squashing (that is, decrease in $s$), suggesting that quality score squashing hurts advertisers. At the same time, retailer profits in the EFLB equilibrium need not necessarily increase with the level of squashing, as in the EFUB equilibrium. This is because any increase in click revenues due to squashing is balanced by the decrease in commission revenues. Finally, because retailer profits remain flat while advertiser surplus decreases, squashing also reduces social welfare in the market.

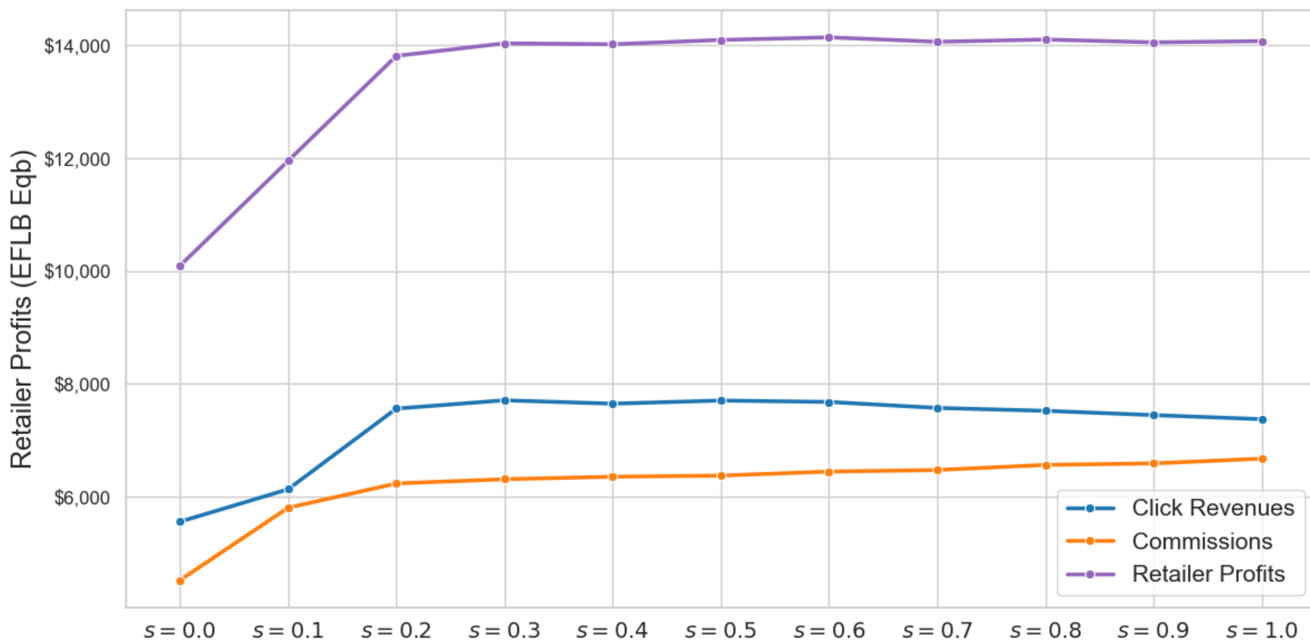


Figure 14: Predictions of (1) click revenues, (2) commission revenues, and (3) total retailer profits, in the EOS equilibrium selection, for quality scoring rules of the form $q_j(s) = \left(\alpha_j(\theta\gamma_j\overline{p}_j f + (1-\theta))\right)^s$, varying $s \in [0, 1]$ at intervals of 0.1.

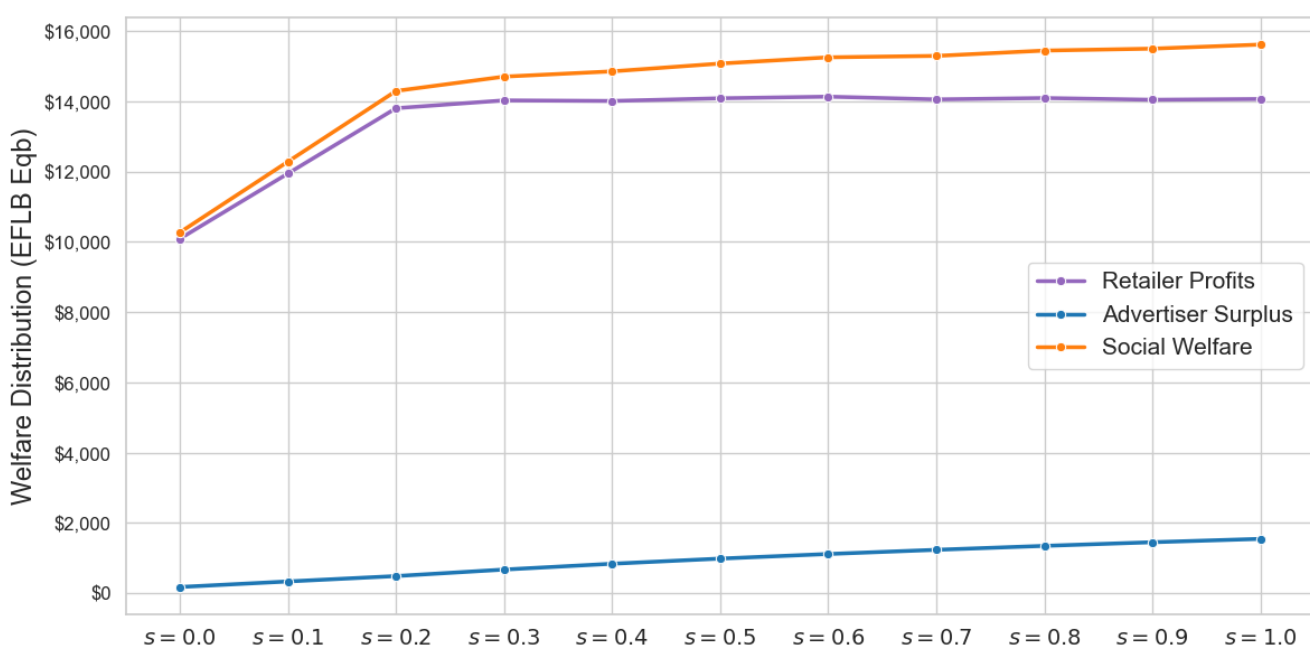


Figure 15: Predictions of (1) click revenues, (2) commission revenues, and (3) total retailer profits, in the EOS equilibrium selection, for quality scoring rules of the form $q_j(s) = \left(\alpha_j(\theta\gamma_j\overline{p}_j f + (1-\theta))\right)^s$, varying $s \in [0, 1]$ at intervals of 0.1.