The Welfare Effects of Sponsored Product Advertising

Chuan Yu*
Stanford University

October 2023

Click Here for the Latest Version

Abstract

In recent years, many digital retail platforms have significantly expanded their advertising businesses, prominently featuring sponsored products in search results. The welfare effects of sponsored product advertising are theoretically ambiguous: while it may enhance efficiency by enabling sellers to reveal information about their products, it may also worsen search results and raise prices. Using data on Amazon searches, purchases, and advertising auction bids, I estimate an equilibrium model to quantify the welfare effects. The model incorporates consumers' search frictions, sellers' pricing and bidding decisions, and the platform's choice of a commission rate. Counterfactual analysis shows that eliminating sponsored product advertising could benefit consumers and sellers if Amazon's commission rate is held fixed but would harm them if the commission rate were to adjust optimally, with consumer surplus falling by 4.3% and seller profits by 8.4%. Sponsored product advertising tends to be more beneficial in markets with fewer established products and greater quality differentiation.

*Department of Economics. Email: chuanyu@stanford.edu. I am particularly grateful to my advisors Liran Einav, Matthew Gentzkow, Michael Ostrovsky, and Ali Yurukoglu for their invaluable support and guidance. I thank Hunt Allcott, Claudia Allende, Lanier Benkard, Shengmao Cao, Jose Ignacio Cuesta, German Gutierrez, Brad Larsen, Pearl Li, Neale Mahoney, Adam Rosenberg, Paulo Somaini, Shoshana Vasserman, Frank Yang, and workshop participants at Stanford and University of San Francisco for helpful comments. I acknowledge Forum Brands for generously providing data. I acknowledge financial support from the SIEPR Program in Regulatory Policy and Leonard W. Ely and Shirley R. Ely Graduate Student Fellowship. All errors are my own.

1 Introduction

Sponsored product advertising has experienced dramatic growth on various digital retail platforms in recent years. When consumers search for products on these platforms, they typically encounter a mix of organic search results generated by the platform's ranking algorithm and sponsored results displaying the winners of real-time ad auctions. Amazon is the leader in this market, with its advertising revenues growing from $2 billion in 2016 to $38 billion in 2022 (Amazon 2017, 2023). Many other platforms offering a wide range of goods and services, such as Walmart.com, Expedia, Instacart, and UberEats, are also rapidly expanding their advertising ventures.[1] This shift represents a significant transformation in their business model.

The proliferation of sponsored product advertising has elicited a variety of reactions from platform users. Some consumers express frustration with the prevalence of sponsored products dominating search results, while others find sponsored products better suited to their needs than organic ones (e.g., Weise 2019; Fowler 2022). Some sellers complain about rising selling costs, while others view sponsored product advertising as an expansion opportunity (e.g., Matsakis 2022; Wang 2022). In September 2023, the U.S. Federal Trade Commission (FTC) launched a lawsuit against Amazon for "illegally maintaining monopoly power," with significant allegations related to its advertising business.[2] Ongoing discussions and the potential for policy regulations concerning sponsored product advertising highlight the importance of carefully examining its impact on the welfare of different market participants.

The welfare effects of sponsored product advertising are theoretically ambiguous. On one hand, sellers can utilize sponsored product advertising to reveal information about their products that may not be reflected in organic rankings, thereby enhancing efficiency by enabling high-quality or low-cost products to reach more consumers. On the other hand, the presence of low-quality sponsored products may harm consumers if they displace higher-quality alternatives. Additionally, sellers may pass on advertising expenses to consumers, resulting in higher prices. Finally, sponsored product advertising may be a substitute for other revenue-raising instruments for the platform. If it became unavailable, the platform might resort to other instruments to raise

[1] For instance, Instacart, a grocery delivery platform whose IPO was one of the most prominent of 2023, earned $740 million from advertising in 2022, accounting for nearly 30% of its total revenues (Browning and Griffith 2023). In 2023, the Interactive Advertising Bureau and Media Rating Council proposed a series of guidelines to standardize measurement for sponsored product advertising. Over 50 platforms contributed to the development of these guidelines.

[2] The FTC asserts that Amazon "degrade[s] the customer experience by replacing relevant, organic search results with paid advertisements—and deliberately increase[s] junk ads that worsen search quality and frustrate both shoppers seeking products and sellers who are promised a return on advertising purchase" (Federal Trade Commission 2023).

revenues, potentially exacerbating distortions elsewhere.

In this paper, I examine the welfare effects of sponsored product advertising on Amazon, the world's largest retail platform. My analysis consists of three main components. First, I introduce a stylized example that illustrates the roles of different economic factors in shaping the welfare effects of sponsored product advertising. Second, I compile extensive datasets containing search results and product information on Amazon, providing descriptive evidence on the forces at play. Finally, I develop and estimate an equilibrium model that incorporates interactions among consumers, sellers, and the platform. Using model estimates, I can simulate a counterfactual scenario in which sponsored products are replaced with organic listings and quantify the welfare effects of sponsored product advertising.

The main dataset for this paper includes a large number of search results from Amazon. I scraped search results for several thousand high-traffic keywords six times a day over a two-month period in 2022; these data allow me to track all organic and sponsored products on the first result page, along with their ranks, for over one million searches. I also collected data for about one million products that appeared in these search results, including daily prices, sales ranks, and total reviews. I utilized an established method to convert daily sales ranks into daily quantities. To validate the accuracy of this conversion, I obtained exact daily sales quantities for several hundred products over two years from a prominent Amazon seller. Finally, I collected aggregate statistics on the bids submitted by auction winners for each keyword in my sample.

I begin my analysis of the data by showing several descriptive findings on factors influencing welfare effects. First, I offer evidence of search frictions by illustrating that a product's position in search results can impact its sales, which motivates sellers to seek higher rankings in search results. Second, by comparing organic and sponsored products in similar locations on the same result pages, I demonstrate that sponsored product advertising enhances the visibility of less established products. Third, I show that top sponsored products have prices that are 16% higher than those of their organic counterparts.

Next, I develop an equilibrium model that incorporates interactions among market participants. In this model, consumers make purchase decisions subject to search frictions, considering only a subset of products in search results. Sellers set prices and submit bids in the ad auction to maximize their expected profits, accounting for the uncertainty arising from organic ranks and auction outcomes. The platform selects a commission rate, which is a percentage of the price paid by sellers, to maximize a combination of commissions, advertising revenues, consumer surplus,

and seller profits. By including the welfare of other participants in its objective, this approach approximates the platform's long-term investment considerations. The commission rate serves as a representation of a broader range of revenue-raising instruments available to the platform.

I identify the extent of search frictions by examining how daily variations in a product's organic ranks impact its sales while accounting for demand persistence. The result indicates the presence of substantial search frictions. To estimate product utility, I rely on consumers' revealed preferences, explicitly incorporating heterogeneity in their consideration sets. In most markets, organic ranks exhibit a moderate, although not perfect, correlation with product utility. Sponsored products tend to have lower average utility than their organic counterparts in similar locations. I derive sellers' marginal costs from the optimality condition in their profit-maximization problems. Despite sponsored products often having lower marginal costs, the additional advertising payment offsets and even reverses this cost advantage. Last, to determine the platform's objective function, I assume that the observed choice of commission rate is optimal.

To evaluate the welfare implications of sponsored product advertising, I simulate a counterfactual scenario in which all sponsored product slots are removed and replaced with organic listings. I examine two cases: one where the commission rate remains unchanged at the existing level and another where the platform can adjust the commission rate in response. I find that eliminating sponsored products while holding the commission rate constant results in a 12.6% increase in consumer surplus and a 14.1% rise in seller profits. These gains result from the removal of low-utility sponsored products and the elimination of the pass-through of advertising costs to prices. The platform collects more commissions, but its total revenues drop by 14.8%.

When the platform can reoptimize the commission rate, the model predicts a 4.7-percentage-point increase in the commission rate following the removal of sponsored products. With sponsored products, higher commission rates lower sellers' margins and willingness to pay for sponsored positions, thereby reducing the platform's advertising revenues. The elimination of sponsored products incentivizes the platform to raise the commission rate. Moving from the status quo to a counterfactual scenario with only organic search results and a higher commission rate leads to a 4.3% reduction in consumer surplus and an 8.4% decrease in seller profits. This reversal in the welfare comparison underscores the importance of policymakers' considering the platform's response when contemplating whether to impose restrictions on sponsored product advertising.

The effect of removing sponsored products on consumer surplus under endogenous commission rates exhibits substantial variation across markets. While it is positive on average, the effect

is negative in nearly 40% of the markets. This heterogeneity is closely related to the accuracy of organic rankings, measured as the correlation between product utility and organic ranks. A major advantage of sponsored product advertising is to allow sellers to reveal information about their products when organic rankings fail to reflect actual product utility. I find that sponsored product advertising tends to be more beneficial in markets with fewer established products or a significant presence of quality differentiation, often associated with less precise organic rankings.

As a final exercise, I consider counterfactual scenarios with different numbers of sponsored positions. The auction mechanism elicits sellers' private information and allows the platform to engage in price discrimination, typically resulting in higher platform revenues. However, whether the efficiency gains benefit other participants remains uncertain. My findings suggest that, while consumers and sellers benefit from sponsored product advertising relative to a scenario with only organic search results, introducing additional sponsored positions can come at the expense of consumers and sellers. Both the consumer- and seller-optimal numbers of sponsored positions are lower than the platform-optimal number. For policymakers concerned with the welfare of consumers and sellers, compared to a complete ban on sponsored product advertising, a cap on the total number of sponsored positions strikes a better balance that preserves the advantages of sponsored product advertising while safeguarding the interests of consumers and sellers.

There are several caveats to my analysis. First, this paper focuses on the short-term impacts of sponsored product advertising and may not capture longer-term effects, such as seller entry and exit. For instance, sponsored product advertising could alleviate the "cold-start" problem (Schein et al. 2002) by enhancing the visibility of new entrants, thereby encouraging more sellers to enter the market. However, the rise in advertising costs could diminish the profitability of selling on Amazon, potentially leading some sellers to leave the platform.

Second, my estimation of product utility relies on consumers' revealed preferences, assuming that consumers can accurately discern product utility. In reality, consumers may not observe product utility perfectly; instead, they might make inferences based on a product's position and whether it is sponsored. Consequently, the perceived utility of the same product may vary when it appears in different positions and could change when sponsored product advertising is removed. Additionally, consumers may exhibit aversions to sponsored products, which could lead to an underestimation of the utility of such products. In the appendix, I explore two extensions that partially relax this assumption, and the findings remain qualitatively robust.

Last, I abstract from several practices employed by the platform, such as the use of relevance

scores to rank sellers in auctions and personalized sponsored results, which aim to enhance the quality of sponsored products. By not incorporating them into my analysis, I may underestimate the benefits of sponsored product advertising. Moreover, my model assumes that organic search rankings are exogenous, which eliminates sellers' incentives to invest in organic ranks and precludes any changes to the ranking algorithm by the platform in counterfactual scenarios.

This paper contributes to several strands of literature. It builds upon extensive research in search advertising and position auctions, primarily focusing on search engines. Pioneering works by Edelman et al. (2007) and Varian (2007) examine the theoretical properties of position auctions. Subsequent studies consider richer settings. For instance, Athey and Ellison (2011) introduce consumer search and endogenize advertisers' valuations. Chen and He (2011) incorporate firms' pricing decisions. Empirical studies have examined sellers' dynamic bidding incentives (Yao and Mela 2011), endogenous valuations of sponsored positions (Chan and Park 2015), and consumer heterogeneity and the effectiveness of search advertising (Blake et al. 2015). This paper adapts the canonical models to the setting of sponsored product advertising, incorporating novel features of the environment, such as price competition and the platform's endogenous response.

Some recent theoretical studies examine the impact of sponsored product advertising on market outcomes and welfare. Ilango (2022) predicts that less prominent firms have a higher pass-through from advertising payment to prices, leading to lower competition and consumer surplus. Motta and Penta (2022) show that brand search advertising can decrease welfare due to reduced competition and higher prices. In this paper, I highlight various factors that influence the welfare effects in a stylized example. Importantly, my framework accounts for the platform's response. I also quantify the magnitudes of different forces in a high-stakes setting and examine the heterogeneity, providing insights that can inform regulatory efforts.

My work is also connected to research exploring the utilization of various instruments to enhance platform revenues. Choi and Mela (2019) analyze the trade-off between advertising revenues and commissions, exploring how various product-ranking and ad-pricing mechanisms impact platform profits. Long et al. (2022) examine how platforms can leverage information revealed in ad auctions to refine organic search rankings and set the commission rate. In my study, I incorporate the welfare of all market participants into the platform's objective and investigate how different platform designs affect the welfare of all participants. A developing literature examines Amazon's practices and their implications for welfare (e.g., Lee and Musolff 2023; Gutierrez 2022; Lam 2023; Farronato et al. 2023). I focus on sponsored product advertising, a rapidly growing

sector within Amazon and numerous other platforms, which is also highly policy-relevant. The framework developed and lessons learned have broader implications beyond this setting.

This paper is broadly related to the literature on slotting fees, which are payments from manufacturers to retailers for shelf space. The literature has debated whether slotting fees are procompetitive (e.g., Sullivan 1997) or anticompetitive (e.g., Marx and Shaffer 2010). Sudhir and Rao (2006) find that the use of slotting fees for allocating shelf space is efficient and helps manufacturers signal private information. Sponsored product advertising is the e-commerce analog, and this paper contributes new evidence using data on placement and payment. It finds that sponsored products deliver lower utility and can displace higher-utility alternatives. However, sponsored product advertising can be beneficial in markets with fewer established products.

The remainder of the paper proceeds as follows. Section 2 describes the institutional setting. Section 3 presents a stylized example to illustrate determinants of the welfare effects. In Section 4, I summarize the data and present descriptive results. I develop and estimate an equilibrium model in Section 5. Section 6 discusses estimation results. In Section 7, I conduct counterfactual exercises to quantify the welfare effects of sponsored product advertising. Section 8 concludes.

2 Setting

This paper focuses on Amazon, the world's largest digital retail platform, which generated a total sales volume of $600 billion in 2021 (Kaziukėnas 2022a). In the U.S., Amazon holds a market share of approximately 40% in the e-commerce retail industry (Lebow 2022). Amazon operates both as a retailer, directly selling products to consumers, and as a marketplace where third-party sellers can offer their own products. As of 2021, the platform hosted over six million sellers worldwide, who contributed 65% of the total sales on the platform (Kaziukėnas 2022b).

When consumers visit Amazon's website or use its mobile app, they usually start by entering a keyword into the search box, which directs them to a page displaying search results. Consumers can browse through the listed products and make their purchase decisions. Each listing in the search results includes a product image, a brief description, price, delivery options, total reviews, and an average consumer rating. Clicking on a listing takes consumers to a product page with additional information. Appendix Figure A.1 shows an example of search results.

There are two main types of search results. The majority of search results are organic, determined by Amazon's ranking algorithm based on various factors, such as relevance to the searched

keyword, sales performance, and consumer reviews. The remaining results are sponsored, for which sellers pay the platform to display their products. Sponsored products are labeled onscreen as "Sponsored" to differentiate them from organic results. A product can appear as both an organic and a sponsored result. Typically, when viewed on a desktop or laptop browser, search results are delivered in 60 positions on each page, usually arranged in 15 rows and 4 columns.[3] In most cases, there are 12 sponsored positions among the 60 total; these can appear at the top, middle, or bottom of the search results.[4] On mobile devices, products are stacked vertically, and consumers can scroll down to view more products.

To advertise a product in search results, a seller selects a set of keywords and submits keyword-specific bids.[5] When a consumer enters a keyword, Amazon conducts a generalized second-price auction among all sellers who bid on that keyword (Varian 2007; Edelman et al. 2007). In principle, sellers are ranked based on a combination of their bids and platform-assigned scores that measure products' relevance to the searched keyword. However, as I do not observe the relevance scores used by Amazon, in my model, I assume that products are ranked solely based on their bids. Under the auction rule, the seller with the r-th highest bid is displayed in the r-th sponsored position. Payment is made for each click, irrespective of whether the click leads to a transaction. When a consumer clicks on a sponsored listing in the r-th sponsored position, the seller pays Amazon an amount equal to the $(r+1)$-th highest bid.

Auction outcomes on Amazon can be highly dynamic. The assortment of products appearing in the sponsored results can vary between two searches of the same keyword conducted within minutes of each other. This variability can result from a range of real-world factors and has been documented and studied in the literature (e.g., Athey and Nekipelov 2011; Ostrovsky and Skrzypacz 2022).[6] To assist sellers in making informed bidding decisions, Amazon provides them with aggregate statistics on winning bids for each keyword. Specifically, it reports estimates of the

[3]There are also nonstandardized positions for displaying products in a carousel layout, such as for products highlighted as "editorial picks" or "climate-pledge friendly." These products are excluded from my analysis. Consumers can also navigate to later pages in the search results. According to Amazon data, 81% of clicks happen on the first page.

[4]Appendix Figure A.2 illustrates the distribution of sponsored positions in search results. In my sample, the most common sponsored positions are 1–4, 17–18, 31–33, and 58–60. There are also less common result pages containing 22 positions in a single column, displaying larger images for products such as electronic devices.

[5]Sellers can choose between automatic or manual targeting. With automatic targeting, a seller does not need to specify keywords; instead, Amazon matches the product to relevant keywords. With manual targeting, there are three match types: *broad match*, where the search term contains the keyword in any order or close variations; *phrase match*, where the search term contains the keyword; and *exact match*, where the search term exactly matches the keyword.

[6]Bidders may set limits on their expenditures, prompting the platform to distribute their participation in auctions over time. The platform may personalize advertising based on consumers' locations or purchase histories. Both bidders and the platform can also experiment with different advertising campaigns or product arrangements.

median, lowest, and highest bids among recent auction winners for each keyword.

When a consumer makes a purchase on the platform, Amazon collects a commission fee from the seller. This fee is calculated as a percentage of the selling price and varies depending on the product category. Appendix Figure A.3 illustrates the commission rates across categories, ranging from 8% to 20%. For most products in my analysis, the commission rate is either 15% or 17%. Additionally, Amazon may collect storage and shipping (fulfillment) fees from sellers, which are usually fixed amounts based on product sizes. They may also collect a Prime membership fee from consumers, which includes free shipping among its benefits.

As previously mentioned, products on Amazon are sold by either third-party sellers or Amazon itself. In the former case, sellers set retail prices and pay commissions to Amazon for each unit sold. In the latter case, Amazon purchases products from suppliers at negotiated wholesale prices and determines retail prices. Both third-party sellers and suppliers can advertise their products on Amazon. In my analysis, I do not differentiate between suppliers and third-party sellers. I assume that they both set retail prices, pay Amazon commissions, and make advertising decisions.[7]

3 A Stylized Example

In this section, I present a stylized example to illustrate the economic factors that drive the welfare effects of sponsored product advertising. This example captures the essential elements: consumer search frictions, seller competition, and a profit-maximizing platform.

3.1 Setup

Consider a market with two sellers, labeled $j = 1, 2$, each offering a single product with quality δ_j. The platform's search results consist of two positions displaying organic or sponsored results. Without loss of generality, I assume that product 1 is ranked higher in the organic rankings, even though this ranking may not necessarily reflect the actual ranking of δ_j. This discrepancy can occur when organic rankings do not fully capture the quality of less established products. The platform collects a commission for each unit sold, calculated as a percentage of the product's price.

A fraction of consumers consider only the product in the first position, while the remaining consumers consider both positions. This assumption approximates real-world search behavior,

[7] In cases where multiple sellers offer the same product, Amazon combines them into a single listing in search results and selects a default seller, usually based on price. While consumers can purchase from other sellers, approximately 80% of sales occur through the default seller (Lee and Musolff 2023). Only the default seller can advertise the product.

where consumers may incur search costs and evaluate only a subset of products in search results. This behavior incentivizes sellers to pursue higher rankings in search results. Consumers have idiosyncratic preferences and may choose either a product that they consider or an outside option.

I first examine a scenario where both positions display organic search results. In this case, product 1 appears in the first position and is considered by all consumers, while product 2 is considered by only a subset of consumers. Each seller sets a price to maximize its profits given its rank in search results. The platform chooses a commission rate that maximizes total commissions.

Next, I consider a scenario where the first position is allocated through an ad auction. The auction outcome is stochastic, with the seller submitting a higher bid having a better chance of winning the auction. The auction winner is placed in the first position and pays a price for each click. The second position displays the top-ranked organic product (product 1). Each seller sets a price and submits a bid to maximize its expected profits. The platform chooses a commission rate to maximize the sum of commissions and advertising revenues. Appendix Figure A.4 visually illustrates the setup, and Appendix A provides formal definitions of consumer demand and the auction rule and derives the equilibria in both scenarios.

3.2 Simulation Results

I simulate the welfare implications of the introduction of a sponsored position by comparing the two scenarios described above. In these simulations, I assume that both products have the same marginal cost. I hold the quality of product 1 constant while varying the quality of product 2.

Sponsored product advertising can affect welfare through multiple channels. To investigate the contribution of each channel to the overall effect, I explore three assumptions concerning how sellers and the platform can respond to the introduction of a sponsored position. In the most restrictive scenario, sellers cannot adjust prices, and the platform cannot modify the commission rate; sellers determine their bids taking prices as given. In the second assumption, sellers have the flexibility to adjust their prices, while the commission rate remains unchanged. In the most flexible assumption, sellers can set new prices, and the platform can select a different commission rate. For each of these assumptions, I calculate the difference in market outcomes and welfare under the equilibrium with solely organic results and those under the equilibrium with a sponsored result.

Figure 1 illustrates the change in consumer surplus following the introduction of a sponsored position.[8] As depicted in the left column with fixed prices and commission rates, consumer sur-

[8] Appendix Figure A.5 presents additional simulation results on seller profits, platform revenues, and total surplus.

plus decreases when product 2 has lower quality than product 1, meaning that organic ranks reflect the actual quality ranking. The intuition is that, due to the uncertainty in auction outcomes, both products can win the sponsored position. When product 2 wins, it displaces the higher-quality product in the more prominent position, disadvantaging consumers who consider only the first product. When product 1 wins, it occupies both the sponsored and organic positions, crowding out product 2 and reducing product variety for consumers considering products in both positions. In either case, consumers are worse off.

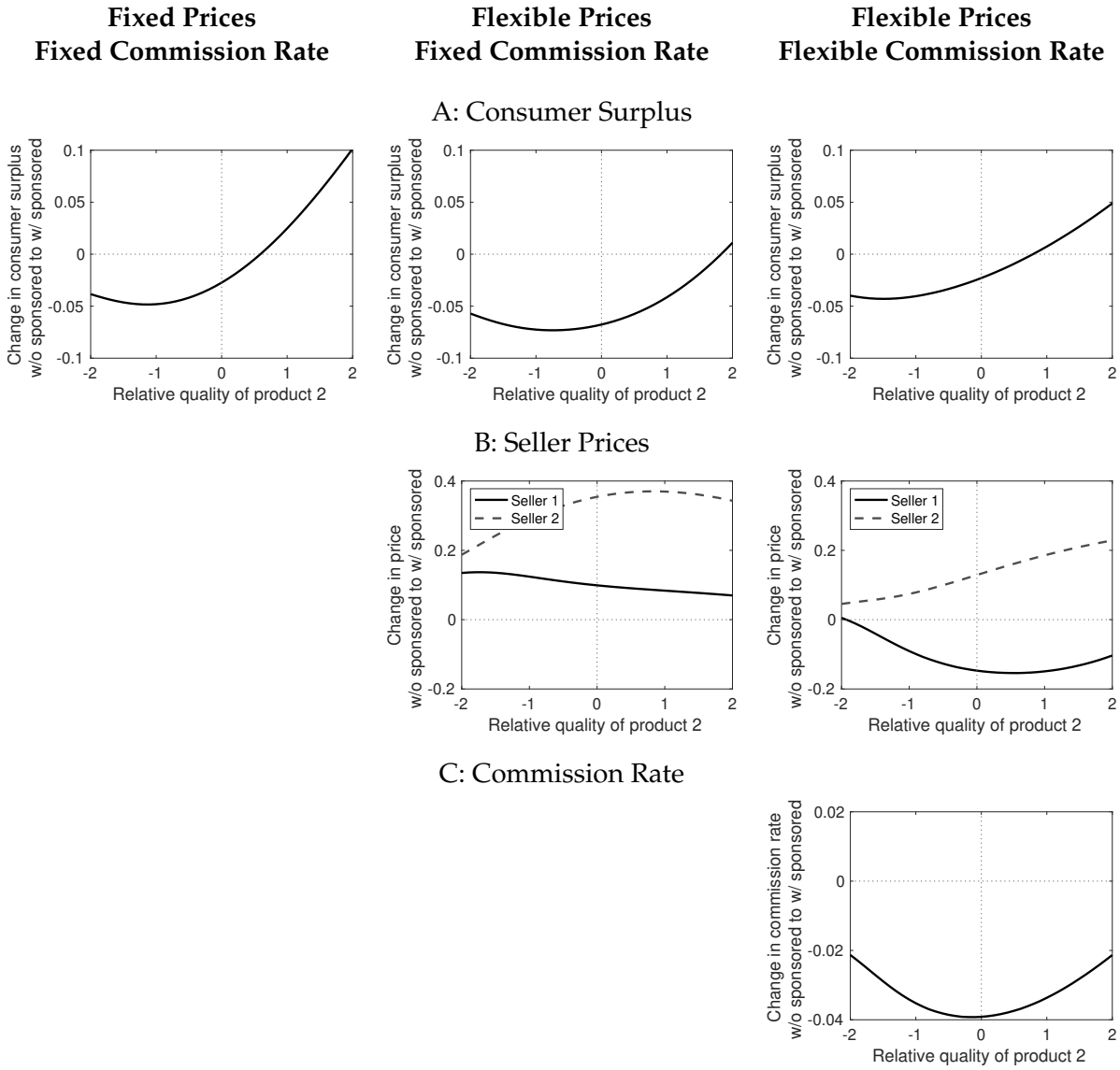
However, when product 2 offers higher quality, sponsored product advertising can enhance consumer surplus. In this scenario, product 2 wins the sponsored position most of the time, as it anticipates greater gains from reaching more consumers. Consumers who consider only the product in the first position now have access to a higher-quality alternative. This result suggests that, in cases where organic rankings fail to accurately reflect product quality, sponsored product advertising can benefit consumers by increasing the visibility of higher-quality but lower-ranked products. This advantage is particularly relevant for less established products.

The second column of Figure 1 shows that consumer welfare diminishes when sellers can adjust their prices after the introduction of a sponsored position while the commission rate is held fixed. This decline occurs because sellers pass on advertising expenses to consumers through higher prices. Since prices are strategic complements, the price of the product that loses the auction also rises. The extent to which advertising expenses are passed on to prices depends on consumers' price sensitivity.

When I further allow the platform to modify its commission rate (right column of Figure 1), a notable finding is that the commission rate is lower with sponsored products, as illustrated in Panel C. Sellers' willingness to pay for the sponsored position is closely linked to their profit margins. A higher commission rate lowers sellers' margins and reduces their bids in the ad auction, leading to lower advertising revenues for the platform. As a result, the introduction of a sponsored position incentivizes the platform to lower its commission rate, which in turn reduces prices and benefits consumers. The magnitude of the change in the commission rate depends on the platform's objectives, which may extend beyond short-term revenues in reality.

This stylized example underscores the theoretical ambiguity in the welfare effects of sponsored product advertising, which are shaped by various factors. These factors include the relationship between product quality and organic rankings, the extent of overlap between sponsored and organic products, consumers' price sensitivity, and the platform's objectives when it determines its

Figure 1: Simulation Results of the Stylized Example



Notes: This figure presents the simulation results of the stylized example. For each panel, the y-axis shows the change in outcomes from the equilibrium with two ranked organic positions and no sponsored positions, to the equilibrium with a sponsored position at the top and an organic position at the bottom. The x-axis represents the relative quality of product 2 compared to product 1. Each column corresponds to a different assumption regarding how sellers and the platform can respond to the introduction of a sponsored position. In the first column, neither sellers can change prices nor can the platform change the commission rate. Sellers choose bids to maximize profits taking their prices as given. In the second column, sellers can set new prices, but the platform cannot change the commission rate. In the third column, sellers can set new prices and the platform can set a new commission rate. Panels A to C display the changes in consumer surplus, the prices of both products, and the commission rate, respectively. See Section 3 for the setup of the stylized example and Appendix Figure A.5 for additional simulation results.

commission rate. Subsequent analyses will measure these factors for Amazon to assess the welfare impact of sponsored product advertising.

4 Data and Descriptive Results

4.1 Data

The main dataset comprises a large set of scraped search results from Amazon. I selected 3,237 high-traffic keywords that span a wide range of product categories and have a daily search volume of 8.8 million.[9] I grouped keywords that generated search results with substantial overlap into the same market, resulting in a total of 546 markets. Appendix B.1 describes the selection of keywords and the construction of markets. Appendix Table A.1 offers examples of several markets and their associated keywords. The descriptive results in this section utilize all keywords, while the estimation in Section 5 focuses on the highest-search-volume keyword within each market.

Between May 1 and July 5, 2022, I conducted six searches per day on Amazon for each keyword in the sample. These searches occurred from 8 a.m. to 6 p.m. Pacific Time, with a two-hour interval between searches.[10] For each search, I observed all sponsored and organic results on the first page and their ranks. I also recorded product characteristics displayed in the search results, including price, total reviews, average consumer rating, eligibility for Prime shipping, and whether the product was labeled "Amazon's Choice." In total, I collected results from over 1.4 million searches.

I collected additional product information from Keepa.com, a platform that monitors products sold on Amazon. For each product that appeared in the collected search results, I identified its initial listing date on Amazon, the category to which it belonged, and its daily Best Sellers Rank (BSR) in that category. Amazon describes the BSR as a ranking system used to assess a product's sales performance relative to that of other products within the same category. I followed an established method outlined in Appendix B.2 to convert a product's daily BSRs into daily quantities. To validate this conversion method, I compared the estimated daily quantities based on BSRs with exact daily quantities for several hundred products over a two-year period. I obtained these exact daily quantities from a prominent Amazon seller. As shown in Appendix Figure A.6, the estimated and exact daily quantities align remarkably well, both in the cross-section and for high-frequency variation over time. This alignment confirms the accuracy of the conversion method.

Finally, I collected the median, lowest, and highest winning bids reported by Amazon for all the keywords used in my analysis as of May 2022. These statistics were updated daily, but upon

[9] I obtained the data on search volumes from Jungle Scout, a prominent e-commerce intelligence service provider.

[10] To facilitate the web-scraping process from Amazon's public domains, I utilized the Rainforest API tool. Each request specified a keyword and scrape time and acted as an anonymous shopper on Amazon.com. This tool routed requests through a vast network across the U.S., allowing the completion of thousands of requests within minutes. This approach provided more representative results than could be obtained through repeated searches in a single location.

manual examination of a sample, I found limited variation over time during my sample period.[11] Appendix Figure A.7 displays the distribution of the median winning bids.[12]

Table 1 presents summary statistics of the data. Panel A reports the search volume and bid statistics for all keywords in the sample. The median keyword has a daily search volume of 1,088 and a median winning bid of \$1.11, with the lowest and highest winning bids being \$0.72 and \$1.31, respectively. There is moderate variation across keywords. Panel B describes the characteristics of searches. An average result page contains 42 organic positions and 11 sponsored ones. Approximately 81% of searches yield 60 results on the first page, and my analysis focuses on these result pages. Panel C summarizes the characteristics of all the products in these search results. The median product has a price of \$28 and an average daily quantity of 46 units and has been listed on Amazon for approximately three years.

4.2 Positions in Search Results and Product Sales

The stylized example presented in Section 3 is built upon the assumption that a larger proportion of consumers consider the product in the top position. This assumption reflects the existence of search frictions and provides sellers with an incentive to be placed higher in search results. In this section, I analyze the impact of a product's position in search results on its sales.

Appendix Figure A.8 illustrates a negative relationship between a product's position in search results and its sales, conditional on product and keyword–day fixed effects. However, this correlation does not imply a causal relationship, as a product's position in search results is not random and can be influenced by its sales performance. Therefore, I use a regression framework that allows me to flexibly control for unobserved demand to partially mitigate the endogeneity problem.

Let q_{jt} represent the sales quantity of product j on day t and pos_{jt} denote its average position in the search results for a specific keyword.[13] I estimate the following regression model:

$$\log(q_{jt}) = \beta_p pos_{jt} + \sum_{\tau=1}^{3} \rho \log(q_{j,t-\tau}) + \phi_{jw} + \phi_{kt} + \varepsilon_{jt}. \tag{1}$$

[11] I obtained the median winning bids for 1,483 keywords on five days and for each keyword computed the coefficient of variation (CV), which is the standard deviation divided by the mean. The median CV is 0.12, indicating low variation.

[12] While I observed the median winning bids for all keywords, I observed the lowest and highest bids only for a subset of keywords. Panel B of Appendix Figure A.7 shows a nearly linear relationship between the median, highest, and lowest winning bids. I imputed the lowest and highest winning bids for keywords with missing data.

[13] A smaller value of pos_{jt} indicates a higher position in search results, with $pos_{jt} = 1$ representing the first position. In cases where a product appears in the search results for multiple keywords, I retain only the keyword with the highest frequency of appearance to ensure that each product is associated with a unique keyword.

Table 1: Summary Statistics

	N	Mean	SD	P25	Median	P75
A: Keyword						
Daily Search Volume	3,237	2,775.28	4,050.84	534.43	1,088.23	3,522.77
Median Bid ($)	3,237	1.36	0.89	0.84	1.11	1.57
Lowest Bid ($)	3,237	0.88	0.58	0.54	0.72	1.01
Highest Bid ($)	3,237	1.60	1.05	0.99	1.31	1.85
B: Search						
Total Positions	1,281,845	52.53	15.15	60	60	60
Organic Positions	1,281,845	41.88	12.58	48	48	48
Sponsored Positions	1,281,845	10.66	2.69	12	12	12
Total Positions = 60	1,281,845	0.81	0.39	1	1	1
C: Product						
Price	64,617,245	83.35	248.49	16.50	27.90	60.77
Daily Quantity	62,939,124	132.67	243.25	11.63	45.52	143.53
Consumer Rating (1-5)	64,434,914	4.44	0.33	4.30	4.50	4.70
Total # Reviews	64,430,254	6,798.86	18,433.52	215.00	1,355.00	5,871.00
Listed Time (Years)	3,836,174	3.87	3.58	1.34	2.97	5.06
Eligibile for Prime Shipping	66,157,677	0.86	0.35	1.00	1.00	1.00
Amazon's Choice	66,157,677	0.02	0.12	0.00	0.00	0.00
Sponsored Product	66,157,677	0.19	0.39	0.00	0.00	0.00

Notes: This table provides summary statistics for the datasets used in the analysis. Panel A presents keyword-level information. Data on search volumes and bids are collected from Jungle Scout. Panel B reports statistics at the search level. I conducted six searches a day for each keyword on Amazon from May 1 to July 5, 2022. Panel C presents product-level characteristics for all products that appeared on the first page of the search results. Daily quantities are estimated based on Amazon's best sellers ranks. See Appendix B.2 for more details.

Here, ϕ_{jw} represents product–week fixed effects, and ϕ_{kt} represents keyword–day fixed effects. In Column 1 of Table 2, I report the estimate of β_p while including product fixed effects instead of product–week fixed effects and without accounting for lagged product sales quantities. This estimate corresponds to the slope depicted in Appendix Figure A.8.

The endogeneity concern arises because a product's unobserved demand may simultaneously affect both its current sales and its organic rank. In Column 2, I introduce product–week fixed effects into the regression to account for short-term seasonality in a product's demand. In Column 3, I further include three-period lagged product sales quantities as controls in the regression, which helps mitigate the influence of persistent demand patterns across days. With these variables included, I observe a smaller coefficient for β_p. Nevertheless, it remains statistically significant with a meaningful magnitude. The estimate in Column 3 suggests that moving up one position higher in the search results for a keyword increases a product's sales by 0.36%.

These estimates aggregate products across all positions, but the impact of positions on sales

Table 2: Relationship Between Positions and Sales

	Dep. Var. = Log(Sales)					
	(1)	(2)	(3)	(4)	(5)	(6)
Position	-0.0125	-0.0054	-0.0036	-0.0558	-0.0213	-0.0169
	(0.00010)	(0.00006)	(0.00005)	(0.00821)	(0.00338)	(0.00255)
L1. Log(Sales)			0.2535			0.3322
			(0.00204)			(0.03155)
L2. Log(Sales)			0.0035			0.0264
			(0.00130)			(0.02160)
L3. Log(Sales)			-0.0555			-0.0503
			(0.00130)			(0.02318)
N	2,733,256	2,681,115	2,606,483	19,500	19,142	18,668
Product FE	Yes	Yes	Yes	Yes	Yes	Yes
Product-Week FE	No	Yes	Yes	No	Yes	Yes
Keyword-Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Products	All	All	All	Top	Top	Top
R^2	0.959	0.985	0.986	0.984	0.994	0.995

Notes: This table examines the relationship between a product's position in search results and its sales quantity. Each observation corresponds to a product on a given day. The dependent variable is the log of the product's total sales quantity. The variable *Position* denotes the average position of the product in the search results for a specific keyword on a particular day, with lower values indicating higher positions in search results. When a product appears in the search results for multiple keywords, I retain only the keyword with the highest frequency of appearance to ensure that each product is linked to a unique keyword. Column 1 presents the estimate of β_p in equation (1) while including product fixed effects instead of product–week fixed effects and without accounting for lagged product sales quantities. Column 2 includes product–week fixed effects in the regression. Column 3 further includes three-period lagged product sales quantities in the regression. The sample includes products that appear in the search results for a keyword in more than 50% of the searches. Columns 4–6 replicate the regressions in columns 1-3 but restrict the analysis to products with an average position smaller than five. Standard errors are clustered at the product level and are reported in parentheses.

may not be uniform across products. For instance, products ranked higher in search results might exhibit a more pronounced sales response to changes in position. In Columns 4–6 of Table 2, I present results from the same regressions as in Columns 1–3, but I restrict the analysis to products with an average position smaller than five. The coefficients for this subset of products are about four times larger, indicating a stronger impact of position on sales for top-ranked products.

However, these regressions do not entirely eliminate the endogeneity issue. While the regressions incorporate flexible controls for recent underlying demand, they do not rule out the possibility of a contemporaneous demand shock that simultaneously affects sales and alters the product's position on the same day. In Section 5.1.2, I will further tackle the endogeneity problem by making structural assumptions about how demand shocks and organic ranks evolve over time.

4.3 Comparison of Sponsored and Organic Products

The stylized example in Section 3 emphasizes the importance of the relationship between products' organic ranks and quality in the welfare effects of sponsored product advertising. When organic rankings fail to accurately reflect product quality, sponsored product advertising can benefit consumers by enabling high-quality products to gain visibility. However, it may harm consumers if low-quality sponsored products displace high-quality organic alternatives. Therefore, I compare sponsored and organic products at the top of search results.

Figure 2 compares sponsored and organic products across four dimensions: the log of total reviews, years listed on Amazon, average consumer rating, and the log of product price. The first two variables can provide insights into a product's presence and visibility on the platform. Although the average consumer rating could serve as a measure of product quality, it is imperfect and may be subject to manipulation. Therefore, it should be interpreted with caution.

This analysis focuses on products that occupy the top eight positions in each search, typically containing four sponsored and four organic products. Let j index products and f index searches. I residualize each outcome variable y_{jf} by search fixed effects and recenter it at its sample mean, denoted as $\tilde{y}_{jf}$. Each panel of the figure displays the distribution of $\tilde{y}_{jf}$ separately for sponsored and organic products, along with the estimated coefficient β_S from the regression:

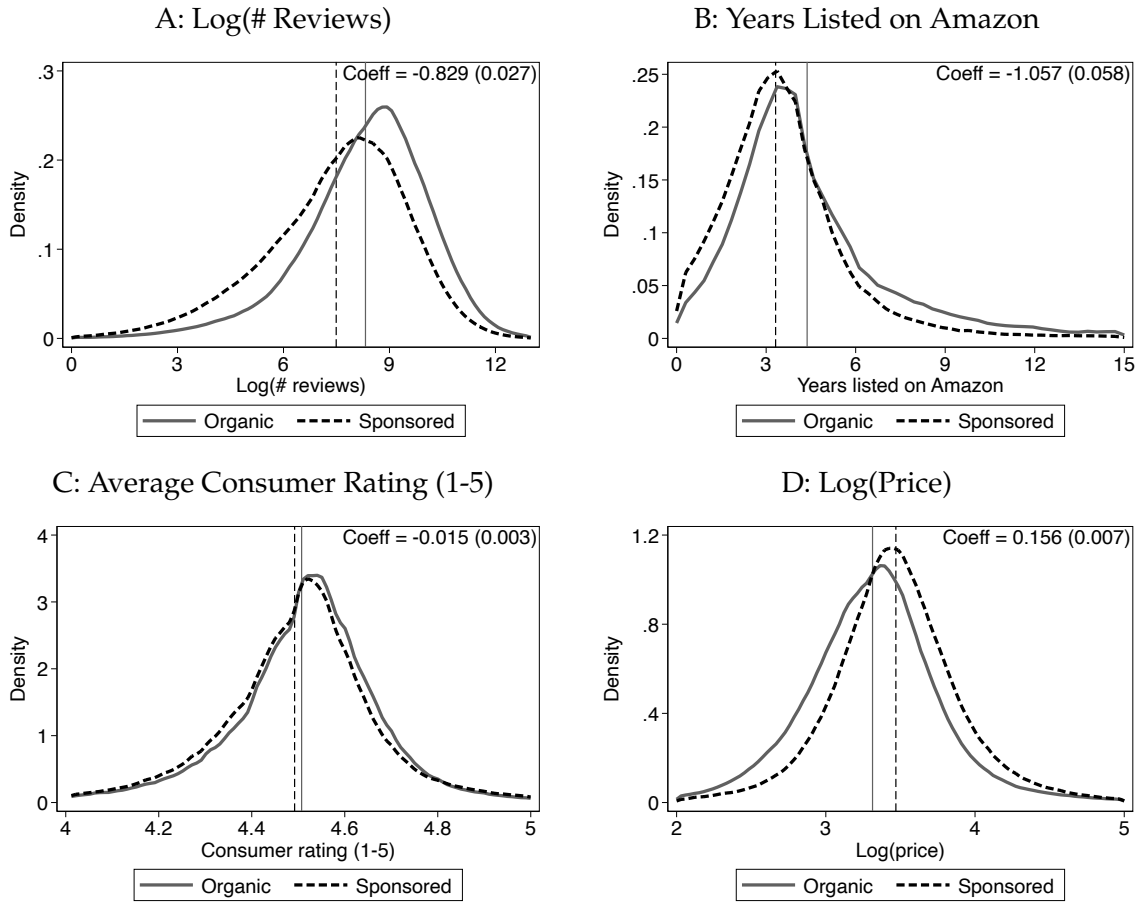
$$\tilde{y}_{jf} = \beta_0 + \beta_S Sponsored_{jf} + \epsilon_{jf}. \tag{2}$$

Here, $Sponsored_{jf}$ is an indicator for sponsored products, and β_S quantifies the average difference between sponsored and organic products in similar positions on the same result pages.

The first two panels of Figure 2 indicate that, on average, top sponsored products have 83% fewer reviews and a one-year shorter listing duration than top organic products. These findings suggest that sponsored product advertising alters the set of products that consumers encounter and enhances the visibility of less-established products. It is plausible that their true utility is not accurately reflected in organic rankings and, as a result, they may benefit more from advertising.

Panel C of Figure 2 demonstrates that top sponsored products have slightly lower average consumer ratings than their organic counterparts. While this comparison may suggest that sponsored products are not significantly worse, it is crucial to note that the average consumer rating is a noisy and potentially manipulable measure of quality. In my analysis in Section 5, I estimate product quality based on consumers' revealed preferences.

Figure 2: Comparing Top Sponsored and Organic Products



Notes: This figure compares sponsored and organic products in top positions. I restrict to products located in the top eight positions in each search, typically including four sponsored and four organic products. For each panel, I residualize the outcome of interest by search fixed effects and recenter it to its sample mean. I then plot the distribution of the residualized outcome separately for sponsored and organic products, with the vertical lines indicating the corresponding means. In Panels A to D, the variables are the log of the number of reviews, years listed on Amazon, the average consumer rating on a scale of 1 to 5, and the log of the product price, respectively. The coefficient reported in each panel is estimated from equation (2). Standard errors are clustered at the product level and reported in parentheses.

The last panel of Figure 2 illustrates that, on average, top sponsored products have prices 16% higher than those of their organic counterparts. This price difference could be attributed to the additional payment required for sponsored placements, raising concerns that if advertising costs lead to higher prices, sponsored product advertising could negatively impact consumers even if it helps certain high-quality products reach more consumers.[14]

Last, sponsored product advertising may harm consumers by causing duplication when the

[14]In Appendix Figure A.9, I estimate equation (2) separately for each keyword in my sample and plot the distribution of β_S across keywords. The relationships between sponsored and organic products hold for the majority of keywords.

same product appears in both a sponsored and an organic position. To assess the extent of overlap between sponsored and organic products, I examine whether each sponsored product also appears in an organic position on the same result page. Appendix Figure A.10 illustrates that over 40% of the top four sponsored products also appear in organic positions on the same result page. This percentage decreases to 27% when I consider all sponsored products.

So far, I have presented evidence of multiple factors at play. While sponsored product advertising could enhance the visibility of less established products, it could also raise prices and reduce product variety. However, these findings do not fully capture the equilibrium effects. If sponsored product advertising were eliminated, sellers would adjust their prices, and the platform would also respond accordingly. In the next section, I will develop and estimate an equilibrium model to account for these various trade-offs.

5 Model and Estimation

I develop and estimate a three-stage equilibrium model. First, the platform selects a commission rate to maximize its objective. Then, sellers make pricing and advertising decisions to maximize expected profits. Finally, consumers, facing search frictions, make purchase decisions. I will describe each stage of the model in reverse order and then discuss the estimation procedure.

5.1 Demand

5.1.1 Model

I define a market as a specific keyword. When consumers enter the keyword, they access the search results and choose from the listed products or an outside option. I focus on the first page of the search results. Since I do not have access to click data, I model the consumer's purchase decision without delving into the details of the browsing process. I also abstract from the broader search process, such as which keyword to use and whether to search for another keyword.

In each market, products are indexed by $j \in \mathcal{J}$. The search results contain N positions indexed by n. Out of these, a total of R positions indexed by r are designated as sponsored positions and allocated through ad auctions. The remaining positions display organic results determined by the platform's ranking algorithm. In the following estimation, $N = 60$ and $R = 12$. Let Γ denote the product arrangement in the search results, where $\Gamma_n \in \mathcal{J}$ refers to the product in the n-th position. Let $\mathcal{J}_n(\Gamma) = \{\Gamma_1, \ldots, \Gamma_n\} \subset \mathcal{J}$ represent the set of products in the top n positions for $1 \leq n \leq N$.

Consumer i's utility from purchasing product j on day t is expressed as follows:

$$u_{ijt} = \phi_j + \psi_t + \psi_j t - \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt} = \delta_{jt} + \epsilon_{ijt}. \tag{3}$$

Here, ϕ_j denotes the time-invariant product quality, and ψ_t is a day fixed effect. $\psi_j t$ represents a product-specific linear time trend, which captures a product's trajectory over time.[15] p_{jt} is the price, and ξ_{jt} is an unobserved demand shock. ϵ_{ijt} is an idiosyncratic preference shock following a type I extreme value distribution. The mean utility of the outside option, or not buying any product on the first page of the search results, is normalized to zero. I assume that consumers can observe δ_{jt} for all products that they consider; therefore, they do not make inferences about product utility based on whether a product is labeled as sponsored.[16]

Consumers may have different consideration sets in the form of the top n products in search results, or $\mathcal{J}_n(\Gamma)$. Let $\lambda_n \in [0,1]$ represent the fraction of consumers whose consideration sets contain the product in the n-th position. All consumers consider the first product, so $\lambda_1 = 1$. Only a subset of consumers consider the second product, so $\lambda_2 \leq 1$. Among consumers who consider the n-th product, only a subset of them consider the $(n+1)$-th product, so $\lambda_{n+1} \leq \lambda_n$. Given that choosing a product beyond the first page is part of the outside option, $\lambda_{N+1} = 0$. In general, a fraction $\lambda_n - \lambda_{n+1} \geq 0$ of consumers have the consideration set $\mathcal{J}_n(\Gamma)$.[17]

Each consumer chooses either a product from her consideration set or the outside option. Given prices of all sellers $\boldsymbol{p}$ and product arrangement Γ, product j's market share on day t is:

$$s_{jt}(\boldsymbol{p}, \Gamma) = \sum_{n=1}^{N} \mathbf{1}\big(j \in \mathcal{J}_n(\Gamma)\big)(\lambda_n - \lambda_{n+1}) \frac{\exp(\delta_{jt})}{1 + \sum\limits_{j' \in \mathcal{J}_n(\Gamma)} \exp(\delta_{jt'})}. \tag{4}$$

Unlike a standard logit model, in equation (4), a product's market share depends on the product arrangement Γ and the population shares of different consideration sets, $\lambda_n - \lambda_{n+1}$.

[15] Appendix G.2 demonstrates that the results are robust to my dropping the linear time trend, substituting it with product–month fixed effects, or incorporating a product-specific quadratic time trend.

[16] The empirical evidence regarding how consumers react to sponsorship disclosure is inconclusive. Sahni and Nair (2020) find that disclosing a restaurant's listing as a paid advertisement increases calls to the restaurant by 77%, while Hui and Liu (2022) find that more salient ad disclosure prompts consumers to substitute away from advertised listings. In Appendix F.1, I explore an extension where some consumers skip sponsored products when making purchases.

[17] This model loosely approximates a sequential search model in which consumers continue their search only if the expected incremental utility outweighs the associated search costs (e.g., see Lam 2023). Consumers with varying search costs may stop at different positions, resulting in distinct consideration sets. However, in my baseline model, the population shares of consumer types, $\lambda_n - \lambda_{n+1}$, are exogenous and do not adapt endogenously to changes in prices or product arrangement. In Appendix F.2, I explore an extension where consideration sets are determined endogenously in a sequential search model and can respond to changes in prices or product arrangement.

5.1.2 Estimation

In applying the model to the data, I introduce several additional assumptions. First, as the dataset consists of multiple searches for a keyword each day, I calculate the average market share across all searches on a given day. Second, while equation (4) defines a product's market share within a specific keyword, referred to as the *focal* keyword, I observe a product's aggregate quantity on the platform. I calibrate the fraction of a product's total sales attributed to the focal keyword in Appendix C.1. Additionally, in Appendix C.2, I calibrate the size of each market using industry reports. These calibrations allow me to construct a product's market share in the focal keyword.

The key primitives that I need to estimate are consumers' search frictions captured by λ_n and the parameter that measures consumers' price sensitivity α. The former determines the population shares of different consideration sets, and the latter plays a significant role in sellers' pricing decisions. I allow these parameters to vary across keywords. To estimate α, I employ a supply-side moment, as described in Section 5.2.2.

I parameterize the fraction of consumers who consider each position in the search results as $\lambda_n = 1/\exp\big(\beta(n-1)\big)$, where $\beta \geq 0$ is a parameter that determines the magnitude of search frictions. When $\beta = 0$, there are no search frictions, and a product's rank in search results does not affect its sales. A larger β implies that a greater share of consumers consider only products in top positions. In this case, all else being equal, ranking higher in search results would result in more sales. Appendix Figure A.11 provides examples of this function under different values of β.

Estimating the causal effect of a product's rank on its sales presents a challenge due to an endogeneity problem. The unobserved demand shock ξ_{jt} in equation (3) can be correlated with the product's organic rank r_{jt}, as a product's organic rank is likely influenced by its sales performance. A greater demand shock leads to more sales, which could be picked up by the platform's ranking algorithm and improve the product's organic rank on the same day.

To identify the parameter β, I employ a common approach found in the literature (e.g., Grennan 2013; Sweeting 2013; Errico and Lashkari 2022). First, I assume that the unobserved demand shock ξ_{jt} follows a first-order autoregressive (AR(1)) process:

$$\xi_{jt} = \rho\xi_{j,t-1} + \eta_{jt}, \;\; \eta_{jt} \sim \text{i.i.d.}, \;\; \mathbb{E}(\eta_{jt}) = 0, \;\; \eta_{jt} \perp \xi_{j,t-1}. \tag{5}$$

Under this assumption, ξ_{jt} consists of a persistent component that can be predicted by the lagged demand shock and a current innovation term η_{jt}.

Second, I assume that:

$$r_{j,t-1} \perp \eta_{jt}. \tag{6}$$

This assumption posits that when the platform determines a product's organic rank, it does not take future innovations in demand into account. This assumption is reasonable because, even if the platform anticipates future changes in a product's demand, it has no incentive to incorporate them into the product's current organic rank, as these changes are realized only in the future.

Last, I assume that, given the current demand shock, the current and lagged organic ranks have a nonzero correlation:

$$\text{Corr}(r_{j,t-1}, r_{jt} | \xi_{jt}) \neq 0. \tag{7}$$

This assumption is valid if the platform's ranking algorithm considers not only the current sales performance but also the sales performance over an extended period, thereby creating persistence in organic ranks.[18] Under this assumption, given the current demand shock ξ_{jt}, there exists variation in the lagged organic rank $r_{j,t-1}$ influenced by demand shocks before $t-1$. This variation carries forward to the current organic rank r_{jt} and affects a product's sales. Importantly, this variation is exogenous to the current innovation and can be used to identify the causal effect.

Under these assumptions, I estimate β and ρ for each keyword using the Method of Moments with the following conditions:

$$\mathbb{E}\begin{pmatrix} \eta_{jt}\xi_{j,t-1} \\ \eta_{jt}r_{j,t-1} \end{pmatrix} = 0. \tag{8}$$

To understand how these two parameters are identified, first consider a product with a constant rank over time. The fluctuations in its daily sales reveal the persistence of demand shocks and help identify ρ. Then, imagine two products with identical lagged demand shocks $\xi_{j,t-1}$. Their current demand shocks would follow the same distribution as given by equation (5). However, these two products may have followed different paths before $t-1$. For example, one product may have had stable demand over an extended period, while the other may have been out of stock for a time. These divergent histories result in variations in $r_{j,t-1}$ between the two products, which then carry over to r_{jt} and lead to different sales quantities. According to equation (6), these differences in $r_{j,t-1}$ are orthogonal to the current innovation and can help identify the parameter β.[19]

[18] Appendix Figure A.12 provides some suggestive evidence supporting this assumption.

[19] In Appendix G.4, I perform a robustness check by exploring various values of β. The range of β that I examine covers the value that aligns with the estimate by Ursu (2018), who uses experimental variation to identify the impact of ranks on consumer search, as well as the value corresponding to the estimate presented in Column 6 of Table 2.

5.2 Supply

5.2.1 Model

Sellers make bidding and pricing decisions every week to maximize their expected profits. In practice, a seller typically makes multiple keyword-specific bidding decisions and a single pricing decision that applies to the entire platform. However, estimating such a model can be challenging because of data and computational limitations. Therefore, I employ a partial-equilibrium approach, focusing on sellers' decisions within the focal keyword. Specifically, when seller j submits a bid b_j for this keyword and sets a price p_j, it assumes its bidding decisions in other keywords are fixed but acknowledges that changing its price can impact its sales across all keywords.

When sellers make bidding and pricing decisions, they consider the uncertainty associated with the realized search results on Amazon, which arises from two sources. First, organic rankings are determined in real time by Amazon's ranking algorithm, influenced by various factors and subject to variations across different searches. I assume that the actual order of organic rankings, denoted as Γ_0, is drawn from an exogenous distribution represented as $G(\cdot)$.

Second, auction outcomes can change frequently over time due to various real-world factors, such as budget constraints, personalized advertising, and experimentation, as discussed in Section 2. To model this uncertainty in auction outcomes, I assume that when a seller submits a bid b_j, the realized bid $\tilde{b}_j$ is stochastic and follows $\tilde{b}_j = \omega_j b_j$, where $\omega_j > 0$ reflects the deviation of the realized bid from the submitted bid. I refer to b_j as the *targeted* bid and $\tilde{b}_j$ as the *realized* bid. The vector $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_J)$ follows an exogenous distribution denoted as $F(\cdot)$.

Let $\boldsymbol{p} = (p_1, \ldots, p_J)$ and $\boldsymbol{b} = (b_1, \ldots, b_J)$ collect the prices and targeted bids of all sellers. The realized product arrangement in the search results, denoted as $\Gamma(\boldsymbol{b}, \boldsymbol{\omega}, \Gamma_0)$, is determined based on targeted bids, deviations in realized bids, and realized organic ranks. Specifically, the seller with the r-th highest realized bid occupies the r-th sponsored position, and the remaining positions are allocated based on realized organic ranks. Seller j's expected profits in a week are

$$\pi_j(\boldsymbol{p}, \boldsymbol{b}) = \left(s_j^e(\boldsymbol{p}, \boldsymbol{b}) + M_j^{other} s_j^{other}(p_j)\right)\left((1-\tau)\, p_j - c_j\right) - ad_j^e(\boldsymbol{p}, \boldsymbol{b}) - ad_j^{other}(p_j), \tag{9}$$

where

$$\begin{aligned} s_j^e(\boldsymbol{p}, \boldsymbol{b}) &= \int s_j\big(\boldsymbol{p}, \Gamma(\boldsymbol{b}, \boldsymbol{\omega}, \Gamma_0)\big) \mathrm{d}F(\boldsymbol{\omega}) \mathrm{d}G(\Gamma_0), \\ ad_j^e(\boldsymbol{p}, \boldsymbol{b}) &= \int ad_j(\boldsymbol{p}, \boldsymbol{b}, \boldsymbol{\omega}, \Gamma_0) \mathrm{d}F(\boldsymbol{\omega}) \mathrm{d}G(\Gamma_0). \end{aligned} \tag{10}$$

Here, τ is the commission rate set by the platform. c_j denotes the marginal cost, which includes production costs (or wholesale prices) and logistics expenses, such as transportation costs, but excludes advertising payment. $s_j^e(\boldsymbol{p}, \boldsymbol{b})$ represents the seller's expected market share within the focal keyword, and $s_j^{other}(p_j)$ denotes its aggregate market share in other keywords. M_j^{other} is the market size in other keywords relative to that in the focal keyword. $ad_j^e(\boldsymbol{p}, \boldsymbol{b})$ and $ad_j^{other}(p_j)$ represent the seller's expected advertising payment in the focal and other keywords, respectively.

Given the realized product arrangement $\Gamma(\boldsymbol{b}, \boldsymbol{\omega}, \Gamma_0)$, seller j's market share within the focal keyword, $s_j\big(\boldsymbol{p}, \Gamma(\boldsymbol{b}, \boldsymbol{\omega}, \Gamma_0)\big)$, is determined by equation (4). The same product may appear in both an organic and a sponsored position as separate listings, and I assume that they receive independent preference shocks. The expected market share, $s_j^e(\boldsymbol{p}, \boldsymbol{b})$, takes into account the uncertainty in auction outcomes and organic ranks in a week and is given by equation (10).

Regarding advertising payment, conditional on the seller's winning the r-th sponsored position, its advertising payment can be calculated as the payment per click (PPC) multiplied by the number of clicks on the sponsored listing. Let $j_r(\boldsymbol{b}, \boldsymbol{\omega})$ refer to the seller with the r-th highest realized bid. In a generalized second-price auction, the PPC for the seller ranked r-th is equal to the bid of the seller ranked $(r+1)$-th, or $\tilde{b}_{j_{r+1}(\boldsymbol{b},\boldsymbol{\omega})}$. Since I do not model a consumer's browsing process, I take a reduced-form approach to model clicks. Specifically, when product j appears in the r-th sponsored position, I denote the average number of clicks per sponsored unit, or the inverse conversion rate, as $\gamma_{jr} \geq 1$. A higher γ_{jr} implies that the product receives more clicks for each sponsored unit sold. Seller j's advertising payment can then be expressed as:

$$ad_j(\boldsymbol{p}, \boldsymbol{b}, \boldsymbol{\omega}, \Gamma_0) = \sum_{r=1}^{R} \mathbf{1}\big(j_r(\boldsymbol{b}, \boldsymbol{\omega}) = j\big) \tilde{b}_{j_{r+1}(\boldsymbol{b},\boldsymbol{\omega})} \gamma_{jr} s_j^S\big(\boldsymbol{p}, \Gamma(\boldsymbol{b}, \boldsymbol{\omega}, \Gamma_0)\big). \tag{11}$$

Here, $s_j^S\big(\boldsymbol{p}, \Gamma(\boldsymbol{b}, \boldsymbol{\omega}, \Gamma_0)\big)$ represents seller j's market share from a sponsored position, and $\mathbf{1}\big(j_r(\boldsymbol{b}, \boldsymbol{\omega}) = j\big)$ is an indicator of seller j winning the r-th sponsored position. The expected advertising payment, $ad_j^e(\boldsymbol{p}, \boldsymbol{b})$, is given by equation (10) and accounts for the uncertainty in search results.

In addition to the focal keyword, sellers may generate sales and incur advertising costs from other keywords. As previously discussed, I do not explicitly model the strategic interactions among sellers in those keywords. Instead, I use a logit model to approximate a seller's market share in other keywords as a function of its price, $s_j^{other}(p_j)$, and assume that the division between sponsored and organic sales mirrors the observed split in the focal keyword. Furthermore, I assume that a seller's advertising payment per sponsored unit in other keywords also mirrors what

is observed in the focal keyword. The relative market size, M_j^{other}, follows the same calibration in Appendix C.2. Appendix D.1 provides details on the construction of these quantities.

In equilibrium, sellers' expectations are rational. Given their expectations, each seller cannot increase its expected profits by submitting a different targeted bid or setting a different price.

5.2.2 Estimation

In this model, each seller makes two weekly decisions: the price p_j and the targeted bid b_j. There are two primitives to be estimated for each seller: the marginal cost c_j and the inverse conversion rate γ_{jr} (i.e., the average number of clicks per sponsored unit). I estimate these primitives by solving the first-order conditions for p_j and b_j in sellers' profit-maximization problems.

The expected profits in equation (9) rely on a seller's rational expectations regarding its market share and advertising payment. I approximate these expectations by calculating averages across a large number of simulated search results. To account for the uncertainty in auction outcomes and organic ranks, which are critical for simulating search results, I will first describe the estimation of the distributions of realized bids and organic ranks. Then, I will simulate search results using the estimated distributions and construct sellers' expected profits. Finally, I will discuss the first-order conditions and the estimation of the price sensitivity parameter.

Bids and Organic Ranks In the model, each seller sets a targeted bid, and sellers are ranked based on their stochastic realized bids. Different searches in a given week are independent realizations of the auction with the same set of targeted bids. I parameterize the realized bid as $\log(\tilde{b}_j) \overset{\text{i.i.d.}}{\sim} N\big(\log(b_j), \sigma_b^2\big)$, where σ_b measures the level of uncertainty in realized bids. I estimate the parameters $\boldsymbol{b}$ and σ_b in each market–week using a Gibbs sampler with data augmentation (Tanner and Wong 1987). The resulting parameters can generate auction outcomes that closely match the observed realizations and align with the aggregate statistics of the winning bids. A detailed description of the estimation procedure can be found in Appendix D.2.

I adopt a similar method to estimate the distribution of organic ranks. In each search, I assume that the platform assigns each product a quality score drawn from a product-specific distribution based on the information gathered by the platform. Products are then ranked based on their realized quality scores, and organic ranks may or may not align with the actual ranking of product utility. I parameterize the realized score as $\log(\tilde{o}_j) \overset{\text{i.i.d.}}{\sim} N(\log(o_j), \sigma_o^2)$ and use a Gibbs sampler to estimate $\boldsymbol{o} = (o_1, \ldots, o_J)$ in each market–week to match the actual realizations of organic ranks.

Expected Profits I simulate 2,000 realizations of organic ranks and realized bids for each market–week based on the estimated distributions. For a given seller j and a specific rank $r \in \{0, 1, \ldots, R\}$, I construct search results where seller j always occupies the r-th sponsored position. When $r = 0$, the seller does not win a sponsored position. To accomplish this, I hypothetically assign the r-th sponsored position to the seller in all realizations and allocate the remaining positions based on simulated organic ranks and the realized bids of other sellers. For each realization, I calculate the seller's market share using equation (4) and determine its advertising payment using equation (11). Then, I compute the average market share and advertising payment across simulations, denoted as $s_j^r(\boldsymbol{p}, \boldsymbol{b})$ and $ad_j^r(\boldsymbol{p}, \boldsymbol{b})$, respectively.

Last, I calculate seller j's probability of winning each sponsored position given other sellers' realized bids, denoted as $q_j^r(\boldsymbol{b})$. I approximate the expected market share and advertising payment in equation (9) as:

$$s_j^e(\boldsymbol{p}, \boldsymbol{b}) = \sum_{r=0}^{R} q_j^r(\boldsymbol{b}) s_j^r(\boldsymbol{p}, \boldsymbol{b}), \quad ad_j^e(\boldsymbol{p}, \boldsymbol{b}) = \sum_{r=1}^{R} q_j^r(\boldsymbol{b}) ad_j^r(\boldsymbol{p}, \boldsymbol{b}). \tag{12}$$

First-Order Conditions I parameterize the inverse conversion rate as $\gamma_{jr} = \exp(\gamma_j \lambda_r)$, where λ_r represents the fraction of consumers who consider the product in the r-th sponsored position, as defined in Section 5.1.1. γ_j captures product-specific conversion rates.[20] The number of clicks is at least one, as consumers who make a purchase always click on that product. I expect γ_j to be positive. A higher position in search results tends to attract more irrelevant traffic, as consumers in the exploration stage are more inclined to click on listings without completing a purchase.

Given a specific value of α, I estimate the seller primitives, c_j and γ_j, by solving two first-order conditions. Intuitively, a seller's pricing decision reveals its margin and, consequently, the sum of its marginal cost and advertising cost. With a fixed margin, a higher bid corresponds to a higher conversion rate, and vice versa. Therefore, the seller's bidding decision helps distinguish between the two types of costs and identifies both parameters. For sellers who never appear in a sponsored position in a given week, I assume that they do not advertise and estimate only c_j for them.

Price Sensitivity I estimate the price sensitivity parameter α in equation (3) using a supply-side moment. Specifically, I leverage data from a survey conducted by Jungle Scout in 2022.[21] This

[20] The parameter γ_j could also encompass sellers' dynamic incentives. For example, a seller can invest in sponsored product advertising to enhance its organic ranks over time, although this possibility is beyond the scope of this paper.

[21] See the report *State of the Amazon Seller 2022* (Jungle Scout 2023).

survey asked approximately 3,500 Amazon sellers of various sizes about their performance on the platform and found a median profit margin of 18% across all sellers. With a specific value of α, I estimate seller primitives and compute each seller's profit margin. Then, I search for the value of α in each market such that the median profit margin matches 18%.[22]

5.3 Platform

5.3.1 Model

The platform selects a commission rate τ to maximize a combination of short-term revenues, consumer surplus, and seller profits. In line with previous studies on multisided platforms (e.g., Gutierrez 2022; Castillo 2022), the platform can invest in user acquisition and retention to maximize its long-run shareholder value. Incorporating the welfare of consumers and sellers into the platform's objective serves to approximate these long-term investment considerations that are not explicitly modeled.

As discussed in Section 2, Amazon also generates revenues from sources beyond commissions, such as storage and shipping fees from sellers and Prime membership fees from consumers. These revenue streams share a common feature with commissions—they serve as substitutes for advertising revenues. The stylized example in Section 3 illustrates how commissions can act as a substitute for advertising revenues, and the same intuition applies to storage and shipping fees: an increase in these fees diminishes sellers' margins and, consequently, the platform's advertising revenues. In the case of Prime membership, a price increase may deter some consumers from subscribing, potentially leading them to explore alternatives. This shift can make top positions in search results less appealing to sellers. The concept of substitution among different revenue instruments is a critical aspect captured in my model. Hence, the commission rate represents just one element within a broader array of revenue-generating tools available to the platform.

Let $CS(\tau)$ and $PS(\tau)$ represent expected consumer surplus and seller profits under the commission rate τ. The platform's short-term revenues encompass commissions $COM(\tau)$ and advertising revenues $AD(\tau)$. Appendix E.1 fleshes out these terms. Given the configuration of sponsored positions in search results, the platform selects τ to maximize the following objective:

$$\max_{\tau} \ \Pi(\tau) = COM(\tau) + AD(\tau) + \mu\big(CS(\tau) + PS(\tau)\big). \tag{13}$$

[22] One caveat is that the number 18% represents the median profit margin across all sellers on the platform and may not necessarily apply to each market. Unfortunately, the survey does not provide category-specific profit margins.

Here, $\mu \geq 0$ is a nonnegative weight, reflecting the platform's consideration of the welfare of other market participants relative to its short-term revenues.

In the model, I do not assume that the platform optimally selects the set of sponsored positions. The number of potential combinations of sponsored positions is vast,[23] making it implausible for the observed choice to be entirely optimal. In reality, the arrangement of sponsored positions on Amazon has been constantly evolving in recent years, likely due to its ongoing experimentation.[24]

5.3.2 Estimation

I estimate the platform's weight on other market participants, μ, using its optimality condition in equation (13). While the model assumes a uniform commission rate, in practice, this rate varies from 8% to 20% across different markets. Let $\tau^{m,0}$ denote the observed commission rate in market m. In most markets within my sample, $\tau^{m,0}$ takes values of either 15% or 17%. The size-weighted average commission rate across these markets, denoted as τ^0, is 15.6%. I assume that the platform chooses an average commission rate, which is then proportionally translated into market-specific commission rates. Specifically, when the platform sets an average commission rate τ, the actual commission rate in market m is calculated as $\tau^m = \frac{\tau^{m,0}}{\tau^0}\tau$.

I maintain the configuration of sponsored positions and vary the commission rate τ. For each market–week, I solve for the new equilibrium under various values of τ and calculate the corresponding consumer surplus, seller profits, commissions, and advertising revenues in the new equilibrium. To identify the weight parameter μ, I take the derivative of the objective function in equation (13) with respect to τ and set it equal to zero at $\tau^0 = 15.6\%$.

6 Results

6.1 Parameter Estimates

Figure 3 presents key parameter estimates. Panel A illustrates the fraction of consumers considering each position in search results. The various lines represent markets with different degrees of search frictions.[25] In the absence of search frictions, consumers would consider all products,

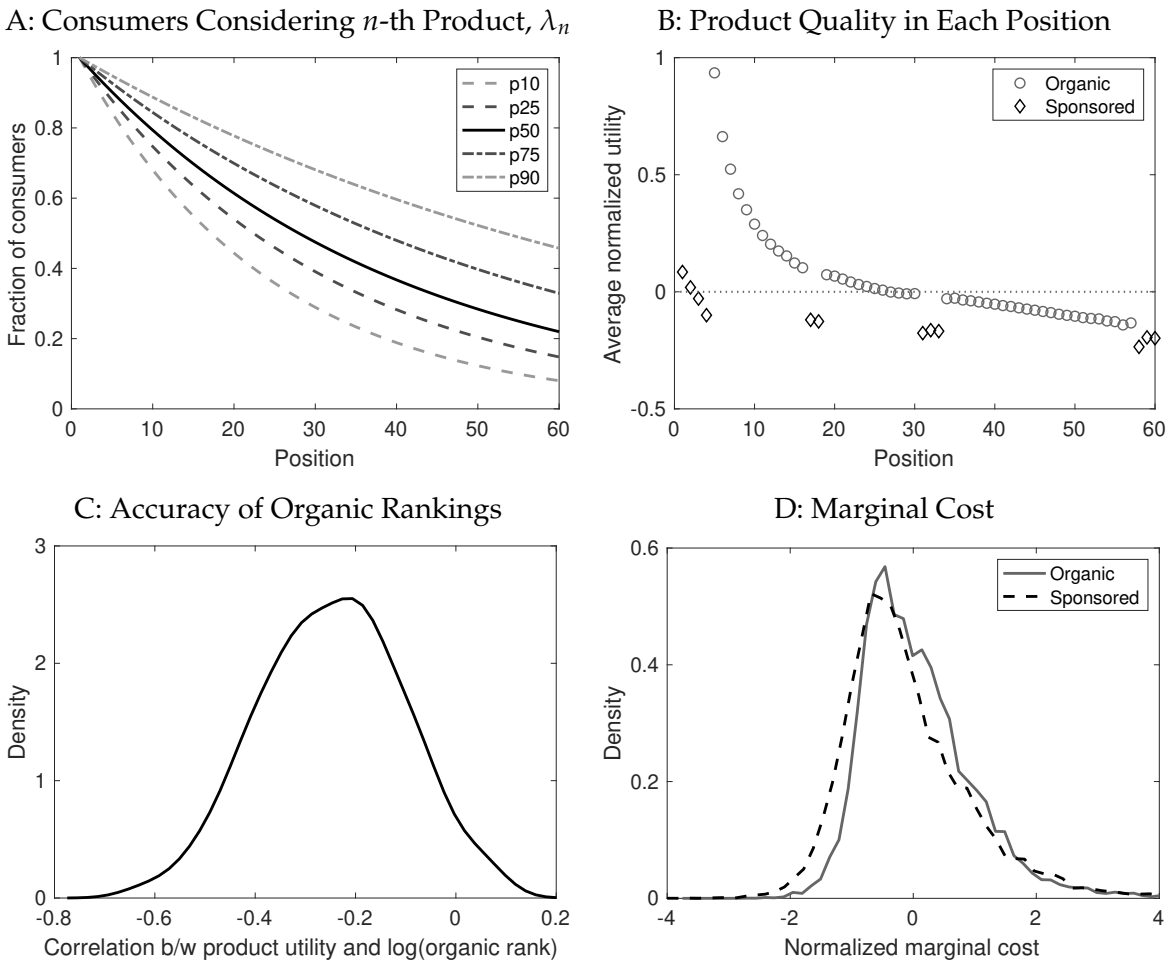
[23] With 60 positions in search results, there are 2^{60} possible sets of sponsored positions, leading to approximately 10^{18}.

[24] As demonstrated in Appendix E.2, there exist deviations in the set of sponsored positions that increase consumer surplus, seller profits, and platform revenues simultaneously and thus benefit the platform regardless of its objective.

[25] To account for sampling errors, I utilize an empirical Bayes procedure to shrink the market-specific parameter $\hat{\beta}$ toward its mean across markets. The details of this procedure are provided in Appendix C.4.

resulting in a flat line in the figure. However, the observed downward-sloping lines indicate a gradual decline in consumer consideration from the top position to the bottom position, which supports the presence of search frictions and is consistent with the finding in Section 4.2.

Figure 3: Estimation Results



Notes: This figure presents key estimation results. In Panel A, I calculate the fraction of consumers whose consideration sets contain the product in the n-th position in the search results, $\lambda_n = 1/\exp(\beta(n-1))$, based on the estimated parameter $\hat{\beta}$ in each market. Different lines correspond to different percentiles of λ_n across markets. In Panel B, I normalize the estimated product utility $\hat{\delta}_{jt}$, as defined in Section 5.1.1, to have zero mean and unit variance across all products in a given market–week. The figure displays the average normalized utility for products that appear in a given position across all searches, separately for sponsored and organic positions. Panel C displays the distribution of the correlation between estimated product utility $\hat{\delta}_{jt}$ and the log of organic ranks in each market, a measure for the accuracy of organic rankings. Panel D displays the distribution of marginal costs, separately for the top four sponsored and four organic products in all searches.

In the median market, approximately 47% of consumers consider a product in the middle of the search results, while 22% consider the last product.[26] A median consumer has 28 products in

[26]My analysis focuses on the first page of search results. Products on subsequent pages are considered part of the

her consideration sets, corresponding to seven rows of products on a desktop or laptop browser.[27] Typically, consumers scroll down the screen multiple times, skip over most products based on the information in the listings, and click on only several listings to inspect their product pages in detail. Therefore, the number 28 should be interpreted as the upper limit on the page position that a consumer may consider rather than the number of products that she carefully examines.

Panel B of Figure 3 compares product utility across different positions in search results. In estimating product utility δ_{jt} based on consumers' revealed preferences, it is crucial to properly account for the position effect. Even when products have the same utility, a product located higher in search results is considered by more consumers and consequently captures a larger market share. A model that assumes a homogeneous consideration set would overestimate the utility of products located in higher positions. The demand model in Section 5.1.1 explicitly incorporates consumers' heterogeneous consideration sets, thereby eliminating the mechanical correlation between positions and utility.

To aggregate the comparison across market–weeks, in this figure, I normalize the estimated utility $\hat{\delta}_{jt}$ to have a zero mean and unit variance across all products within a specific market–week. The figure then illustrates the average normalized utility of products in each position. This comparison reveals two notable patterns. First, in organic positions, there exists a clear correlation between higher rankings and higher average utility, and this relationship is generally monotonic. The average utility declines rapidly in the initial few positions. For instance, the average utility for products in the first organic position is 0.94 standard deviations higher than the average utility of all products. However, this difference diminishes to 0.35 standard deviations for products in the fifth organic position.

Second, sponsored products exhibit significantly lower average utility than their organic counterparts in similar positions. For instance, products in the first sponsored position, which is also the first position on the entire page, have slightly higher utility than an average product but fall significantly short of the average utility of products in the first organic position. Sponsored products located beyond the third sponsored positions consistently exhibit lower average utility than an average product. These observations indicate that, on average, organic ranks on Amazon ef-

outside option. Some of the 22% may represent consumers who consider products beyond the first page. According to Amazon's data, 81% of clicks occur on the first page of search results. See the article published by *Search Engine Journal*, "Amazon's Search Engine Ranking Algorithm: What Marketers Need to Know" (accessed July 2023).

[27] Ursu (2018) utilizes an experiment on Expedia to examine the impact of ranks on consumer search and finds that a median consumer considers approximately 34 products. Lam (2023) adopts an identification strategy similar to that in my paper and reports that a median consumer sees 15–20 products in the "Home & Kitchen" category on Amazon.

fectively reflect product utility, while sponsored products tend to exhibit lower utility.

In Panel C of Figure 3, I calculate the correlation between product utility and the log of organic ranks for each market. This measure helps me assess the accuracy of organic rankings. The figure shows a median correlation of -0.25, which suggests that while organic rankings, on average, reasonably reflect product utility, they are far from perfect. Notably, there is considerable variation across markets, with the correlations spanning from -0.65 to 0.05. I further examine this heterogeneity in Section 7.3.

Panel D of Figure 3 illustrates the distribution of normalized estimated marginal costs for the top eight products in search results, separately for sponsored and organic products. Top sponsored products exhibit significantly lower average marginal costs. However, this cost advantage does not translate into lower prices for consumers. In fact, as shown in Panel D of Figure 2, prices for top sponsored products are 16% higher than those for top organic products. This price difference is primarily attributed to the additional advertising payment required for securing sponsored placements.

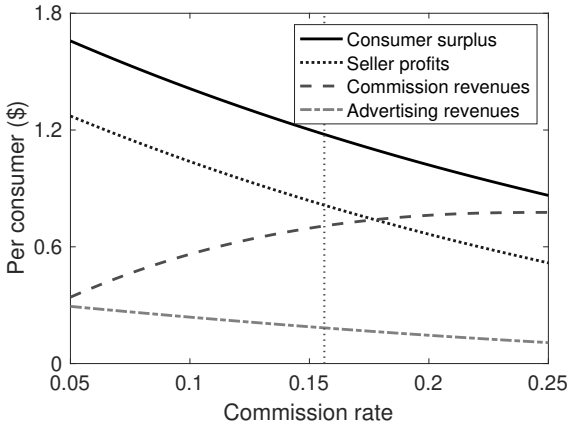
In Figure 4, I present the estimation of the platform's weight on the welfare of other market participants. As the platform increases its commission rate, consumer surplus and seller profits decrease, while commission revenues increase in a concave manner. Notably, advertising revenues decline as the commission rate rises. This relationship is important: higher commission rates reduce sellers' willingness to pay for sponsored positions and lower advertising revenues for the platform. Using these functions, I determine the parameter μ to be 0.116, which satisfies the platform's first-order condition.[28]

6.2 Model Fit

In Figure 5, I validate the model estimates by comparing them with industry reports. Panel A focuses on three metrics commonly used in the advertising industry: (i) the click-through rate, calculated as total clicks divided by total impressions; (ii) the conversion rate, calculated as total sponsored units sold divided by total clicks; and (iii) the advertising cost of sales, calculated as the total advertising payment divided by total sponsored sales. I compute these metrics for each market based on my model estimates and compare them to industry benchmarks. Remarkably, the

[28]This value is relatively low compared to estimates in the literature. For instance, Castillo (2022) finds that Uber acted as if it were maximizing rider surplus in 2017, and Gutierrez (2022) finds that Amazon put a median weight of 1.11 on consumer welfare and 0.39 on seller welfare in 2018–2020. The settings are different, and the weights likely decline as a platform becomes more established.

Figure 4: Outcomes Under Varying Commission Rates



Notes: This figure displays consumer surplus, seller profits, commission revenues, and advertising revenues under different commission rates. I hold the set of sponsored positions at the observed configuration and vary the commission rate. Under each commission rate, I solve for the new equilibrium in each market–week, compute the outcome measures, and aggregate across market–weeks using market sizes.

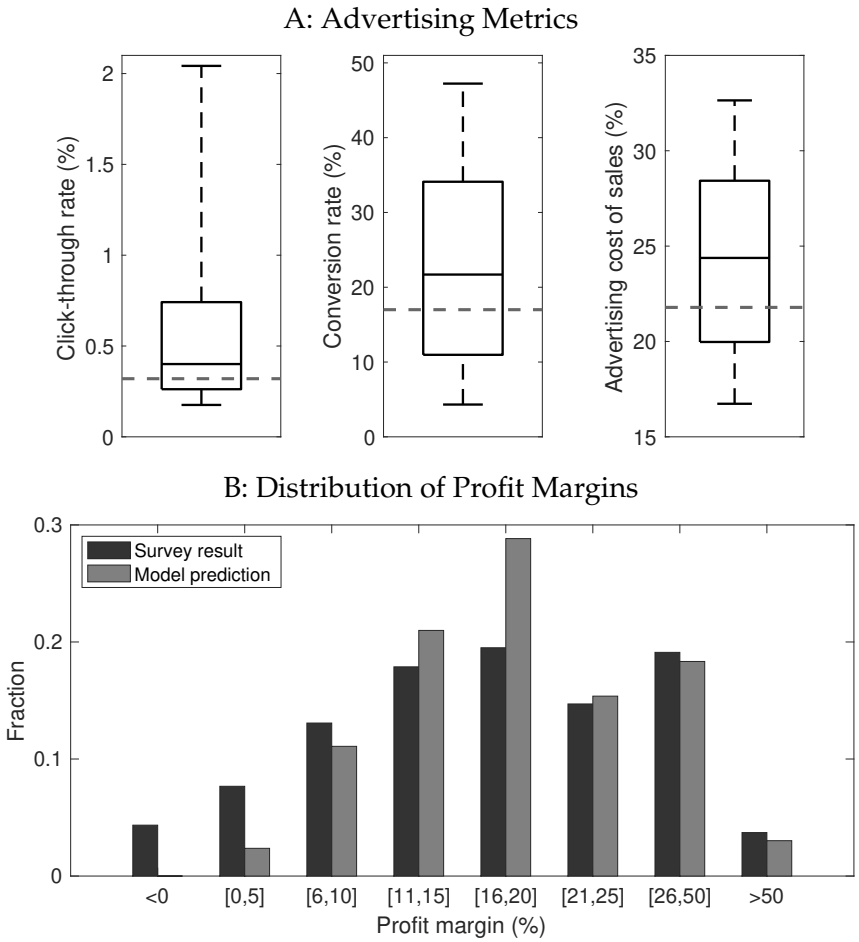
model's predictions closely align with the industry estimates, despite the estimation process not directly targeting these metrics. For instance, in a median market, the model suggests an average click-through rate of 0.40%, which is close to the industry estimate of 0.32%.[29]

The Jungle Scout survey provides data on the fractions of sellers with profit margins falling into eight different intervals.[30] Using the estimates from my model, I calculate the profit margin of each seller and then compute the corresponding fractions. Panel B of Figure 5 compares these predicted fractions from the model with those reported in the survey. In general, the model's predictions align reasonably well with the survey results. For example, the survey identifies 17.9% of sellers with profit margins between 11% and 15%, while the model predicts a similar number of 21.0%. However, the survey reports that 4.4% of sellers have negative profits, which the model precludes by construction. These sellers might have dynamic incentives, such as setting low prices to invest in improving their organic ranks. These incentives could be reflected in the estimates of their marginal costs, which may be lower than their actual costs.

[29] According to Pacvue's estimates (accessed December 2022), in the second quarter of 2022, the average click-through rate on Amazon was 0.32%, conversion rate was 17.0%, and advertising cost of sales was 21.8%.

[30] These intervals are $<$0%, 0–5%, 6–10%, 11–15%, 16–20%, 21–25%, 26–50%, and $>$50%.

Figure 5: Model Validation



Notes: This figure serves as a cross-validation of model estimates by comparing them to industry data. Panel A presents a comparison of three key metrics commonly used in the search advertising domain, as derived from model estimates and industry reports. These metrics include (i) click-through rate (CTR), calculated as the total number of clicks divided by the total number of impressions; (ii) conversion rate (CVR), calculated as the total number of sponsored units sold divided by the total number of clicks; and (iii) advertising cost of sales (ACoS), calculated as the total advertising payments divided by the total sponsored sales. Each box in Panel A represents the distribution of one of these metrics across markets, with the middle line indicating the median, the edges of a box indicating the 25th and 75th percentiles, and the outer lines indicating the 5th and 95th percentiles. The dashed line represents the industry estimate from Pacvue in the second quarter of 2022 (accessed December 2022). Panel B compares the distribution of profit margins for sellers, as inferred from model estimates, with data collected in a Jungle Scout survey. Jungle Scout conducted this survey in 2022, collecting information on the proportion of sellers whose profit margins fell within eight different intervals. I compute the profit margin of each seller in my sample using the model's estimates and calculate the corresponding fractions. The figure compares these predicted fractions with those reported in the survey.

7 Welfare Effects of Sponsored Product Advertising

In this section, I quantify the welfare effects of sponsored product advertising on Amazon. I start by evaluating the overall effects on consumers, sellers, and the platform and then break down the

change in consumer surplus into different components. Next, I examine the heterogeneity in the welfare effects across markets. Last, I explore alternative sets of sponsored positions.

7.1 Overall Effects

I examine a counterfactual scenario in which all sponsored product slots are removed and replaced with organic listings. To simulate this scenario, I generate 2,000 realizations of organic ranks and populate the search results with the top $N = 60$ products. I then find the new equilibria in each market under varying commission rates and aggregate across markets. I consider two cases that differ in how the platform responds to the removal of sponsored positions: one where the platform maintains a fixed commission rate of 15.6% and another where the platform adjusts its commission rate to optimize its objective. By comparing this counterfactual scenario to the status quo, I can assess the welfare effects of *removing* sponsored product advertising.

Table 3 summarizes the welfare effects of removing sponsored product advertising. The first three columns report consumer surplus, seller profits, commission revenues, and advertising revenues under both the status quo and the counterfactual scenario with a fixed commission rate. The elimination of sponsored positions results in a 12.6% increase in consumer surplus and a 14.1% rise in seller profits. While there is a moderate uptick in commission revenues because of increased sales, the platform experiences a substantial 14.8% decline in total revenues due to the loss of all advertising revenues. The total surplus, which encompasses consumer surplus, seller profits, and platform revenues, increases by 4.6%. These findings suggest that when the commission rate remains fixed, the presence of sponsored product advertising on Amazon has an overall negative impact on consumers and results in a significant transfer from sellers to the platform.

The finding that the removal of sponsored positions leads to an increase in consumer surplus is consistent with the estimation results. As illustrated in Panel B of Figure 3, organic ranks on Amazon serve as reliable predictors of product utility. In contrast, sponsored products consistently exhibit lower utility than their organic counterparts when occupying similar positions. Despite sponsored products having lower marginal costs, the additional advertising payment effectively negates their cost advantages. Furthermore, the overlap between sponsored and organic products, as depicted in Appendix Figure A.10, exacerbates the loss of consumer surplus.[31]

[31] In the baseline analysis, I assume that the second appearance of the same product in search results does not receive an independent idiosyncratic preference shock when I calculate consumer surplus. Consequently, any duplication of product listings would be a waste of space. In Appendix G.6, I explore an alternative assumption where the second appearance also receives its independent preference shock; the welfare results remain qualitatively the same.

Table 3: Welfare Effects of Removing Sponsored Product Advertising

	(1)	(2)	(3)	(4)	(5)
	Status Quo	Counterfactual: Without Sponsored Product Advertising			
		Fixed Commission		Endogenous Commission	
	Level ($)	Level ($)	Change (%)	Level ($)	Change (%)
Consumer Surplus	1.18	1.33	12.6%	1.13	-4.3%
Seller Profits	0.81	0.93	14.1%	0.75	-8.4%
Platform Revenues	0.89	0.76	-14.8%	0.83	-7.2%
Commission Revenues	0.71	0.76	7.4%	0.83	16.9%
Advertising Revenues	0.18	0.00	-100.0%	0.00	-100.0%
Total Surplus	2.88	3.02	4.6%	2.70	-6.3%

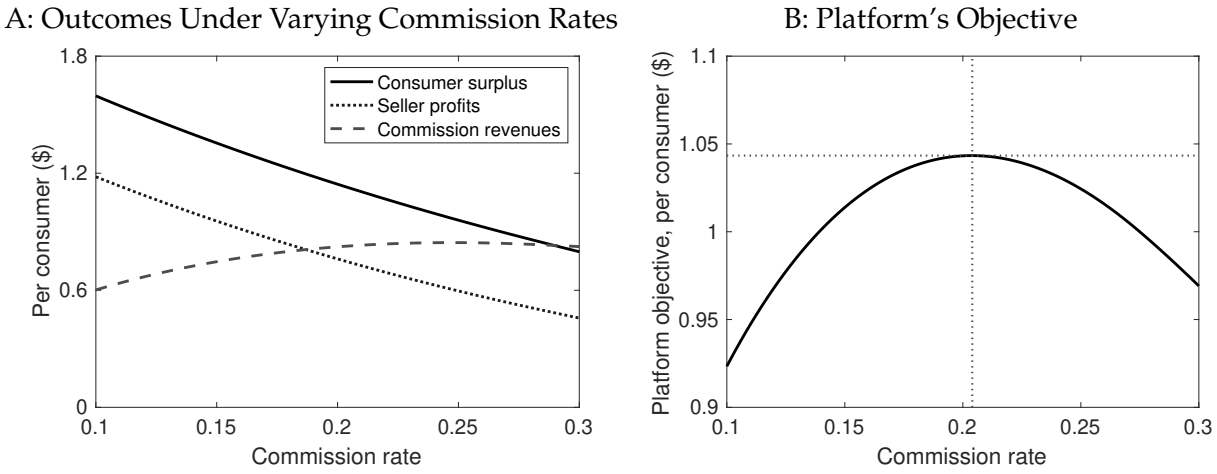
Notes: This table compares outcome measures between two scenarios: the status quo with sponsored positions (column 1) and a counterfactual scenario where all sponsored positions are replaced with organic ones (columns 2-5). In columns 2 and 3, the commission rate remains fixed at the observed rate from the status quo. In columns 4 and 5, the platform re-optimizes the commission rate to maximize its objective. Columns 1, 2, and 4 present the outcome measures in dollars per consumer, while columns 3 and 5 calculate the percentage changes from the status quo to the respective counterfactual scenarios. The results are aggregated across market–weeks based on market sizes. Figure 6 illustrates the optimal commission rate in the counterfactual scenario.

Now turn to the scenario where the platform has the flexibility to adjust its commission rate following the removal of sponsored positions, Figure 6 unveils an optimal commission rate of 20.3%, a significant increase over the current rate of 15.6%. This adjustment arises because, in the presence of sponsored product advertising, higher commission rates result in lower advertising revenues, which incentivizes the platform to maintain a relatively low commission rate. The model thus predicts an increase in the commission rate when this incentive is removed.

The last two columns of Table 3 illustrate the welfare effects with the reoptimized commission rate. Compared to their counterparts under the status quo, consumer surplus decreases by 4.3%, and seller profits decline by 8.4%. While the additional commissions collected can partially offset the loss in advertising revenues, total platform revenues still decrease by 7.2%. Overall, the total surplus drops by 6.3%, primarily driven by the decline in total transactions on the platform.

The comparison between the two scenarios, one with a fixed commission rate and the other with endogenous commission rates, underscores the importance of choosing the right counterfactual scenario in assessing the impact of sponsored product advertising. Simply comparing the status quo to a scenario with no sponsored product advertising with everything else held constant may not accurately reflect Amazon's response. If sponsored product advertising were prohibited, Amazon would have a strong economic incentive to increase other fees. Therefore, it is crucial to compare the impact of different revenue-raising methods. My model compares the introduction of sponsored positions and the raising of the commission rate, and the results indicate that the for-

Figure 6: Optimal Commission Rate Without Sponsored Product Advertising



Notes: This figure identifies the optimal commission rate in the counterfactual scenario where all sponsored positions are replaced with organic ones. Panel A plots consumer surplus, seller profits, and commission revenues under different commission rates. The results are aggregated across market–weeks using market sizes. Panel B constructs the platform's objective, a linear combination of revenues, consumer surplus, and seller profits, using the weight estimated in Section 6.1. The vertical dotted line indicates the commission rate that maximizes the platform's objective.

mer not only generates more revenues for the platform but also exerts a smaller negative impact on consumers and sellers. In reality, while Amazon might not frequently modify its commission rate, it does adjust other fees, such as fulfillment fees or its Prime membership price, on an annual basis. As discussed in Section 5.3.1, the commission rate in my model represents a broader set of instruments available to the platform.

How should we interpret this result? In theory, a platform could maximize its revenues through tailored product placement and individualized seller fees. However, this approach requires extensive information and is impractical in real-world scenarios. To allocate positions in search results, the platform generally has two primary options: to rely on existing information with sellers paying a uniform fee or to conduct auctions for a subset of positions with winners paying their bids. The auction mechanism elicits private information from sellers and allows the platform to engage in price discrimination, often resulting in higher platform revenues. My finding suggests that this enhanced efficiency can also benefit other market participants.

7.2 Breaking Down the Change in Consumer Surplus

Section 7.1 evaluates the overall impact of removing sponsored product advertising. To gain a deeper understanding of the various factors at play, I break down the overall change in consumer

surplus into distinct components. Specifically, I analyze three intermediate cases in addition to the status quo and the counterfactual scenario, as defined below.

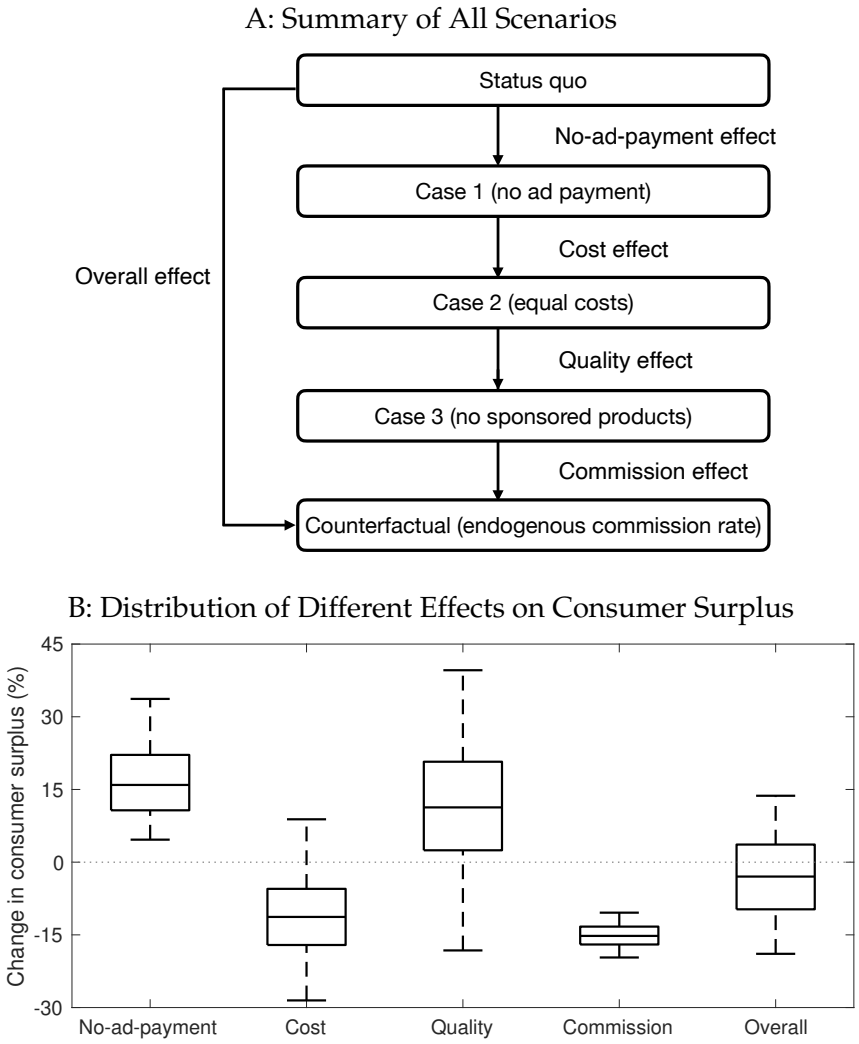
- **Status quo**: The commission rate is 15.6%. Sponsored products pay advertising payment.
- **Case 1 (no ad payment)**: The commission rate is 15.6%. The product arrangement is the same as in the status quo, but sponsored products do not pay advertising payment.
- **Case 2 (equal costs)**: I hypothetically adjust the marginal costs of sponsored products by a market-specific constant to equalize the average marginal cost for sponsored and organic products within each market. All other conditions remain the same as in Case 1.
- **Case 3 (no sponsored products)**: The commission rate is 15.6%. All positions show organic results.
- **Counterfactual (endogenous commission rate)**: The commission rate is 20.3%. All positions show organic results.

I refer to the percentage change in consumer surplus from the status quo to Case 1 as the *no-ad-payment effect*, which signifies the impact of eliminating the pass-through from the advertising payment to prices. Likewise, I define the change from Case 1 to Case 2 as the *cost effect*, reflecting the impact of cost disparities between sponsored and organic products. Transitioning from Case 2 to Case 3 represents a shift in the set of products available to consumers. Given that any cost differences have already been removed, the change in consumer surplus can be understood as the *quality effect*. Last, the change from Case 3 to the counterfactual scenario is labeled the *commission effect*, quantifying the impact of a higher commission rate. Panel A of Figure 7 summarizes the definitions of these effects.

Panel B of Figure 7 illustrates the distribution of these four effects across markets.[32] The no-ad-payment effect is consistently positive, as the removal of advertising payment always lowers prices and benefits consumers. In contrast, the commission effect stands out as highly negative, indicating a significant distortion caused by higher commission rates. Meanwhile, the cost effect tends to be negative, aligning with the finding in Panel C of Figure 3 that sponsored products often have lower marginal costs than organic products.

[32] Appendix Figure A.13 provides a similar breakdown for seller profits, platform revenues, and total surplus. Appendix Figure A.14 demonstrates that alternative sequences for the decomposition exercise yield similar results.

Figure 7: Decomposition of Welfare Effects of Removing Sponsored Product Advertising



Notes: This figure decomposes the aggregate effect of removing sponsored product advertising on consumer surplus. Besides the status quo and counterfactual scenario, I consider three intermediate cases. The first case assumes the same commission rate and product arrangement as the status quo, but sponsored products are not required to pay advertising payments. The second case hypothetically shifts the marginal costs of sponsored products so that the average marginal cost is the same for sponsored and organic products in each market–week. Everything else remains the same as in the first case. The third case assumes an unchanged commission rate but with all positions displaying organic products. The percentage change in consumer surplus from the status quo to the first case is defined as the *no-ad-payment effect*, from the first to the second case defined as the *cost effect*, from the second to the third case defined as the *quality effect*, and from the third case to the counterfactual scenario with an endogenous commission rate defined as the *commission effect*. Panel A summarizes the definitions of different effects. In Panel B, each box presents the distribution of one of these four effects and the overall effect across market–weeks, with the middle line indicating the median, the edges of a box indicating the 25th and 75th percentiles, and the outer lines indicating the 5th and 95th percentiles. Appendix Figure A.13 plots the decomposition of the aggregate changes in seller profits, platform revenues, and total surplus.

Notably, the quality effect is generally positive but exhibits substantial variation across markets. In approximately 20% of these markets, this effect turns negative. This suggests that, in some

cases, transitioning from the existing mix of products to exclusively organic results may reduce the average quality of products available to consumers. When all these effects are combined, they result in a negative overall effect in the majority of markets. However, in 38% of the markets, the overall effect is positive. In the following section, I will explore this heterogeneity in more detail.

7.3 Heterogeneity Across Markets

Section 7.2 reveals substantial variation across markets in the effect of removing sponsored product advertising on consumers. This section explores the factors contributing to this heterogeneity.

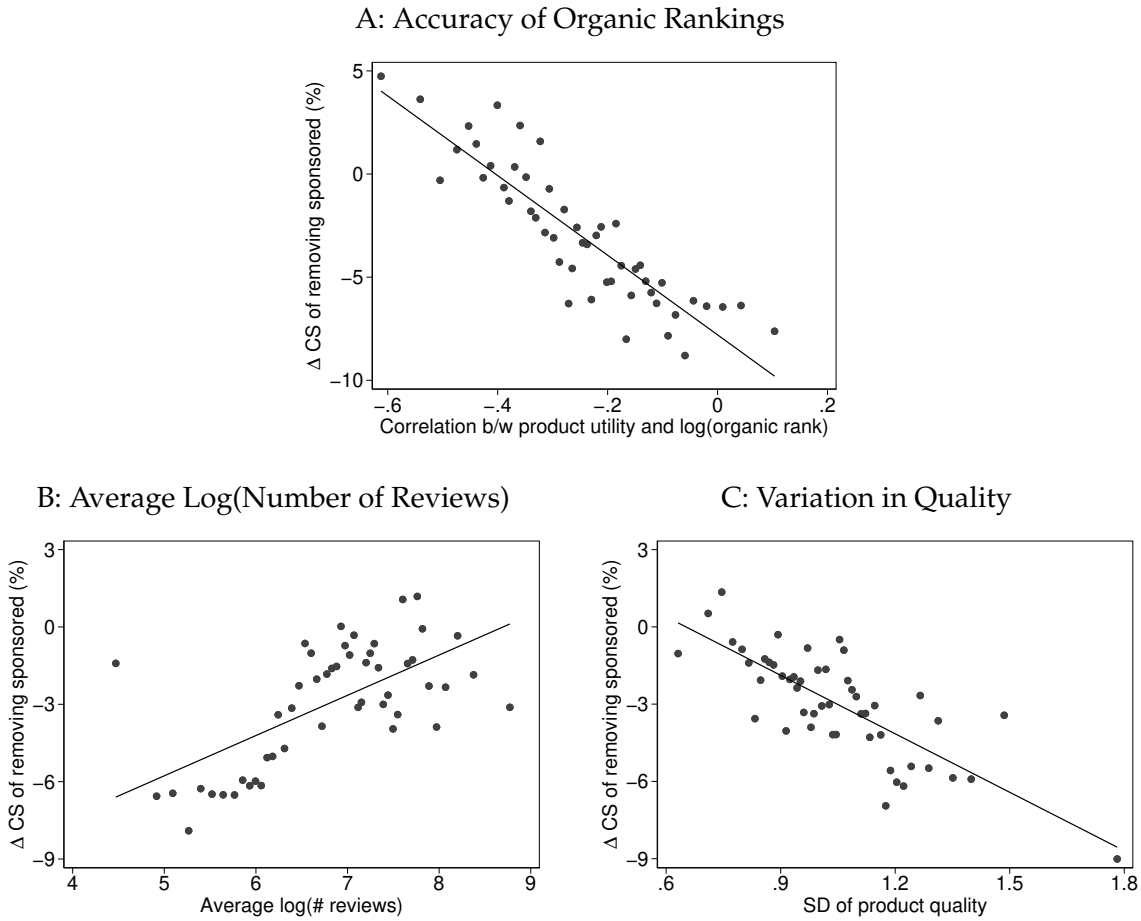
As highlighted in Section 3, one primary advantage of sponsored product advertising is enabling sellers to reveal information about their products when organic rankings fail to accurately reflect product utility. Panel A of Figure 8 confirms this intuition, showing a positive correlation between the impact of eliminating sponsored product advertising on consumers and the accuracy of organic rankings within a given market. This accuracy is measured as the correlation between product utility and organic ranks, as illustrated in Panel C of Figure 3. In markets where organic rankings are less accurate, removing sponsored products tends to be more harmful to consumers.

I further examine market-level characteristics that could predict the accuracy of organic rankings within a market. A conceivable factor is that organic rankings may be less accurate in markets with a substantial number of young products that lack established sales records on the platform. Panel B of Figure 8 shows a positive correlation between the welfare impact and the log of total reviews averaged across all products in a given market. When the average number of reviews is low, the platform may lack the necessary information to construct accurate organic rankings, and sponsored product advertising can bridge this information gap. Consequently, removing sponsored product advertising in such situations may have a more detrimental effect on consumers.

Organic rankings may also be less accurate in markets with significant quality differentiation. Conversely, markets with relatively homogeneous or standardized products likely have limited information not reflected in their organic rankings. To investigate this hypothesis, I compute the standard deviation of estimated product utility $\hat{\delta}_{jt}$ across all products within a given market, which serves as a proxy for quality differentiation. Panel C of Figure 8 demonstrates a negative correlation between the impact of removing sponsored product advertising and the degree of quality differentiation, providing support for this hypothesis.

This analysis emphasizes that the impact of sponsored product advertising on consumer surplus varies with market conditions. Sponsored product advertising tends to generate greater gains

Figure 8: Heterogeneity in the Welfare Effect on Consumer Surplus Across Markets



Notes: This figure examines the heterogeneity in the effect on consumer surplus across markets. The y-axis represents the percentage change in consumer surplus from the status quo with sponsored positions to the counterfactual scenario without sponsored positions under endogenous commission rates, which corresponds to the "overall effect" plotted in Figure 7. In Panel A, the x-axis represents the correlation between estimated product utility and the log of organic rank in a market, a measure of the accuracy of organic rankings. A more negative correlation indicates more accurate organic rankings. In Panel B, the x-axis represents the log of the number of reviews averaged across all products in a market. In Panel C, the x-axis represents the standard deviation of estimated product utility $\hat{\delta}_{jt}$ across all products in a market, a measure of product differentiation. All panels show binned scatter plots using 50 bins.

in markets characterized by fewer established products and higher levels of quality differentiation.

7.4 Alternative Numbers of Sponsored Positions

For policymakers concerned about the rise in sponsored product advertising, a complete removal of sponsored products may not be a practical or desirable policy choice. Instead, a more pragmatic policy target might involve an intermediate approach. In this final analysis, I examine the effects of changing the number of sponsored positions in search results. Starting from the baseline sce-

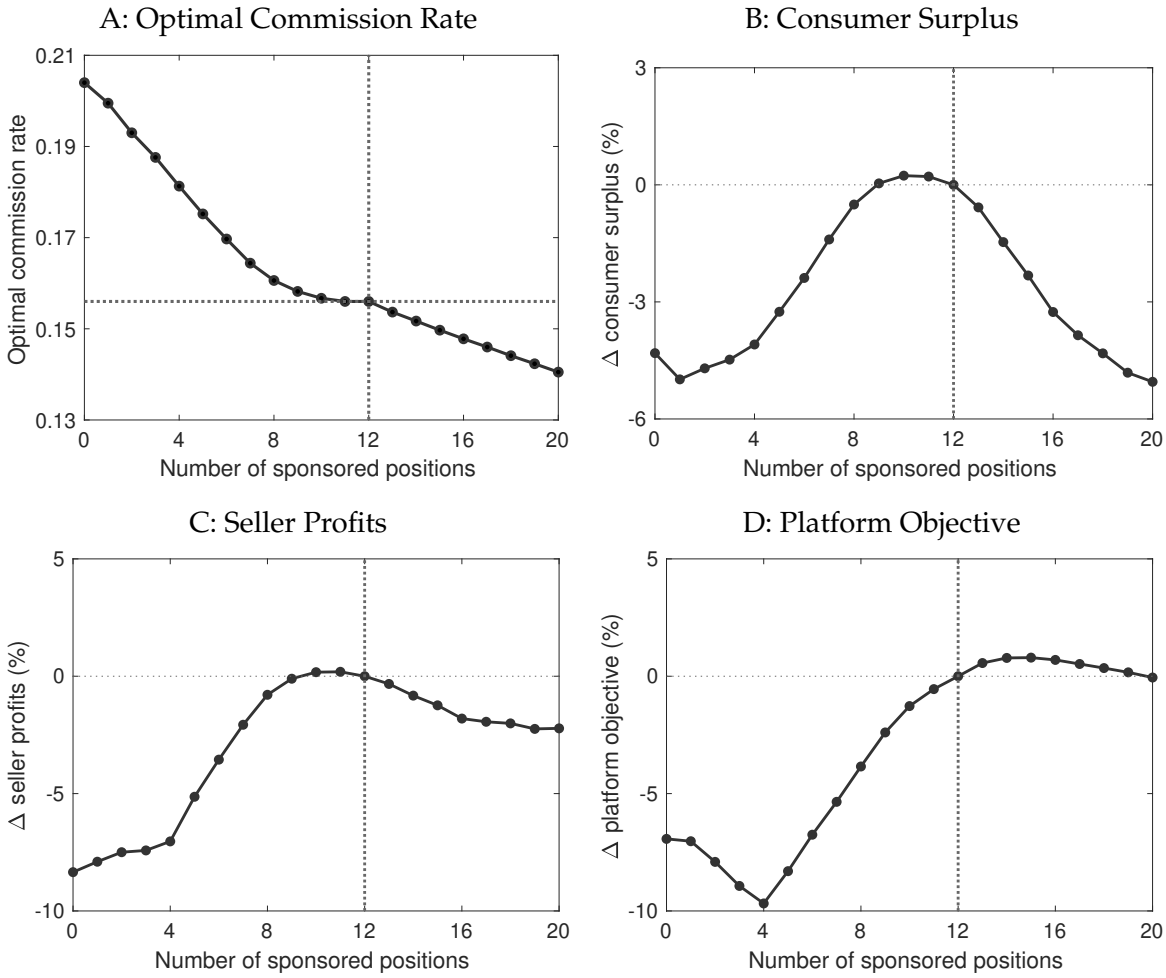
nario with 12 sponsored positions, I explore two directions. First, I gradually replace sponsored positions with organic ones, starting from the lowest sponsored position and proceeding to the highest. This approach represents an intermediate step between maintaining the status quo and eliminating all sponsored positions. Second, I incrementally introduce more sponsored positions in the middle of search results, starting by substituting the 5th organic position with a sponsored one and continuing down to the 12th organic position. This exercise allows me to speculate on Amazon's potential strategies in the future. Panel A of Figure 9 shows that as the number of sponsored positions increases, the optimal commission rate consistently decreases.

Panel B of Figure 9 reveals a nonmonotonic relationship between consumer surplus and the number of sponsored positions in search results. When the number of sponsored positions decreases, consumer surplus initially experiences a slight increase. However, as more sponsored positions are removed, consumer surplus begins to decline. Conversely, when additional sponsored positions are introduced, consumer surplus gradually decreases. There exists a similar trend for seller profits as illustrated in Panel C.[33]

The nonmonotonic relationship emphasizes that sponsored product advertising does not always benefit consumers and sellers. When I compare the status quo to a scenario with no sponsored positions at all, the findings in Section 7.1 indicate that both consumers and sellers, on average, benefit from the presence of sponsored product advertising. However, the introduction of additional sponsored positions, while increasing platform revenues, comes at the expense of consumer and seller welfare. Notably, both the consumer- and seller-optimal numbers of sponsored positions are slightly lower than the current number. In contrast, as illustrated in Panel D of Figure 9, the platform-optimal number is higher than the current number, suggesting that the platform may have an incentive to introduce several more sponsored positions to maximize its objectives. For policymakers primarily concerned with the welfare of consumers and sellers, a potential solution might involve implementing a limit on the number of sponsored positions rather than completely banning sponsored product advertising. This approach seeks to strike a balance that preserves the benefits of sponsored product advertising while safeguarding the interests of consumers and sellers.

[33] Appendix Figure A.15 presents the results under a fixed commission rate. As more sponsored positions are added, consumer surplus and seller profits consistently decrease. Platform revenues initially increase and then level off.

Figure 9: Welfare Effects Under Varying Numbers of Sponsored Positions



Notes: This figure examines the welfare effects of altering the number of sponsored positions in search results. I change the set of sponsored positions in two directions: (i) a gradual removal of sponsored positions, starting from the bottom and progressing to the top; and (ii) a gradual addition of more sponsored positions in the middle, starting by replacing the ninth organic position and moving downward. Panel A illustrates the optimal commission rate corresponding to various numbers of sponsored positions. Panels B to D depict the aggregate changes in consumer surplus, seller profits, and platform revenues relative to the status quo with 12 sponsored positions. Appendix Figure A.15 presents the results under a fixed commission rate.

8 Conclusion

In recent years, there has been a significant expansion in sponsored product advertising across various digital retail platforms offering a wide array of products and services. These platforms, which traditionally relied primarily on sales commissions for revenue, have embraced sponsored product advertising as an important part of their revenue streams. This shift has brought about a transformation in their business models. This evolution of the digital marketplace has profound

implications for all market participants.

This paper introduces a comprehensive framework for analyzing the welfare effects of sponsored product advertising. It begins with a stylized example that highlights the theoretical ambiguity regarding the welfare impact of sponsored product advertising. The empirical analysis leverages extensive datasets on Amazon searches, purchases, and advertising auction bids. Using estimates obtained from an equilibrium model, the study reveals that, under a fixed commission rate, removing sponsored product advertising on average benefits both consumers and sellers on Amazon. These effects are driven by the removal of low-utility sponsored products and the elimination of the pass-through of advertising costs to prices.

However, the removal of sponsored product advertising would prompt the platform to respond by increasing its commission rate, resulting in negative overall effects for consumers and sellers. These findings highlight the importance of considering various factors, including the platform's strategic response, when evaluating the implications of sponsored product advertising. For policymakers primarily concerned with the welfare of consumers and sellers, implementing a limit on the number of sponsored positions emerges as a more suitable policy than instituting an outright ban on sponsored product advertising.

I want to emphasize some limitations in my analysis that could motivate further research. First, my dataset covers a relatively short period, which constrains my ability to explore longer-term effects, such as seller entry and exit dynamics. Future studies with more extended time coverage could provide a more comprehensive understanding of these effects. Second, certain aspects of my analysis rely on observational data at an aggregate level, which necessitates specific assumptions in modeling and estimation. For instance, my model assumes perfect consumer awareness of product utility. Researchers with access to microdata on consumers' search behavior or experimental variations in product arrangement could potentially estimate a more flexible model. Additionally, I observe only a subset of keywords for which products appear. For keywords that I do not observe, I have to make restrictive assumptions regarding consumer demand and seller competition. Finally, my analysis abstracts from several practices employed by the platform, such as the relevance scores used to rank sellers in auctions and personalized sponsored results, which are designed to enhance the quality of sponsored results. By not incorporating these in my analysis, I may underestimate the potential benefits of sponsored product advertising.

References

Amazon (2017). 2016 Amazon.com Annual Report. https://s2.q4cdn.com/299287126/files/doc_financials/annual/2016-Annual-Report.pdf.

Amazon (2023). 2022 Amazon.com Annual Report. https://s2.q4cdn.com/299287126/files/doc_financials/2023/ar/Amazon-2022-Annual-Report.pdf.

Athey, S. and Ellison, G. (2011). Position Auctions with Consumer Search. *Quarterly Journal of Economics*, 126(3):1213–1270.

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

Blake, T., Nosko, C., and Tadelis, S. (2015). Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment. *Econometrica*, 83(1):155–174.

Browning, K. and Griffith, E. (2023). Instacart Reveals $428 Million in Profit but Slowing Growth Ahead of I.P.O. *New York Times*. https://www.nytimes.com/2023/08/25/technology/instacart-profit-slowing-growth-ipo.html.

Castillo, J. C. (2022). Who Benefits From Surge Pricing? Working Paper.

Chan, T. Y. and Park, Y.-H. (2015). Consumer Search Activities and the Value of Ad Positions in Sponsored Search Advertising. *Marketing Science*, 34(4):606–623.

Chen, Y. and He, C. (2011). Paid Placement: Advertising and Search on the Internet. *Economic Journal*, 121(556):F309–F328.

Choi, H. and Mela, C. F. (2019). Monetizing Online Marketplaces. *Marketing Science*, 38(6):948–972.

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

Errico, M. and Lashkari, D. (2022). The Quality of US Imports and the Consumption Gains From Globalization. Working Paper.

Farronato, C., Fradkin, A., and MacKay, A. (2023). Self-Preferencing at Amazon: Evidence From Search Rankings. Working Paper 30894, National Bureau of Economic Research.

Federal Trade Commission (2023). FTC Sues Amazon for Illegally Maintaining Monopoly Power. https://www.ftc.gov/news-events/news/press-releases/2023/09/ftc-sues-amazon-illegally-maintaining-monopoly-power.

Fowler, G. A. (2022). It's Not Your Imagination: Shopping on Amazon Has Gotten Worse. *Washing-*

ton Post. https://www.washingtonpost.com/technology/interactive/2022/amazon-shopping-ads/.

Grennan, M. (2013). Price Discrimination and Bargaining: Empirical Evidence From Medical Devices. *American Economic Review*, 103(1):145–177.

Gutierrez, G. (2022). The Welfare Consequences of Regulating Amazon. Working Paper.

Hui, X. and Liu, M. (2022). Quality Certificates Alleviate Consumer Aversion to Sponsored Search Advertising. CESifo Working Paper.

Ilango, A. (2022). Sponsored Search: Theory and Evidence on How Platforms Exacerbate Product Market Concentration. Working Paper.

Jungle Scout (2023). State of the Amazon Seller 2022. https://www.junglescout.com/amazon-seller-report/2022-results/.

Kaziukėnas, J. (2022a). Amazon GMV Reached $600 Billion in 2021. *Marketplace Pulse*. https://www.marketplacepulse.com/articles/amazon-gmv-reached-600-billion-in-2021.

Kaziukėnas, J. (2022b). Amazon Tops Six Million Third-Party Sellers. *Marketplace Pulse*. https://www.marketplacepulse.com/articles/amazon-reaches-six-million-third-party-sellers.

Lam, H. T. (2023). Platform Search Design and Market Power. Working Paper.

Lebow, S. (2022). Amazon Will Capture Nearly 40% of the US Ecommerce Market. *Insider Intelligence*. https://www.insiderintelligence.com/content/amazon-us-ecommerce-market.

Lee, K. H. and Musolff, L. (2023). Entry Into Two-Sided Markets Shaped by Platform-Guided Search. Working Paper.

Long, F., Jerath, K., and Sarvary, M. (2022). Designing An Online Retail Marketplace: Leveraging Information From Sponsored Advertising. *Marketing Science*, 41(1):115–138.

Marx, L. M. and Shaffer, G. (2010). Slotting Allowances and Scarce Shelf Space. *Journal of Economics & Management Strategy*, 19(3):575–603.

Matsakis, L. (2022). Small Businesses Say They Are Hurt by Rising Costs to Advertise on Amazon. *NBC News*. https://www.nbcnews.com/tech/small-businesses-say-are-hurt-rising-costs-advertise-amazon-rcna16685.

Motta, M. and Penta, A. (2022). Market Effects of Sponsored Search Auctions. CEPR Discussion Paper No. DP17401.

Ostrovsky, M. and Skrzypacz, A. (2022). Pure-Strategy Equilibrium in the Generalized First-Price Auction. Working Paper.

Sahni, N. S. and Nair, H. S. (2020). Does Advertising Serve as a Signal? Evidence From a Field Experiment in Mobile Search. *Review of Economic Studies*, 87(3):1529–1564.

Schein, A. I., Popescul, A., Ungar, L. H., and Pennock, D. M. (2002). Methods and Metrics for Cold-Start Recommendations. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 253–260.

Sudhir, K. and Rao, V. R. (2006). Do Slotting Allowances Enhance Efficiency or Hinder Competition? *Journal of Marketing Research*, 43(2):137–155.

Sullivan, M. W. (1997). Slotting Allowances and the Market for New Products. *Journal of Law and Economics*, 40(2):461–494.

Sweeting, A. (2013). Dynamic Product Positioning in Differentiated Product Markets: The Effect of Fees for Musical Performance Rights on the Commercial Radio Industry. *Econometrica*, 81(5):1763–1803.

Tanner, M. A. and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American statistical Association*, 82(398):528–540.

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

Varian, H. R. (2007). Position Auctions. *International Journal of Industrial Organization*, 25(6):1163–1178.

Wang, Z. (2022). I'm an Amazon Advertiser. Here's How I Used My Ad Budget To Pull In $100 Million in Sales in 2021. https://www.businessinsider.com/amazon-advertiser-hero-cosmetics-breaks-down-how-it-spends-its-ad-budget-2022-5.

Weise, K. (2019). Amazon Knows What You Buy. And It's Building a Big Ad Business From It. *New York Times*. https://www.nytimes.com/2019/01/20/technology/amazon-ads-advertising.html.

Yao, S. and Mela, C. F. (2011). A Dynamic Model of Sponsored Search Advertising. *Marketing Science*, 30(3):447–468.

Appendix

A Details on Stylized Example **A.1**

B Details on Data Collection **A.3**

B.1 Selection of Keywords and Construction of Markets A.3

B.2 Conversion of Best Sellers Ranks to Sales Quantities A.5

C Details on Demand Estimation **A.7**

C.1 Calibration of Fraction of Sales From Focal Keywords A.7

C.2 Calibration of Market Size A.9

C.3 Sample Restrictions A.10

C.4 Estimation and Empirical Bayes Procedure A.10

D Details on Supply Estimation **A.11**

D.1 Construction of Profits From Other Keywords A.11

D.2 Gibbs Sampler and Estimation of Bids and Organic Ranks A.14

E Details on Estimation of Platform's Weight **A.18**

E.1 Definition of Platform's Objective Function A.18

E.2 Alternative Set of Sponsored Positions A.19

F Model Extensions **A.19**

F.1 Ad Blindness and Quality of Sponsored Products A.19

F.2 Endogenous Consideration Sets and Search Costs A.23

F.3 Market-Specific Commission Rates A.27

G Robustness Checks **A.27**

G.1 Parametrization of Consumer Search Frictions A.27

G.2 Parametrization of Utility Function A.27

G.3 Alternative Autoregressive Models A.28

G.4 Alternative Extent of Search Frictions A.28

G.5 Distributional Assumption of Realized Bids A.29

G.6 Assumption on Duplicated Listings A.30

H Additional Figures and Tables **A.30**

A Details on Stylized Example

The stylized example presented in Section 3 illustrates the economic determinants that influence the welfare implications of sponsored product advertising. It captures the essential elements in sponsored product advertising: consumer search frictions, seller competition, and a profit-maximizing platform. In this section, I provide formal definitions of consumer demand and auction rules and derive the equilibria.

Consider a market with two sellers indexed by $j = 1, 2$, each offering a single product with quality δ_j, a marginal cost c_j, and a price p_j. The platform's search results consist of two positions, displaying either organic or sponsored results. Without loss of generality, I assume product 1 is always above product 2 in the organic ranking, although this may not align with the ranking of δ_j. The platform collects a commission for every unit sold, equal to a percentage of the price, denoted by $\tau \in (0, 1)$.

Consumer i's utility of purchasing product j follows a logit demand specification:

$$u_{ij} = \delta_j - \alpha p_j + \epsilon_{ij}, \tag{A.1}$$

where ϵ_{ij} is an idiosyncratic preference shock that follows a type I extreme value distribution. A fraction $\lambda \in (0, 1)$ of consumers only consider the product in the first position, while the remaining consumers consider products in both positions. This assumption implies the existence of search frictions and incentivizes sellers to be ranked higher in the search results. Consumers choose either a product they consider or the outside option, whose mean utility is normalized to be 0. Let $s_j(p_j)$ denote the market share of product j among consumers who only consider product j, and $s_j(p_1, p_2)$ denote the market share of product j among consumers considering both products. Specifically,

$$s_j(p_j) = \frac{\exp(\delta_j - \alpha p_j)}{1 + \exp(\delta_j - \alpha p_j)}, \quad s_j(p_1, p_2) = \frac{\exp(\delta_j - \alpha p_j)}{1 + \sum_{j'=1,2} \exp(\delta_{j'} - \alpha p_{j'})}.$$

Organic Only In the absence of sponsored product advertising, both positions display organic results. By assumption, product 1 occupies the first position and is considered by all consumers, whereas product 2 appears in the second position and is only considered by $(1 - \lambda)$ consumers who consider both options. Let $m_j = (1 - \tau)p_j - c_j$ denote seller j's margin and $\boldsymbol{p} = (p_1, p_2)$. Each

seller sets a price p_j to maximize its profits, taking into account its rank in the search results:

$$\begin{aligned}\max_{p_1}\ \pi_1^N(\boldsymbol{p};\tau) &= \big(\lambda s_1(p_1) + (1-\lambda)s_1(p_1,p_2)\big)m_1,\\ \max_{p_2}\ \pi_2^N(\boldsymbol{p};\tau) &= (1-\lambda)s_2(p_1,p_2)m_2.\end{aligned} \tag{A.2}$$

Let $p_j^N(\tau)$ and $s_j^N(\tau)$ denote the equilibrium prices and market shares under a commission rate τ. The platform chooses an optimal commission rate τ^N that maximizes its commissions:

$$\tau^N = \arg\max_{\tau}\ \tau\big(s_1^N(\tau)p_1^N(\tau) + s_2^N(\tau)p_2^N(\tau)\big). \tag{A.3}$$

With Sponsored Product Advertising Consider the case where the first position is allocated through a first-price auction on a pay-per-click basis.[34] The second position still shows an organic result, displaying product 1 by assumption. When seller j submits a bid b_j, the realized bid $\tilde{b}_j$ is stochastic and follows $\log(\tilde{b}_j) \overset{\text{i.i.d.}}{\sim} N(\log(b_j), \sigma^2)$. The seller with a higher realized bid occupies the first position. Given both sellers' bids $\boldsymbol{b} = (b_1, b_2)$, the probability that each seller wins the auction is

$$q_1(\boldsymbol{b}) = \Pr(b_1\omega_1 > b_2\omega_2) = \Phi\Big(\frac{\log(b_1/b_2)}{\sqrt{2}\sigma}\Big), \quad q_2(\boldsymbol{b}) = 1 - q_1(\boldsymbol{b}). \tag{A.4}$$

Conditional on winning, the seller pays its realized bid for each click on its listing. Let $F(\cdot)$ denote the distribution of $\boldsymbol{\omega} = (\omega_1, \omega_2)$. The expected payment per click is

$$\begin{aligned}\ell_1(\boldsymbol{b}) &= \frac{1}{q_1(\boldsymbol{b})}\int_{b_1\omega_1 > b_2\omega_2} b_1\omega_1 \mathrm{d}F(\boldsymbol{\omega}) = \frac{b_1}{q_1(\boldsymbol{b})}\int \Phi\Big(\frac{\log(b_1/b_2)}{\sigma} + u\Big)\exp(\sigma u)\phi(u)\mathrm{d}u,\\ \ell_2(\boldsymbol{b}) &= \frac{1}{q_2(\boldsymbol{b})}\int_{b_1\omega_1 < b_2\omega_2} b_2\omega_2 \mathrm{d}F(\boldsymbol{\omega}) = \frac{b_2}{q_2(\boldsymbol{b})}\int \Phi\Big(\frac{\log(b_2/b_1)}{\sigma} + u\Big)\exp(\sigma u)\phi(u)\mathrm{d}u.\end{aligned} \tag{A.5}$$

I assume the number of clicks is equal to a linear combination of the number of sponsored units and the number of impressions, with weights $\gamma_1 \geq 0$ and $\gamma_2 \geq 0$, respectively. Each seller sets a price p_j and submits a bid b_j to maximize its expected profits:

$$\begin{aligned}\max_{p_1,b_1}\ \pi_1^S(\boldsymbol{b},\boldsymbol{p};\tau) &= \big(q_1(\boldsymbol{b})s_1(p_1) + q_2(\boldsymbol{b})(1-\lambda)s_1(p_1,p_2)\big)m_1 - q_1(\boldsymbol{b})\ell_1(\boldsymbol{b})\big(\gamma_1 s_1(p_1) + \gamma_2\big),\\ \max_{p_2,b_2}\ \pi_2^S(\boldsymbol{b},\boldsymbol{p};\tau) &= q_2(\boldsymbol{b})\big(\lambda s_2(p_2) + (1-\lambda)s_2(p_1,p_2)\big)\big(m_2 - \gamma_1\ell_2(\boldsymbol{b})\big) - \gamma_2 q_2(\boldsymbol{b})\ell_2(\boldsymbol{b}).\end{aligned} \tag{A.6}$$

[34] The stylized example assumes a first-price auction to generate reasonable results with two bidders. The model in Section 5 considers a generalized second-price auction following Amazon's practice.

Let $p_j^S(\tau)$, $b_j^S(\tau)$, and $s_j^S(\tau)$ denote the equilibrium prices, bids, and market shares under a commission rate τ. Define $q_j^S(\tau) = q_j(\boldsymbol{b}^S(\tau))$ and $\ell_j^S(\tau) = \ell_j(\boldsymbol{b}^S(\tau))$. The platform chooses an optimal commission rate τ^S that maximizes the sum of commissions and advertising revenues:

$$\begin{aligned}\tau^S = \arg\max_{\tau}\ & \tau\big(s_1^S(\tau)p_1^S(\tau) + s_2^S(\tau)p_2^S(\tau)\big) + q_1^S(\tau)\ell_1^S(\tau)\Big(\gamma_1 s_1\big(p_1^S(\tau)\big) + \gamma_2\Big) \\ & + q_2^S(\tau)\ell_2^S(\tau)\Big(\gamma_1\lambda s_2\big(p_2^S(\tau)\big) + \gamma_1(1-\lambda)s_2\big(p_1^S(\tau), p_2^S(\tau)\big) + \gamma_2\Big). \qquad \text{(A.7)}\end{aligned}$$

Simulations Throughout all simulations, I set $\lambda = 0.4$, $\alpha = 0.8$, $\sigma = 0.5$, $\gamma_1 = 1$, and $\gamma_2 = 0.1$. In the main results discussed in Section 3, I assume both products have the same marginal cost $c_1 = c_2 = 4$. I hold the quality of product 1 constant at $\delta_1 = 4$ while varying product 2's quality δ_2 within a range from 2 to 6, centered around the value of δ_1. Figure 1 and Appendix Figure A.5 depict the simulation results. In the supplementary analysis presented in Appendix Figures A.16 and A.17, I assume both products have the same quality $\delta_1 = \delta_2 = 4$. I hold the marginal cost of product 1 constant at $c_1 = 5$ while varying product 2's marginal cost c_2 within a range from 3 to 7, centered around the value of c_1.

Given any parameter, I first solve the equilibrium without sponsored product advertising, characterized by the platform's commission rate τ^N and sellers' prices $\boldsymbol{p}^N(\tau^N)$. To assess how different channels contribute to the effect of introducing a sponsored position, the simulations explore three distinct assumptions concerning the responses of sellers and the platform following the introduction of a sponsored position. Under the first and most restrictive assumption, neither sellers can adjust prices nor can the platform modify the commission rate. So the commission rate remains at τ^N and sellers determine their bids taking prices $\boldsymbol{p}^N(\tau^N)$ as given. Under the second assumption, sellers can set new prices, but the commission rate remains fixed at τ^N. Finally, under the most flexible assumption, both sellers and the platform can choose new prices and commission rate. For each of these assumptions, I compute the change in welfare measures, moving from the equilibrium with only organic results to the one with a sponsored position.

B Details on Data Collection

B.1 Selection of Keywords and Construction of Markets

The main dataset introduced in Section 4.1 is a large set of scraped search results on Amazon for 3,237 keywords. In this section, I describe the procedure of constructing the set of keywords, with

the objective of identifying high-search-volume keywords on Amazon and categorizing them into different markets using a data-driven method.

I started from the frequently searched keywords reported in Yin and Jeffries (2021). The authors collected the 300 most popular keywords from the first to the third quarters of 2020 reported by Amazon across nine categories, including "Automotive," "Baby," "Beauty," "Electronics," "Grocery," "Office Products," "Softlines," "Toys," and "Amazon.com." This compilation yielded a total of 2,700 unique keywords.[35]

To further enrich this dataset, I searched these keywords on Jungle Scout, a renowned marketing service provider for Amazon sellers. For each keyword, Jungle Scout furnished an extensive list of related keywords, ranging from a few hundred to more than ten thousand, alongside their estimated 30-day search volumes for the prior month.[36] Within this expanded dataset of keywords, I kept keywords with a 30-day search volume of at least 100,000 and then manually eliminated keywords falling into four categories: (i) general terms, like "my orders" or "prime"; (ii) keywords that are too broad, like "car accessories" or "home decor"; (iii) keywords specific to a brand, like "Nintendo switch" or "iPhone"; and (iv) keywords related to a specific time, like "Christmas lights," or "2021 monthly planner."

To identify a comprehensive set of high-search-volume keywords, I executed an iterative process, repeatedly following the aforementioned steps: identifying relevant keywords, keeping those with high search volumes, and excluding keywords falling into those categories. This iterative process ended when no new relevant keywords could be identified. The process took place in April 2022 and yielded an extensive catalog of 904 keywords that had a 30-day search volume of at least 100,000 and satisfied the exclusion criteria outlined earlier, which I referred to as *focal keywords*.

Within the compiled list of focal keywords, some keywords could be close substitutes, potentially encompassing products within the same market. I implemented a data-driven method to define markets. In essence, I grouped keywords that shared a significant number of overlapping products in their search results within the same market. Specifically, I conducted searches for each focal keyword on Amazon over three separate days in April 2022 and kept all products appearing on the first page of the search results. Let $\mathcal{J}_k$ denote the set of products that appeared in the search results of keyword k. I defined the similarity of two keywords k_1 and k_2 as the fraction of

[35] The original data can be found here.

[36] I focused on the *exact* search volume provided by Jungle Scout, which accounted for searches with identical phrases, excluding plurals, misspellings, additional or omitted words, or synonyms.

overlapping products:

$$s_{k_1,k_2} = \frac{|\mathcal{J}_{k_1} \cap \mathcal{J}_{k_2}|}{\sqrt{|\mathcal{J}_{k_1}||\mathcal{J}_{k_2}|}}. \tag{B.8}$$

This metric reflected the extent of product overlap between the two keywords.

I represented the set of focal keywords using a graph, where each focal keyword was depicted as a node. To establish connections between nodes (keywords), I introduced edges between two nodes if they shared a minimum of 10% of their respective products, i.e., $s_{k_1,k_2} \geq 0.1$. I defined a *market* as a connected component within this graph, which partitioned the focal keywords into distinct markets. An average market encompassed 1.7 focal keywords. Two keywords not directly linked by an edge in the graph could still belong to the same market if they were connected through other intermediary keywords within the network.

For each focal keyword, Jungle Scout provided a list of *relevant keywords* along with a relevance score, rated on a scale of 1-100, which indicated the degree of relevance each keyword had to the focal keyword. I retained all relevant keywords that met two criteria: they had a relevance score of at least 30, and their 30-day search volume had to be at least 10,000. When a single market encompassed multiple focal keywords, I merged their lists of relevant keywords. In the rare instances where a keyword appeared in multiple markets, I only retained that keyword in the market with the highest relevance score.

This process produced a final list of 3,237 keywords categorized into 546 markets. These keywords collectively accounted for a total daily search volume of nearly nine million. Appendix Table A.1 presents examples of markets and keywords.

B.2 Conversion of Best Sellers Ranks to Sales Quantities

A crucial element for demand estimation is the daily sales quantities of each product. However, Amazon does not publicly disclose sales quantities for any product sold on Amazon. Instead, it assigns a Best Sellers Rank (BSR) to each product based on its recent sales performance, which is updated several times a day. In this section, I outline how I utilize a well-established tool to convert a product's BSR to an estimate of its daily sales quantity. Additionally, I validate the accuracy of this conversion method using actual sales data.

Amazon describes the BSR as a ranking system designed to assess a product's sales performance relative to other products within the same category over a recent period of time. The BSR is defined within one of approximately 20 product categories on Amazon, such as "Baby Prod-

ucts," "Beauty & Personal Care," "Clothing, Shoes & Jewelry," "Health & Household," "Home & Kitchen," "Office Products," and "Pet Supplies," among others. While Amazon maintains confidentiality regarding the exact algorithm employed to calculate the BSR, it is generally believed that the BSR takes into account both recent and historical sales, with more recent sales carrying significantly higher weights.

Prior studies have documented that when the number of products is sufficiently large, there exists a relatively stable and predictable relationship between a product's sales rank and its sales quantities (e.g., Chevalier and Goolsbee 2003; Brynjolfsson et al. 2011). This relationship is often modeled using functional forms such as power functions or splines.

I utilized a tool developed by Jungle Scout. Jungle Scout provided marketing services to many Amazon sellers of varying sizes selling products in different categories, thus has access to a substantial amount of actual sales data from Amazon sellers. This tool used these actual sales data to fit a mapping between the BSR and daily sales quantities in each category. I obtained these mappings estimated by Jungle Scout on May 16, 2022. Appendix Figure A.18 presents mappings for six common categories.[37] Using these mappings, I converted the BSR to the sales quantity for each product on each day in my sample.

I validated the conversion from the BSR to daily sales quantities using actual daily sales data obtained in collaboration with a prominent Amazon seller, Forum Brands. This dataset encompassed several hundred products spanning various product categories on Amazon and covered two years. For each product, I compared its estimated daily sales quantities derived from the BSR-to-quantity mappings with its actual daily sales quantities.

Appendix Figure A.6 depicts these two series for several randomly chosen products with a wide range of sales volumes. For most products, the two series align remarkably well. I also computed the correlation between estimated and actual daily quantities for each product. Appendix Figure A.19 displays the distribution of the correlation across all products sold by the company. The median correlation is 0.81, suggesting that the conversion method is reasonably accurate and captures most day-to-day variation in product sales.

[37] The BSR can range as high as 600,000, so it was impractical to collect estimated quantities for every BSR value. I implemented a sampling strategy within each product category. I selected a set of BSR values with roughly equal spacing on the logarithmic scale, such as 1, 2, 3, ..., 10, 12, 14, ...,20, 25, 30, ..., 50, 60, 70, ..., 100, 120, 140, ..., 200, 250, 300, ..., 500, 600, 700,..., 1000, 1200, 1400, ..., 2000, 2500, 3000, ..., 5000, 6000, 7000, ... up to the point where the estimated sales quantity is 0. I then collected the corresponding estimated quantities for these selected BSR values. I then employed a log-linear interpolation method to impute daily sales quantities for all potential BSR values. I conducted spot checks to verify that this imputation process did not introduce any meaningful measurement errors.

C Details on Demand Estimation

C.1 Calibration of Fraction of Sales From Focal Keywords

The demand model introduced in Section 5.1.1 calculates the market share of each product in one keyword. As in Appendix B.1, I refer to the keyword with the highest search volume in each market as the "focal keyword." However, what I observe in the data is a product's total sales quantity on the platform. In reality, a product can appear in the search results of multiple keywords. Disregarding these additional keywords in the estimation can lead to biased estimates of search frictions.

To illustrate this bias, consider a product that ranks poorly in the focal keyword. Search frictions are so high that few consumers have this product in their consideration sets. However, this same product might rank higher in other keywords that are a better match, resulting in sales from those keywords. Attributing all of its sales solely to the focal keyword would lead to an incorrect conclusion that search frictions are low.

In this section, I describe the method used to calibrate the fraction of a product's total sales quantity attributed to the focal keyword, represented by $\kappa_{jt} \in (0,1)$. Let q_{jt} denote the total sales quantity of product j on day t, and M_t the market size of the focal keyword, calibrated in Appendix C.2. For each keyword, I conducted a total of $F = 6$ searches each day, indexed by f. These searches are assumed to represent what consumers typically encounter on that day, and I calculate the average market share across all of these searches. Therefore, equation (4) can be reformulated as follows:

$$s_{jt}(\boldsymbol{p}) = \frac{\kappa_{jt} q_{jt}}{M_t} = \frac{1}{F} \sum_{f=1}^{F} \sum_{n=1}^{N} \mathbf{1}(j \in \mathcal{J}_{nt}^f)(\lambda_n - \lambda_{n+1}) \frac{\exp(\delta_{jt})}{1 + \sum_{j' \in \mathcal{J}_{nt}^f} \exp(\delta_{j't})}. \tag{C.9}$$

Here, $\mathcal{J}_{nt}^f$ represents the set of products in the top n positions within search f on day t.

The calibration of κ_{jt} relies on external data. I randomly selected 480 products across various product categories, which appeared in the search results of the focal keywords with a wide range of ranks. I refer to this sample of products as the "prediction sample." I leveraged a tool developed by Jungle Scout. For each product j, this tool provided a comprehensive list of keywords for which the product had recently appeared on the first page of search results, indexed by $k \in \mathcal{K}_j$. In addition, the tool also reported the rank of this product in each keyword, indicated by n_{jk}, and the

search volume of the keyword, represented by V_k.

Similar to Section 5.1.2, I parametrize the fraction of consumers whose consideration sets include the product in the n-th position of the search results as $\lambda_n = 1/\exp(\beta(n-1))$, where $\beta \geq 0$ determines the extent of search frictions. I assume that the mean utility of a product in the n-th position is $\delta_n(\gamma) = -\gamma \log(n)$. Consequently, product j's total sales quantity from any arbitrary set of keywords $\mathcal{K}$ can be calculated as

$$q_j(\mathcal{K}; \beta, \gamma) = \sum_{k \in \mathcal{K}} V_k \sum_{n=n_{jk}}^{N} (\lambda_n - \lambda_{n+1}) \frac{\exp(\delta_{n_{jk}})}{1 + \sum_{n'=1}^{n} \exp(\delta_{n'})}. \tag{C.10}$$

Let $k_j^0 \in \mathcal{K}_j$ represent the focal keyword. I can determine the fraction of product j' total sales quantity attributed to the focal keyword with the following equation:

$$\kappa_j(\beta, \gamma) = \frac{q_j(\{k_j^0\}; \beta, \gamma)}{q_j(\mathcal{K}_j; \beta, \gamma)}. \tag{C.11}$$

The next step involves predicting κ_{jt} for all products in the estimation sample by leveraging the relationships observed in the prediction sample. To achieve this, I utilize several predictors that can be constructed both in the prediction sample and in the estimation sample. As discussed in Section 3 and Appendix B.1, in addition to the 546 focal keywords, I also keep track of search results for several thousand other keywords, with an average of about five keywords in each market. Let $\overline{K}$ represent the set of all tracked keywords (so $k_j^0 \in \overline{K}, \forall j$). For each product j, I construct the following predictors:

- Log of the rank of the product in the focal keyword, $\log(r_{jk_j^0})$.
- Indicator for the product appearing in other tracked keywords, $\mathbf{1}(\overline{K} \cap (\mathcal{K}_j \setminus \{k_j^0\}) \neq \emptyset)$.
- The search volume of the focal keyword as a fraction of the search volume of all tracked keywords that this product appears in, $V_{k_j^0} / \sum_{k \in \overline{K} \cap \mathcal{K}_j} V_k$.
- Log of the average rank of the product in other tracked keywords $\log(\frac{\sum_{k \in \overline{K} \cap (\mathcal{K}_j \setminus \{k_j^0\})} r_{jk}}{\sum_{k \in \overline{K} \cap (\mathcal{K}_j \setminus \{k_j^0\})} 1})$.[38]

I regress $\kappa_j(\beta, \gamma)$ on these four variables. Appendix Table A.2 presents the regression results under different values of β. Notably, the R^2 ranges from 0.35 to 0.38, indicating a good fit of the

[38] If $\overline{K} \cap (\mathcal{K}_j \setminus \{k_j^0\} = \emptyset$, this variable is set to a constant. The specific value of this constant does not affect the analysis since there's already an indicator variable $\mathbf{1}(\overline{K} \cap (\mathcal{K}_j \setminus \{k_j^0\}) \neq \emptyset)$ in the regression.

model. The signs of the coefficients align with expectations. For instance, the fraction tends to be lower when a product ranks low in the focal keyword but high in other tracked keywords. Panel A of Appendix Figure A.20 compares the predicted fractions to the actual ones, confirming that the predictions perform reasonably well and capture the observed patterns.

The final step involves applying the prediction model to the estimation sample to calibrate κ_{jt} for each product on each day. This calibration process depends on two parameters, β and γ. β is also a parameter to be estimated and γ depends on the estimated product utility. I iterate between the calibration and the demand estimation until the values of β and γ used in the calibration match the estimated β and implied γ averaged across all markets. Panel B of Appendix Figure A.20 plots the distribution of κ_{jt} under the final calibration across all markets, products, and days, where I restrict κ_{jt} within the range of $[0.001, 1]$.

C.2 Calibration of Market Size

The market size is another input to a product's market share in equation (C.9). In this section, I describe the calibration of the market size. Estimating market size for each keyword is challenging, as consumers may search for multiple keywords during the same visit to Amazon or the same keyword across different visits when looking for a product. While industry reports commonly suggest that Amazon holds around a 50% market share of online purchases (Lam 2023), this share can be significantly different at the *search* level.

To address this challenge, I calibrate the market size using aggregate data from industry reports. The calibration depends on two essential statistics: (i) the total number of units sold on Amazon and (ii) the total number of searches on Amazon.

Amazon's earnings release reveals that total sales on Amazon's marketplaces amounted to approximately $118.4 billion in the second quarter (Q2) of 2022,[39] and North America accounted for 62%. My analysis focuses on Amazon.com, a domain primarily used in the US market. Across all markets in my sample, the sales-weighted average price is $35.5. Therefore, I estimate the total quarterly units sold in North America to be roughly 2.1 billion units ($\$118.4 \times 62\% \div \$35.5$).

Furthermore, there were approximately 2.4 billion monthly visits to Amazon.com in May 2022.[40] I assume that, on average, consumers perform two keyword searches during each visit.

[39] According to Amazon's 2022 Q2 earnings release, available at this link (page 17), Amazon's online stores (i.e., first-party products) generated revenues of $50.9 billion in 2022 Q2, and 57% of units were sold by third-party sellers. Since the prices of products sold by third-party sellers and by Amazon were similar, total sales amounted to $118.4 billion.

[40] Source https://www.statista.com/statistics/623566/web-visits-to-amazoncom/.

Hence, the total number of searches is 14.4 billion (2.4 billion visits $\times 2$ searches per visit $\times 3$ months). Combining these two statistics, I can estimate the average inside market share as 2.1 billion units sold divided by 14.4 billion searches, or 14.6%. As a result, I calibrate the total market size of each market, denoted as M_t, as follows:

$$M_t = \frac{1}{0.146} \sum_{j \in \mathcal{J}} \kappa_{jt} q_{jt}. \tag{C.12}$$

C.3 Sample Restrictions

This section outlines the sample restrictions applied in the demand estimation. I impose several restrictions at the product level. First, I exclude products with missing data for sales quantities or prices. Second, I drop products that appear in the search results only once out of six times on any given day. Third, I exclude products with prices exceeding five times or falling below 20% of the median price in each market. Fourthly, I remove products with calibrated fractions of sales attributed to the focal keyword less than 5%. Last, I drop products that appear in the bottom one-third of search results but have market shares exceeding the 90th percentile of all products in each market, or appear in the bottom two-thirds of search results but have market shares exceeding the 95th percentile of all products. This last restriction is based on spot checks and aims to eliminate products that may be popular in other markets but less relevant in the focal market.

I also make several restrictions at the market level. I drop markets (i) where the search results do not consistently contain 60 positions, (ii) where over 40% of products are excluded based on the previous criteria, or (iii) where the median bid exceeds 15% of the median price, which likely reflects miscoded data.

C.4 Estimation and Empirical Bayes Procedure

I estimate the parameters (β, ρ) using the Method of Moments separately in each market with the moment conditons in equation (8). The estimation in each market takes the following steps.

1. Given β, invert market shares s_{jt} to get δ_{jt} using equation (C.9).

2. Regress δ_{jt} on product dummies, time dummies, product-specific linear time trends, and p_{jt}. Calculate the residuals as the unobserved demand shock ξ_{jt}.

3. Given ρ, calculate the innovation $\eta_{jt} = \xi_{jt} - \rho \xi_{j,t-1}$.

4. Construct the moment in equation (8) and search for parameters (β, ρ) that minimize $|m|$.

To adjust for sampling errors, I employ an empirical Bayes procedure to shrink the market-specific parameter $\hat{\beta}$ in the demand estimation towards its mean value across markets. Let m index markets. Let $\hat{\beta}^m$ denote the estimated parameter in market m and $se(\hat{\beta}^m)$ its standard error. I compute the shrunken parameter using the following formula:

$$\tilde{\beta}^m = \frac{\sigma_\beta^2}{\sigma^2 + se(\hat{\beta}^m)^2}\hat{\beta}^m + \frac{se(\hat{\beta}^m)^2}{\sigma_\beta^2 + se(\hat{\beta}^m)^2}\bar{\beta}, \tag{C.13}$$

where $\bar{\beta}$ represents the mean of $\hat{\beta}^m$ across markets, and σ_β denotes the standard deviation of $\hat{\beta}^m$. This procedure effectively adjusts market-specific parameter estimates to account for variability and ensures more robust estimates.

D Details on Supply Estimation

D.1 Construction of Profits From Other Keywords

Sellers on the platform typically make multiple keyword-specific bidding decisions and a single pricing decision that applies to all keywords. However, estimating such a model is not feasible due to data and computational limitations. On the data side, I do not observe search results from all possible keywords. Even if I did observe search results from a comprehensive set of keywords, estimating a model where each seller simultaneously makes many interdependent decisions is computationally challenging.

To address this problem, I adopt a partial-equilibrium approach, focusing on sellers' decisions within the focal keyword, which is the keyword with the highest search volume in a market (for more details, see Appendix B.1). Specifically, when seller j submits a bid b_j for a particular keyword and sets a price p_j, it takes its bidding decisions in other keywords as given, but recognizes that changing its price can impact its sales across all keywords.

As introduced in Section 5.2.1, seller j's expected total profits can be expressed as:

$$\pi_j(\boldsymbol{p}, \boldsymbol{b}) = \left(s_j^e(\boldsymbol{p}, \boldsymbol{b}) + M_j^{other} s_j^{other}(p_j)\right)\left((1-\tau)\, p_j - c_j\right) - ad_j^e(\boldsymbol{p}, \boldsymbol{b}) - ad_j^{other}(p_j). \tag{D.14}$$

In this equation, $s_j^e(\boldsymbol{p}, \boldsymbol{b})$ and $ad_j^e(\boldsymbol{p}, \boldsymbol{b})$ represent seller j's expected market share and advertising payments within the focal keyword, as defined in equation (10). $s_j^{other}(p_j)$ denotes seller j' aggre-

gate market share in other keywords, M_j^{other} represents the relative market size of other keywords compared to the focal keyword, and $ad_j^{other}(p_j)$ indicates seller j's total advertising payments in other keywords. This section derives these three elements: $s_j^{other}(p_j)$, M_j^{other}, and $ad_j^{other}(p_j)$.

Let s_j^0 and $s_j^{0,other}$ denote seller j's market share in the status quo within the focal keyword and in other keywords, respectively. While I can directly calculate s_j^0 based on the demand model and realized search results of the focal keyword, I cannot calculate $s_j^{0,other}$ due to a lack of search results in other keywords. Therefore, I assume

$$s_j^{0,other} = s_j^0. \tag{D.15}$$

This assumption is grounded in the idea that the competitive environment and strategic interactions among sellers in other keywords closely resemble those in the focal keyword. After all, these sellers are all part of the same market. Hence, I consider the market outcome observed in the focal keyword as a reasonable proxy for the market outcomes in other keywords.

In Appendix C.1, I calibrate the fraction of product j's total sales that originate from the focal keyword on day t, denoted as $\kappa_{jt} \in [0,1]$. When a product does not appear in the search results on a given day, I set $\kappa_{jt} = 0$. I calculate the average of this fraction across all days within a week, denoted as κ_j. The relative market size of other keywords compared to the focal keyword can then be derived as

$$\kappa_j = \frac{s_j^0}{s_j^0 + M_j^{other} s_j^{0,other}} = \frac{1}{1 + M_j^{other}} \quad \Rightarrow \quad M_j^{other} = \frac{1-\kappa_j}{\kappa_j}. \tag{D.16}$$

A product can generate both sponsored and organic sales within the focal keyword or in other keywords. Let $s_{j,O}^0$ and $s_{j,S}^0$ represent seller j's market shares for organic and sponsored sales within the focal keyword in the status quo, respectively. Let $s_{j,O}^{0,other}$ and $s_{j,S}^{0,other}$ denote the corresponding market shares in other keywords. By definition, these quantities satisfy the relationships $s_{j,O}^0 + s_{j,O}^0 = s_j^0$ for the focal keyword and $s_{j,O}^{0,other} + s_{j,S}^{0,other} = s_j^{0,other}$ for other keywords. I can derive $s_{j,O}^0$ and $s_{j,S}^0$ within the focal keyword using the demand model. To extend this division to other keywords, I assume that the split between sponsored and organic sales in other keywords mirrors the observed split in the focal keyword, with one modification described below.

To approximate the division between sponsored and organic sales in other keywords, I rely on the ratio of seller j's sponsored sales to total sales within the focal keyword in the status quo. This ratio is calculated as $r_j = s_{j,S}^0 / s_j^0$ and falls within the range [0,1]. It quantifies the relative importance of sponsored sales for seller j within the focal keyword. I extend this ratio to other

keywords using the following formula:

$$s_{j,S}^{0,other} = \min\{0.5, \frac{s_{j,S}^{0}}{s_{j}^{0}}\} \times s_{j}^{0,other}, \;\; s_{j,O}^{0,other} = s_{j}^{0,other} - s_{j,S}^{0,other}. \tag{D.17}$$

It is worth highlighting that in equation (D.17), I cap the fraction of sponsored sales at 0.5. This constraint ensures that, in the status quo, no product can have more than 50% of its sales originating from sponsored positions in other keywords. The reason for imposing this limitation is to prevent the model from implying an unreasonably high proportion of sponsored sales on Amazon, which would not align with observed data. Specifically, I can calculate the ratio between sponsored and organic sales aggregated across all keywords as

$$\frac{\sum_{j} p_j(s_{j,S}^{0} + M_{j}^{other} s_{j,S}^{0,other})}{\sum_{j} p_j(s_{j,O}^{0} + M_{j}^{other} s_{j,O}^{0,other})}. \tag{D.18}$$

Without the cap imposed in equation (D.17), $s_{j,S}^{0,other} = s_{j,S}^{0}$ and $s_{j,O}^{0,other} = s_{j,S}^{0}$, resulting in a ratio between sponsored and organic sales in all keywords of 34%, significantly higher than the actual ratio of around 20% in 2022. By imposing a cap of 0.5 in equation (D.17), the ratio in equation (D.18) has a more reasonable value of 21%.

Having established seller j's organic and sponsored market shares in other keywords in the status quo, I now need to specify how these shares respond to changes in the product's price. I define $s_{j,O}^{other}(p_j)$ and $s_{j,S}^{other}(p_j)$ as seller j's organic and sponsored market shares in other keywords under any given price p_j. I employ a logit model to approximate both functions. For either organic ($G = O$) or sponsored ($G = S$) market share and for any price p_j, I assume the following relationship:

$$\log\left(s_{j,G}^{other}(p_j)\right) = \log(s_{j,G}^{0,other}) - \alpha(p_j - p_j^0), \;\; G \in \{O, S\}. \tag{D.19}$$

Here, p_j^0 represents product j's price in the status quo, and α characterizes consumers' price sensitivity and is the same parameter used in equation (3).

The final step is to construct seller j's advertising payments in other keywords, $ad_j^{other}(p_j)$. Let $\ell_j(\boldsymbol{p}, \boldsymbol{b})$ represent the advertising payments per sponsored unit for product j in the focal keyword, which is defined as $\ell_j(\boldsymbol{p}, \boldsymbol{b}) = ad_j^e(\boldsymbol{p}, \boldsymbol{b})/s_{j,S}^e(\boldsymbol{p}, \boldsymbol{b})$. Similarly, I define $\ell_j^{other}(p_j)$ as the expected advertising payments per sponsored unit in other keywords for product j. I can then express

seller j's advertising payments in other keywords as

$$ad_j^{other}(p_j) = M_j^{other} \ell_j^{other} s_{j,S}^{other}(p_j), \tag{D.20}$$

where M_j^{other} is determined using equation (D.16) and $s_{j,S}^{other}(p_j)$ is defined by equation (D.19).

Auctions in different keywords are independent, so a seller's bid in the focal keyword affects only its sales and advertising payments within that keyword without impacting other keywords. In the partial equilibrium I consider, each seller takes its bidding decisions in other keywords as given. Therefore, for a given price, the expected advertising payments in other keywords do not affect the seller's bidding decision in the focal keyword. In equilibrium, I assume the expected advertising payments per sponsored unit in other keywords are the same as those in the focal keyword:

$$\ell_j^{other}(p_j) = \ell_j(\boldsymbol{p}, \boldsymbol{b}). \tag{D.21}$$

It is worth emphasizing that equation (D.21) holds only in equilibrium. When a seller makes its bidding decision in the focal keyword, it does not consider its sponsored sales or advertising payments in other keywords.

D.2 Gibbs Sampler and Estimation of Bids and Organic Ranks

When sellers make weekly bidding and pricing decisions, they consider the uncertainty stemming from the unpredictability of realized search results on Amazon. These search results are subject to two primary sources of uncertainty. First, organic ranks are determined by Amazon's real-time ranking algorithm and can vary from one search to another. Second, auction outcomes can change frequently over time, driven by many real-world factors. This section describes the methodology to estimate the uncertainty associated with realized search results. Specifically, this involves separately estimating sellers' bids and the distribution of organic ranks.

Instead of observing individual sellers' bids, I only have access to realized auction outcomes from numerous searches, along with aggregate statistics such as the median, lowest, and highest winning bids for each keyword. The aim is to infer sellers' bidding decisions and the associated uncertainty most likely to produce the observed auction outcomes.

As discussed in Section 5.2.1, I assume when seller j submits a bid b_j, the realized bid in search

f is subject to randomness and follows a log-normal distribution:

$$\log(\tilde{b}_j^f) \overset{\text{i.i.d.}}{\sim} N(\log(b_j), \sigma_b^2). \tag{D.22}$$

Sellers are ranked based on the value of $\tilde{b}_j^f$. Let r_j^{f*} represent the underlying rank of product j in search f. However, only the top R sellers are displayed in sponsored positions, and thus, I can only observe the rank when $r_j^{f*} \leq R$. The observed rank, denoted by r_j^f, is determined as follows:

$$r_j^f = \begin{cases} r_j^{f*}, & \text{if } r_j^{f*} \leq R \\ \emptyset, & \text{if } r_j^{f*} > R \end{cases}. \tag{D.23}$$

The parameters of interest are $\boldsymbol{\theta} = \left\{\{b_j\}_{j=1}^J, \sigma_b\right\}$. The data consists of observed auction ranks $\{r_j^f\}_{j,f}$ and two aggregate statistics: the median winning bid denoted as $\overline{M}$ and the range of winning bid, calculated as the highest winning bid minus the lowest one and represented as $\overline{R}$.

In this model, the relative ranking of sellers remains unchanged if I replace b_j with $a_1(b_j)^{a_2}$ and σ_b with $a_2\sigma_b$ for any $a_1, a_2 > 0$. However, this transformation will impact the median and range of winning bids. The estimation follows two steps. I first fix two parameters in $\boldsymbol{\theta}$ (e.g., b_1 and σ_b) and estimate the remaining parameters using the observed auction ranks. I then estimate the two parameters to match the two aggregate statistics.

In the first step of estimation, a conceptually straightforward approach involves using the maximum likelihood estimator (MLE). Let $(R_1, \ldots, R_J)$ be a vector of random variables representing the observed rank of each product. The likelihood function is:

$$f(\boldsymbol{\theta}) = \prod_{f=1}^{F} \Pr\left((R_1, \ldots, R_J) = (r_1^f, \ldots, r_J^f)\right) = \prod_{f=1}^{F} \Pr(\tilde{b}_{j_{1f}}^f > \tilde{b}_{j_{2f}}^f > \cdots > \tilde{b}_{j_{Rf}}^f > \max_{j: r_j^f = \emptyset} \tilde{b}_j^f), \tag{D.24}$$

where j_{rf} refers to the product that ranks the r-th in the auction of search f. However, the likelihood function lacks an analytical expression. For a large number of bidders (ranging from 30-150 in my data), numerical integration or approximation methods may not be feasible and could introduce significant errors.

I estimate the model parameters using the Gibbs sampler. The Gibbs sampler is a Markov chain Monte Carlo (MCMC) algorithm to generate a sequence of observations approximated from a specified multivariate probability distribution. In my model, the Gibbs sampler is tractable be-

cause the econometric model is fully parameterized. A notable advantage of the Gibbs sampler is that it allows for data augmentation of latent variables (Tanner and Wong 1987). I augment the realized bids in each search, $\tilde{b}_j^f$, as pseudo parameters of the model. This augmentation approach allows me to break down simulations from the posterior distribution into smaller steps. Specifically, conditional on draws of b_j and the observed data, the posterior distribution of $\log(\tilde{b}_j^f)$ is a truncated normal distribution. Conditional on draws of $\tilde{b}_j^f$, the posterior distribution of $\log(b_j)$ is a normal distribution.

In the first step of estimation, I fix the parameters $b_1 = 1$ and $\sigma_b = 0.1$, resulting in a set of parameters to be estimated denoted as $\tilde{\boldsymbol{\theta}} = \{\{b_j\}_{j=2}^J, \{\tilde{b}_j^f\}_{j,f}\}$. For any parameter $\delta \in \tilde{\boldsymbol{\theta}}$, I use $\delta|\tilde{\boldsymbol{\theta}}^-$ to denote the parameter δ conditional on all other parameters in $\tilde{\boldsymbol{\theta}}$ and all the observed data. I employ a conventional diffuse prior distribution for each parameter in $\tilde{\boldsymbol{\theta}}$: $\log(b_j) \sim N(0, \Sigma_0)$ and $\log(\tilde{b}_j^f) \sim N(0, \Sigma_0)$ with $\Sigma_0^{-1} = 0$ for all j and f. I can then express the posterior distribution of each parameter, given all other parameters and the observed data, as follows:

$$\log(b_j)|\tilde{\boldsymbol{\theta}}^- = N(\frac{1}{F}\sum_{f=1}^{F}\log(\tilde{b}_j^f), \frac{\sigma_b^2}{F}), \quad j = 2, \ldots, J,$$

$$\log(\tilde{b}_j^f)|\tilde{\boldsymbol{\theta}}^- \sim TN\left(\log(b_j), \sigma_b^2, \log(\underline{b}_j^f), \log(\overline{b}_j^f)\right), \quad j = 1, \ldots, J,$$

$$\underline{b}_j^f = \begin{cases} 0, & r_j^f = \emptyset \\ \max\{\tilde{b}_{j'}^f\}_{r_{j'}^f = \emptyset'}, & r_j^f = R \\ \tilde{b}^f_{j_{r_j^f+1,f}}, & r_j^f < R \end{cases}, \quad \overline{b}_j^f = \begin{cases} \min\{\tilde{b}_{j'}^f\}_{r_{j'}^f \neq \emptyset'}, & r_j^f = \emptyset \\ \tilde{b}^f_{j_{r_j^f-1,f}}, & 1 < r_j^f \leq R \\ \infty, & r_j^f = 1 \end{cases}. \tag{D.25}$$

Here, $TN(\mu, \sigma^2, a, b)$ represents a truncated normal distribution with support $[a, b]$. I ran the algorithm for 5,000 iterations. A common practice in Markov chain Monte Carlo (MCMC) algorithms is to run a sufficient number of iterations to ensure that the posterior distributions have converged and stabilized. In this case, most posterior distributions became stable after a few thousand iterations, indicating convergence to the target distributions. Appendix Figure A.21 shows two examples of such sequences.

Following a standard procedure, I discarded the initial half of the MCMC iterations, often referred to as the "burn-in" period, and calculated parameter estimates by averaging the values across the latter half of the iterations. This approach helps minimize potential bias introduced during the initial iterations, ensuring that the parameter estimates are based on samples drawn

from the actual posterior distributions.

In the second step, I calculate the aggregate statistics for the winning bids $\{\tilde{b}_j^f\}_{j:r_j^f \neq \emptyset}$, including the median M_f, the second lowest bid L_f, and the second highest bid H_f.[41] The relative ranking of sellers is preserved if I replace b_j with $a_1(b_j^f)^{a_2}$ and σ_b with $a_2\sigma_b$, in which case $\tilde{b}_j^f$ is replaced with $a_1(\tilde{b}_j^f)^{a_2}$, as well as the statistics M_f, L_f, and H_f. I search for the parameters (a_1, a_2) that yield aggregate statistics $(\overline{M}, \overline{R})$ matching the observed values:

$$\min_{a_1,a_2}(\frac{1}{F}\sum_{f=1}^{F} a_1 M_f^{a_2} - \overline{M})^2 + (\frac{1}{F}\sum_{f=1}^{F}(a_1 H_f^{a_2} - a_1 L_f^{a_2}) - \overline{R})^2. \tag{D.26}$$

After obtaining the estimated values $\hat{a}_1$ and $\hat{a}_2$, I transform all estimated parameters using these values. Specifically, $\hat{b}_j$ is rescaled to $\hat{a}_1(\hat{b}_j)^{\hat{a}_2}$ and σ_b is rescaled to $\hat{a}_2\sigma_b$.

Appendix Figure A.22 summarizes the estimates of the auction model. Panel A depicts the distribution of the winning bids, pooling data from all market-week pairs. Notably, the average winning bid is \$1.24, closely mirroring industry estimates.[42] In Panel B, I plot the parameter that measures uncertainty, σ_b, against the number of unique auction winners. A higher σ indicates greater uncertainty in realized bids, resulting in more unique auction winners. Panel C illustrates the strong alignment between predicted frequencies of winning the auction and observed winning frequencies, with a correlation coefficient of 0.966.

I apply a similar methodology to estimate the distribution of organic ranks. I do not have access to Amazon's proprietary ranking algorithm, so I need to approximate this process. I assume when a consumer enters a keyword, the platform assigns each product a stochastic quality score drawn from a product-specific distribution:

$$\log(\tilde{o}_j) \overset{\text{i.i.d.}}{\sim} N(\log(o_j), \sigma_o^2). \tag{D.27}$$

Products are ranked based on their realized quality scores. This assumption parallels equation (D.22) with the submitted bid b_j replaced with o_j, making the process similar in nature to the determination of auction outcomes. A product with a higher o_j is more likely to be ranked higher in the organic search results. However, there exists a crucial distinction. In the case of auction outcomes, sellers can influence the distribution of their realized bids by strategically choosing their submitted bids. In contrast, organic ranks are exogenous to the model, and sellers cannot

[41] To avoid potential distortions from extreme values, I exclude the minimum and maximum bids from this analysis.
[42] For instance, according to the 2022 Q2 report by Pacvue, the average PPC on Amazon was estimated to be \$1.24.

manipulate them. This distinction reflects the fundamental difference between paid advertising placements and organic search rankings.

I employ the Gibbs sampler to estimate the vector of parameters $\boldsymbol{o} = (o_1, \ldots, o_J)$ using realizations of organic ranks within a given week. I normalize the parameters with $o_1 = 1$ and $\sigma_o = 0.1$. Unlike the bids, the absolute level and dispersion of quality scores do not possess economic significance. Appendix Figure A.23 indicates that the predicted frequency of products appearing in organic positions closely aligns with the observed frequency, with a correlation coefficient of 0.972.

E Details on Estimation of Platform's Weight

E.1 Definition of Platform's Objective Function

This section provides a formal definition of the platform's objective function introduced in Section 5.3.1. Similar to other papers studying multi-sided platforms (e.g., Gutierrez 2022, Castillo 2022), to maximize the long-run shareholder value, the platform can invest in consumer and seller acquisition and retention by incorporating the welfare of these participants in its objective.

The platform selects a commission rate τ to maximize a combination of short-term revenues, consumer surplus, and seller profits. Let $\boldsymbol{p}(\tau)$ and $\boldsymbol{b}(\tau)$ denote the equilibrium prices and bids under commission rate τ, respectively. Consumer surplus under a given product arrangement, Γ is defined as follows

$$CS(\Gamma;\tau) = \frac{1}{\alpha}\sum_{n=1}^{N}(\lambda_n - \lambda_{n+1})\log\left(1 + \sum_{j\in\mathcal{J}_n(\Gamma)}\exp(\delta_j)\right) \tag{E.28}$$

The expected consumer surplus is

$$CS(\tau) = \int CS\Big(\Gamma\big(\boldsymbol{b}(\tau),\boldsymbol{\omega},\Gamma_0\big);\tau\Big)\mathrm{d}F(\boldsymbol{\omega})\mathrm{d}G(\Gamma_0). \tag{E.29}$$

Expected seller profits are given by

$$PS(\tau) = \sum_{j\in\mathcal{J}} s_j^e\big(\boldsymbol{p}(\tau),\boldsymbol{b}(\tau)\big)\big((1-\tau)p_j(\tau) - c_j\big) - ad_j^e\big(\boldsymbol{p}(\tau),\boldsymbol{b}(\tau)\big), \tag{E.30}$$

where $s_j^e(\boldsymbol{p},\boldsymbol{b})$ and $ad_j^e(\boldsymbol{p},\boldsymbol{b})$ represent seller j's expected market share and advertising payments, as defined in equation (10). The platform's short-term revenues consist of commissions and ad-

vertising revenues

$$\begin{aligned} COM(\tau) =& \tau \sum_{j \in \mathcal{J}} s_j^e\big(\boldsymbol{p}(\tau), \boldsymbol{b}(\tau)\big) p_j(\tau), \\ AD(\tau) =& \sum_{j \in \mathcal{J}} ad_j^e\big(\boldsymbol{p}(\tau), \boldsymbol{b}(\tau)\big). \end{aligned} \tag{E.31}$$

The platform chooses the commission rate τ that maximizes the following objective function:

$$\max_{\tau} \; COM(\tau) + AD(\tau, \boldsymbol{n}^S) + \mu\big(CS(\tau) + PS(\tau)\big), \tag{E.32}$$

where $\mu \geq 0$ represents the platform's weight on consumer surplus and seller profits relative to its short-term revenues. All these quantities are aggregated across markets using the market sizes calibrated in Appendix C.2.

E.2 Alternative Set of Sponsored Positions

In the model, I do not assume that the platform optimally selects the set of sponsored positions. The number of possible configurations is enormous, making it implausible for the observed choice to be fully optimal. The layout of sponsored positions on Amazon has exhibited significant variability in recent years, likely driven by Amazon's ongoing experimentation.

To offer additional evidence, I explore small deviations from the existing configuration of sponsored positions. Specifically, I replace a single sponsored position with an organic one and evaluate the resulting change in consumer surplus, seller profits, and platform revenues. As depicted in Appendix Figure A.24, replacing the first or the second sponsored position with an organic one can simultaneously increase consumer surplus, seller profits, and platform revenues. Therefore, regardless of the weight placed on consumers and sellers, the platform would be better off from these deviations, which implies that the current layout is not entirely optimal.

F Model Extensions

F.1 Ad Blindness and Quality of Sponsored Products

Ad blindness is a well-documented phenomenon in which website visitors consciously or unconsciously ignore advertising elements (Benway and Lane 1998). This term originally referred to

banner-like advertisements in areas distinct from the sought-after content. On digital retail platforms, although sponsored products are often intermingled with organic search results, making it more challenging for consumers to disregard them, ad blindness could also occur. Platforms are legally required to differentiate between sponsored and organic results. For instance, on Amazon, sponsored products are clearly labeled as "Sponsored" on the screen.

In the baseline model, I estimate product quality based on consumers' revealed preferences. When consumers are unaware of the distinction between sponsored and organic results (e.g., they overlook the "Sponsored" label or lack an understanding of its meaning) or hold rational expectations regarding the quality of sponsored products, the estimates are valid. However, if consumers irrationally avoid sponsored products, skipping them without considering them at all, it could lead to an underestimation of the quality of these products and potentially bias the results. In this section, I employ two methods to investigate how ad blindness might impact the findings.

F.1.1 A Fixed Fraction of Consumers Skip Sponsored Positions

Let λ_n represent the fraction of consumers considering the product in the n-th *organic* position. However, due to ad blindness or irrational aversion, a fixed fraction $\Omega \in [0, 1]$ of consumers consistently skip sponsored products, resulting in only a fraction $(1 - \Omega)\lambda_n$ of consumers considering the product in the n-th *sponsored* position. Unfortunately, since I do not have microdata on consumers' search behavior or experimental variation in product arrangement or labeling, I cannot directly determine the value of Ω from aggregate demand data. Instead, I explore how different values of Ω might impact the results. Researchers with access to more detailed data could employ a more flexible model to estimate this parameter.

I re-estimate the model, including consumer demand, product quality, seller primitives, and the platform's weight, and conduct counterfactual simulations under different values of Ω, ranging from 0 to 0.6 with increments of 0.1. Appendix Figure A.25 presents the quality estimates for products appearing in each position under different values of Ω. The first panel, with $\Omega = 0$, corresponds to the baseline result and replicates Panel B of Figure 3. As the value of Ω increases, the estimated quality of sponsored products gradually improves. Notably, when Ω exceeds 0.4, sponsored products have better quality than their organic counterparts in similar positions, except for products at the very top.

Appendix Figure A.26 summarizes the welfare implications of removing sponsored positions under various values of Ω. As in Section 7.1, I consider two scenarios: one with a fixed commission

rate and one with endogenous commission rates. There are two types of consumers: "skippers," who consistently skip all sponsored products, and "non-skippers," who consider all products. I calculate consumer surplus separately for these two groups.

Panel A illustrates that removing sponsored positions benefits skippers across all values of Ω. The effect shrinks under endogenous commission rates, primarily due to the higher prices resulting from a higher commission rate. For non-skippers, the impact under a fixed commission rate depends on the value of Ω. When Ω exceeds 50%, the initially positive effect is reversed. It is because, at higher values of Ω, sponsored products are estimated to have better quality, and their removal negatively affects non-skippers. Under endogenous commission rates, for the same reason, the negative effect on non-skippers becomes more pronounced as Ω increases.

When examining the aggregate effect, which weighs the impacts on both skippers and non-skippers based on their respective shares, the overall change in consumer surplus remains positive under a fixed commission rate and negative under endogenous commission rates. In the latter case, the losses experienced by non-skippers outweigh the benefits enjoyed by skippers, leading to a negative overall impact on consumer surplus.

Panel B illustrates the effects on seller profits and platform revenues. As Ω increases, the impact of removing sponsored positions on seller profits becomes more positive under a fixed commission rate. With endogenous commission rates, the negative effect becomes significantly smaller. This change occurs because, at higher values of Ω, the estimate of the platform's weight on consumers and sellers is higher, leading to a smaller increase in the commission rate after sponsored positions are removed.

Furthermore, as Ω increases, the negative impact of removing sponsored positions on the platform's total revenues becomes less pronounced. Under endogenous commission rates, when Ω is very high, removing sponsored product advertising does not significantly reduce the platform's revenues. Therefore, if a substantial portion of consumers consistently skip sponsored products, it is plausible that the platform could benefit from eliminating sponsored positions.

In summary, under the reasonable range of values for Ω considered (i.e., $\Omega \in [0, 0.6]$), all the results remain qualitatively consistent with those in the baseline model.

F.1.2 Predicted Quality Based on Observables

To mitigate the potential biases in estimating the quality of sponsored products resulting from ad blindness, I construct an alternative measure of product quality that relies on objective observ-

ables, which aims to provide a more robust quality measure for sponsored products.

In constructing this measure, I first select products that never appear in a sponsored position within a week and pool the estimated product quality, $\hat{\delta}_{jt}$, for these products across all market-week pairs. Next, I regress $\hat{\delta}_{jt}$ on market fixed effects and four observables: the log of the number of reviews, the average consumer rating, an indicator for Prime eligibility, and an indicator for Amazon's Choice. The signs of the coefficients on these observables align with expectations: products with more reviews and a higher consumer rating tend to have higher quality, and products eligible for Prime and chosen as Amazon's Choice also tend to have higher quality. I then extend this relationship to all products, including those appearing in sponsored positions. The resulting predicted product quality is denoted as $\tilde{\delta}_{jt}$.

In Panel A of Appendix Figure A.27, I compare product quality based on revealed preferences, $\hat{\delta}_{jt}$, with product quality based on observables, $\tilde{\delta}_{jt}$. While these two measures exhibit a moderately high correlation of 0.699, the variation in the latter measure is considerably smaller. Panel B displays the average observable-based quality for all products in each position. Compared to the baseline result in Panel B of Figure 3, the quality difference between organic and sponsored products in similar positions shrinks significantly and is even reversed in the bottom positions. This change suggests that the observable-based quality measure could mitigate potential biases in quality estimates based on consumers' revealed preferences.

When consumers' preferences are biased, it becomes crucial to distinguish between consumers' perceived utility and actual utility (Allcott 2013; Train 2015). The former determines consumer choices, while the latter determines consumer welfare. In this setting, consumers' perceived utility is represented by $\hat{u}_{ijt} = \hat{\delta}_{jt} + \epsilon_{ijt}$, while I assume the actual utility follows $\tilde{u}_{ijt} = \tilde{\delta}_{jt} + \epsilon_{ijt}$. The discrepancy between the two measures may arise from consumers' ad blindness or irrational aversion to sponsored products.

When maximizing profits, sellers are only concerned with consumers' decision utility. Similarly, I assume that when the platform chooses the commission rate, it only cares about consumers' decision utility rather than their actual utility. Therefore, all estimation results except for product quality remain valid. The only distinction arises in how I calculate consumer surplus. Specifically, let s_{jt} denote product j's market share. Consumer surplus is given by

$$CS = \frac{1}{\alpha}\mathbb{E}(\max_j \tilde{u}_{ijt}) = \frac{1}{\alpha}\mathbb{E}(\max_j \hat{u}_{ijt}) + \frac{1}{\alpha}\sum s_{jt} \times (\tilde{\delta}_{jt} - \hat{\delta}_{jt}), \tag{F.33}$$

where the last term captures the welfare losses arising from biased preferences.

Appendix Figure A.28 presents the welfare results. Panel A illustrates consumer surplus in the status quo and counterfactual scenarios, calculated using both the baseline and the observable-based quality measures. Consumer surplus experiences a decline under the observable-based quality measure due to the discrepancy between perceived and actual utility. Panel B depicts the change in consumer surplus following the removal of sponsored positions. Notably, since sponsored products have higher observable-based quality compared to the baseline quality, eliminating sponsored positions leads to a less pronounced positive effect under a fixed commission rate and a more substantial negative impact under endogenous commission rates. Importantly, all findings remain qualitatively robust under this alternative quality measure.

F.2 Endogenous Consideration Sets and Search Costs

In the baseline model, consumers have heterogeneous consideration sets in the form of the top n products in the search results. However, the population share of each consumer type, denoted as $\lambda_n - \lambda_{n+1}$, is treated as exogenous and does not respond endogenously to changes in prices or product arrangements. In this section, I endogenize the formation of consideration sets using a sequential search model. This approach allows λ_n to change in a counterfactual scenario where all sponsored positions are eliminated.

F.2.1 Model

The sequential search model is built upon the framework developed by Lam (2023). In this model, consumers navigate the search results following a fixed order from the top of the page to the bottom, gradually unveiling the product arrangement and incorporating new products into their consideration sets. I want to highlight that this model focuses on a stage of the search process different from the optimal search model developed by Weitzman (1979). In the latter, consumers start with a known utility component for a predetermined set of products. They explore the products in an *unrestricted* order to disclose the remaining utility associated with those options. However, in my model, consumers begin with neither a utility component nor a predefined set of options. Prior to reaching a product, a consumer must pass through all the products located above it, which is a reasonable assumption given the typical search process on a platform.

Before reaching a specific position, consumers do not observe the quality of the product, denoted as δ_j, in that position. Instead, they form expectations regarding a product's quality based

on its location in the search results and whether it is a sponsored result. Let $\tilde{\delta}_n$ represent the expected quality of the product in the n-th position and j_n the realized product appearing in the n-th position. Consumers behave as if they knew product quality perfectly such that $\delta_{j_n} = \tilde{\delta}_n$.

Let $\mathcal{J}_n = \{j_1, ..., j_n\}$ denote the set of the top n products. Consumer i with a consideration set $\mathcal{J}_{n_1}$ will add new products to her consideration set to form a new consideration set $\mathcal{J}_{n_2} \supset \mathcal{J}_{n_1}$ (i.e., $n_2 > n_1$), incurring a search cost of $(n_2 - n_1)s_i$, if and only if the following condition holds:

$$\mathbb{E}_\epsilon[\max\{\delta_{j_n} + \epsilon_{ij_n}\}_{n \le n_1} \cup \{\tilde{\delta}_n + \epsilon_{ij_n}\}_{n_1 < n \le n_2}] - \mathbb{E}_\epsilon[\max\{\delta_{j_n} + \epsilon_{ij_n}\}_{n \le n_1}] > (n_2 - n_1)s_i. \tag{F.34}$$

In this equation, $(n_2 - n_1)$ is the number of steps between two consideration sets, and s_i is consumer i's *per-step* search cost drawn from a distribution F. The left-hand side of equation (F.34) represents the incremental expected utility gained from enlarging the consideration set, while the right-hand side accounts for the total search cost incurred when transitioning from $\mathcal{J}_{n_1}$ to $\mathcal{J}_{n_2}$. The consumer will continue searching if there exists $n_2 > n_1$ for which equation (F.34) holds, i.e., the incremental expected utility surpasses the cost of searching. Otherwise, she will stop her search and maintain her current consideration set.

The search process in this model is dynamic due to the uncertainty in the realizations of δ_{j_n}. A consumer may stop searching if she encounters a high realization of δ_{j_n}, whereas she might continue if she encounters a low realization. The idiosyncratic preference shock, denoted as ϵ_{ij}, follows a type I extreme value distribution and is only realized after the search process.

To determine whether to stop searching, I define a threshold function as follows:

$$s(n_1) = \max_{n_2 > n_1} \frac{\mathbb{E}_\epsilon[\max\{\delta_{j_n} + \epsilon_{ij_n}\}_{n \le n_1} \cup \{\tilde{\delta}_n + \epsilon_{ij_n}\}_{n_1 < n \le n_2}] - \mathbb{E}_\epsilon[\max\{\delta_{j_n} + \epsilon_{ij_n}\}_{n \le n_1}]}{n_2 - n_1}. \tag{F.35}$$

Given a consideration set $\mathcal{J}_n$ formed after reaching the n-th product, consumer i decides to stop searching if and only if her per-step search cost s_i exceeds $s(n)$. Thus, consumers with a per-step search cost $s_i \in \big(s(n), \min_{n'<n} s(n')\big]$ have a consideration set $\mathcal{J}_n$.[43] The fraction of consumers who consider the n-th product, denoted as λ_n, is determined by the following expression:

$$\lambda_n = \sum_{n'=n}^{N} \int \mathbb{1}\Big(s_i \in \big(s(n), \min_{n'<n} s(n')\big]\Big) \mathrm{d}F(s_i). \tag{F.36}$$

[43] The set $\big(s(n), \min_{n'<n} s(n')\big]$ can be empty, suggesting that no consumers have a consideration set being $\mathcal{J}_n$. This can happen, for example, if δ_{j_n} has a very low realization.

Equations (F.35) and (F.36), combined with consumers' expectations regarding product quality $\{\tilde{\delta}_n\}_{n=1}^N$, the realized product quality $\{\delta_{j_n}\}_{n=1}^N$, and the distribution of consumers' per-step search cost F, together determine the population shares of different consideration sets.

A limitation of the model is that I assume sellers and the platform do not respond to changes in consumers' search behaviors in counterfactual scenarios. In reality, sellers and the platform are likely to adapt their strategies based on consumers' behaviors, which can trigger a feedback loop where consumers may further adjust their search behaviors in response. This dynamic interplay creates a complex environment where each player's decision depends on the choices of the others. The equilibrium, thus, represents a "fixed point" in which each player's choice is optimal given other players' choices. However, estimating such a model is challenging and is beyond the scope of this paper. The current model, while simplified, provides valuable insights into the formation of consideration sets and offers potential avenues for future research.

F.2.2 Estimation

The estimation process involves two steps: (1) constructing consumers' expectations regarding product quality in each position $\{\tilde{\delta}_n\}_{n=1}^N$, and (2) estimating the distribution of consumers' per-step search cost F to match observed population shares of different consideration sets.

I assume consumers hold the following expectations regarding product quality

$$\begin{aligned}\tilde{\delta}_n^{O,m} &= \bar{\delta}^{O,m} - \beta^O \log(n),\\ \tilde{\delta}_n^{R,m} &= \bar{\delta}^{R,m} - \beta^R \log(n).\end{aligned} \tag{F.37}$$

In this equation, m indexes markets, O refers to organic positions, and R stands for sponsored positions. The parameters $\bar{\delta}^{O,m}$ and $\bar{\delta}^{R,m}$ indicate the expected quality of the product in the first organic or sponsored position in market m. $\beta^O > 0$ and $\beta^R > 0$ measure the decrease in expected product quality as the rank increases, which I assume to be common across markets. I estimate these relationships and construct consumers' expectations.

I assume consumers' per-step search cost s_i follow a log-normal distribution, i.e., $\log(s_i) \sim N(\mu, \sigma^2)$. Given parameters (μ, σ), I can predict $\tilde{\lambda}_{nf}$ using equations (F.35) and (F.36) for each search indexed by f. I then calculate $\tilde{\lambda}_n$ as the average across all searches within a market. I estimate parameters (μ, σ) to minimize $\sum_{n=1}^N (\tilde{\lambda}_n - \hat{\lambda}_n)^2$, where $\hat{\lambda}_n = 1/\exp(\hat{\beta}(n-1))$ represents the fraction of consumers considering each position implied by the estimates in Section 5.1.1

F.2.3 Results

Panel A of Appendix Figure A.29 plots the distribution of the average per-step search cost across markets. The average cost is converted into dollars and can be calculated as $\exp(\mu + \sigma^2/2)/\alpha$, where α is the price sensitivity in consumers' utility defined in equation (3). The average per-step search cost is concentrated around a few cents, suggesting that moving from the first to the 20th position in the search results incurs an estimated search cost of approximately half a dollar.

I use the estimated distribution of consumers' per-step search cost to predict the proportion of consumers considering each position in three different scenarios:

1. $\tilde{\lambda}_n^0$: the status quo with sponsored positions;
2. $\tilde{\lambda}_n^1$: the scenario with sponsored positions removed under a fixed commission rate;
3. $\tilde{\lambda}_n^2$: the scenario with sponsored positions removed under endogenous commission rates.

For each set of λ_n, the average number of products in consumers' consideration sets can be calculated as $\sum_{n=1}^{N} n(\lambda_n - \lambda_{n+1})$.

Panel B of Appendix Figure A.29 evaluates the model fit by comparing the average numbers of products in consumers' consideration sets under $\hat{\lambda}_n$, the estimates in Section 5.1.2, and under $\tilde{\lambda}_n^0$, the values predicted above. The two numbers closely align across all markets, indicating a good fit. In Panel C. I compare the average numbers of products under $\tilde{\lambda}_n^0$ and $\tilde{\lambda}_n^1$. When sponsored positions are removed, consumers, on average, search for 2.7 products less (30.9 vs. 28.2). The removal of sponsored positions results in consumers encountering high-quality products more quickly, leading them to stop searching sooner. Panel D compares the average numbers of products under $\tilde{\lambda}_n^1$ and $\tilde{\lambda}_n^2$. With an increase in the commission rate, consumers, on average, search for 1.3 products more (28.2 vs. 29.5). The higher prices may force consumers to explore more options before making a purchase decision.

In Appendix Figure A.30, I evaluate the overall effect on consumer surplus with consideration sets endogenized and search costs incorporated. The first block illustrates consumer surplus without accounting for search costs, the second block quantifies the total search costs incurred, and the third block presents the net consumer surplus as the difference between consumer surplus and search costs. The results align qualitatively with the baseline findings: while removing sponsored positions increases net consumer surplus when holding the commission rate fixed, it decreases consumer surplus under endogenous commission rates.

F.3 Market-Specific Commission Rates

On Amazon, commission rates have limited variation across different markets. For most markets in my sample, the commission rate is either 15% or 17%. In this section, I investigate the potential impact of Amazon implementing market-specific commission rates.

For each market, I solve the optimal commission rate that maximizes the platform's objective function, both with and without sponsored positions. As shown in Panel A of Appendix Figure A.31, the optimal commission rate without sponsored positions consistently surpasses the rate when sponsored positions are present, confirming the intuition in the baseline result. In Panel B, I compute the averaged predicted commission rate separately for markets with observed commission rates of 15% and 17%. Remarkably, the average predicted commission rate is higher in the second group of markets and is very close to the actual 17%. In Panel C, the welfare effects of removing sponsored positions under market-specific commission rates are qualitatively the same as the baseline results.

G Robustness Checks

G.1 Parametrization of Consumer Search Frictions

In Section 5.1.2, I parametrize the fraction of consumers whose consideration sets contain the n-th product in search results as $\lambda_n = 1/\exp(\beta(n-1))$ for $1 \leq n \leq N$, where the parameter $\beta \geq 0$ governs the extent of search frictions. I explore two alternative parametrizations: $\lambda_n = 1-\beta(n-1)$ and $\lambda_n = 1-\beta\log(n)$. In both models, the parameter β quantifies the degree of search frictions, ensuring consistency in the estimation process. Appendix Figure A.32 demonstrates that the estimates obtained using these two functional forms are similar to the baseline results.

G.2 Parametrization of Utility Function

In Section 5.1.2, I parametrize consumers' utility function in equation (3), where I include a product-specific linear time trend that captures whether a product is on an upward or downward trajectory over time. Appendix Figure A.33 considers three alternative parametrizations: dropping the linear time trend, substituting it with product-month fixed effects, or incorporating a product-specific quadratic time trend. The estimates are very close to the baseline results.

G.3 Alternative Autoregressive Models

In Section 5.1.2, the identification strategy hinges on the assumption that the unobserved demand shock ξ_{jt} follows an AR(1) process. To explore the robustness of the estimates, I consider a more generalized autoregressive model. Specifically, I assume ξ_{jt} follows an AR(p) process given by

$$\xi_{jt} = \sum_{\tau=1}^{p} \rho_\tau \xi_{j,t-\tau} + \eta_{jt}, \;\; \eta_{jt} \sim \text{i.i.d.}, \;\; \mathbb{E}(\eta_{jt}) = 0, \;\; \eta_{jt} \perp \xi_{j,t-\tau}, \forall \tau \geq 1 \tag{G.38}$$

In the baseline analysis, I set $p = 1$, which leads to equation (5). Panel A of Appendix Figure A.34 illustrates that the estimates of β under alternative values of $p = 2, 3, 4, 5$ closely resemble the estimates derived under the $p = 1$ specification. Panel B presents the estimates of ρ_τ when $p = 5$. Notably, while ρ_1 averages around 0.4, ρ_2 diminishes to less than 0.1 in most markets, and ρ_3 approaches 0. This empirical pattern suggests that the AR(1) model is a reasonable approximation, reinforcing the validity of the baseline findings.

G.4 Alternative Extent of Search Frictions

An important factor in estimating seller primitives and the platform's weight and in conducting counterfactual simulations is the extent of search frictions. This factor influences how a product's position in search results impacts its sales. Given the absence of experimental variation, it is challenging to establish the causal effect of positions on sales. In Section 5.1.2, I parametrize the fraction of consumers whose consideration sets contain the n-th product as $\lambda_n = 1/\exp(\beta(n-1))$. To identify the parameter β, I examine how changes in a product's daily positions affect its sales while accommodating demand persistence. This identification approach relies on two assumptions: (i) unobserved demand shocks on the platform follow an AR(1) process, and (ii) organic ranks move more slowly than demand shocks. To address concerns regarding the sensitivity of the welfare results to this parameter estimate, I conduct a robustness check to evaluate the impact of varying values of β on the welfare outcomes.

In this robustness exercise, I consider six values of $\beta = 0.015, 0.02, 0.025, 0.03, 0.035, 0.04$,[44] and I assume all markets share the same β. It is noteworthy that these values offer a range that encompasses various alternative estimates for β. First, Ursu (2018) rely on experimental variation to identify the impact of ranks on consumer search. The results depicted in Figure 2 of Ursu (2018)

[44] Appendix Figure A.11 provides examples of this function under different values of β. In the demand estimation, the average $\hat{\beta}$ across markets is 0.027.

suggest a β value of 0.021. Second, Lam (2023) adopts a similar identification strategy as in my paper to estimate consumers' search behavior in the "Home & Kitchen" category on Amazon. The results illustrated in Figure 11 of Lam (2023) imply a β value of 0.04. Third, Column 6 of Table 2 in my paper presents a reduced-form estimate that corresponds to a β value of 0.016.

Under each value of β, I re-estimate the model, including product quality, seller primitives, and the platform's weight, and conduct counterfactual simulations. Appendix Figure A.35 presents the quality estimates for products appearing in each position under different values of β. When β is small, the slope between quality and position is notably steep, possibly reflecting the mechanical correlation between positions and quality when search frictions are not adequately accounted for. As β increases, the slope becomes flatter.

Appendix Figure A.36 presents the welfare implications of removing sponsored positions. Under a fixed commission rate, the impact on consumers and sellers is more sensitive to the degree of search frictions. When search frictions are relatively high (i.e., β is larger), removing sponsored positions is more beneficial for consumers and sellers. Higher search frictions amplify the welfare losses of low-quality sponsored products displacing high-quality organic alternatives. Therefore, eliminating sponsored positions would result in more substantial welfare gains.

Under endogenous commission rates, the welfare effects are more consistent across varying values of β. The intuition is that when search frictions are high, advertising revenues play a more prominent role in the platform's total revenues. Consequently, I estimate that the platform places a lower weight on consumers and sellers in those cases. When sponsored positions are removed, the platform will raise its commission rate by a slightly larger magnitude. This force offsets the larger benefits brought by high-quality organic options, resulting in relatively stable overall effects.

The findings of this sensitivity analysis underscore that, although the magnitude of the welfare effects may exhibit variations with different values of β, the qualitative conclusions remain robust across reasonable parameter ranges.

G.5 Distributional Assumption of Realized Bids

In the baseline model, I assume that when a seller submits a bid b_j, the realized bid follows a log-normal distribution: $\log(\tilde{b}_j) \overset{\text{i.i.d.}}{\sim} N(\log(b_j), \sigma_b^2)$, where σ_b measures the magnitude of uncertainty in realized bids. I explore an alternative distributional assumption where the realized bid follows a normal distribution: $\tilde{b}_j \overset{\text{i.i.d.}}{\sim} N(b_j, \sigma_b^2)$. Appendix Figure A.37 presents the estimates of the auction model under this assumption. Compared to the baseline estimates in Appendix Figure A.22, this

alternative model exhibits a slightly worse fit to the data: the correlation between predicted and observed frequencies of winning the auction drops from 0.966 to 0.916.

G.6 Assumption on Duplicated Listings

In the baseline analysis, I assume the second appearance of the same product in search results does not receive an independent idiosyncratic preference shock when calculating consumer surplus. Under this assumption, duplicated listings resulting from a product appearing in both sponsored and organic positions reduce product variety, a source of inefficiency associated with sponsored product advertising. I explore an alternative assumption where the second appearance receives its independent preference shock. Appendix Table A.3 and Appendix Figure A.38 present the welfare outcomes under this alternative assumption, confirming the qualitative robustness of my findings. As this duplication is no longer a source of inefficiency, the impact of removing sponsored positions on consumer surplus differs in magnitude, resulting in a smaller positive effect under a fixed commission rate and a larger negative impact under endogenous commission rates.

H Additional Figures and Tables

References

Allcott, H. (2013). The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market. *American Economic Journal: Economic Policy*, 5(3):30–66.

Benway, J. P. and Lane, D. M. (1998). Banner Blindness: Web Searchers Often Miss "Obvious" Links. *Itg Newsletter*, 1(3):1–22.

Brynjolfsson, E., Hu, Y., and Simester, D. (2011). Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. *Management Science*, 57(8):1373–1386.

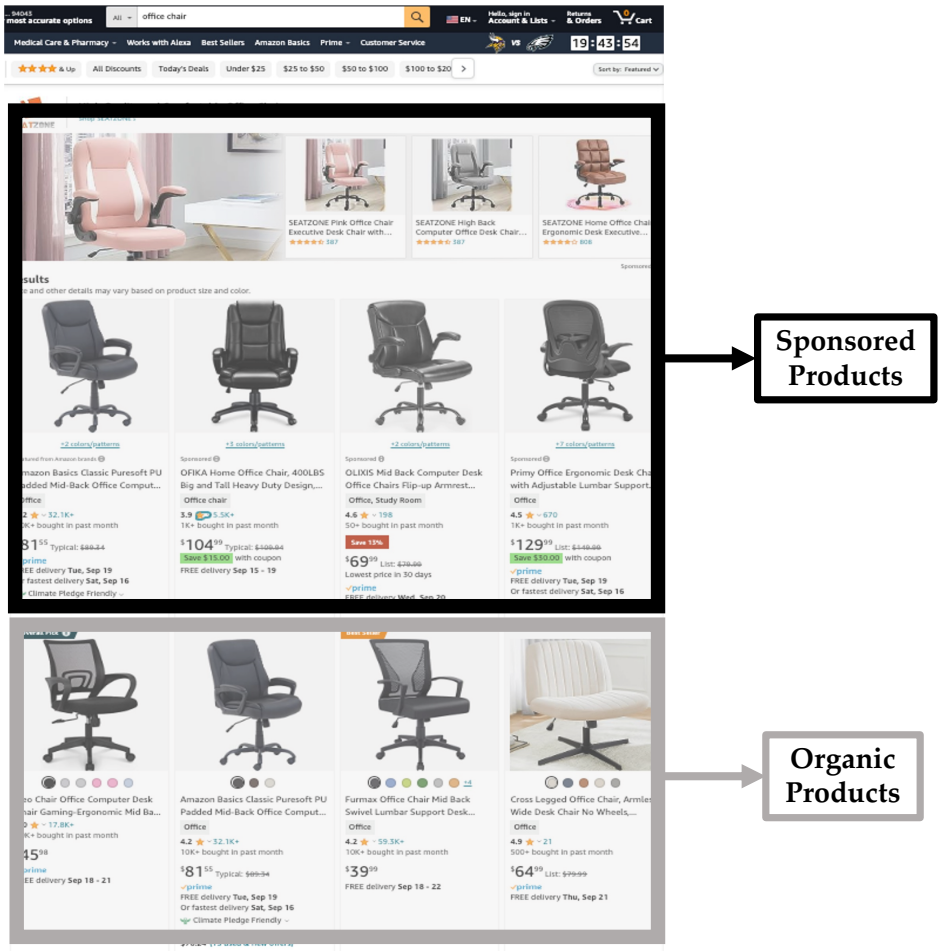
Chevalier, J. and Goolsbee, A. (2003). Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com. *Quantitative Marketing and Economics*, 1:203–222.

Train, K. (2015). Welfare Calculations in Discrete Choice Models When Anticipated and Experienced Attributes Differ: A Guide with Examples. *Journal of Choice Modelling*, 16:15–22.

Weitzman, M. L. (1979). Optimal Search for the Best Alternative. *Econometrica*, 47(3):641–654.

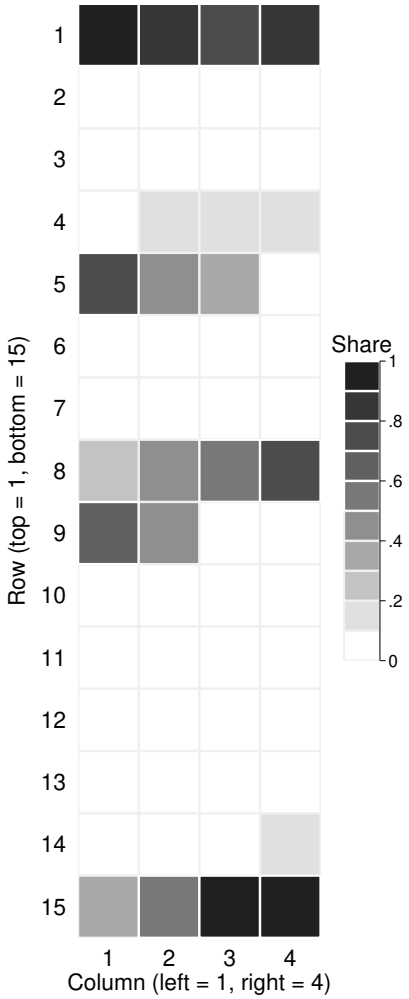
Yin, L. and Jeffries, A. (2021). How We Analyzed Amazon's Treatment of Its "Brands" in Search Results. https://themarkup.org/amazons-advantage/2021/10/14/how-we-analyzed-amazons-treatment-of-its-brands-in-search-results. Accessed in April 2023.

Figure A.1: Example of Amazon Search Results



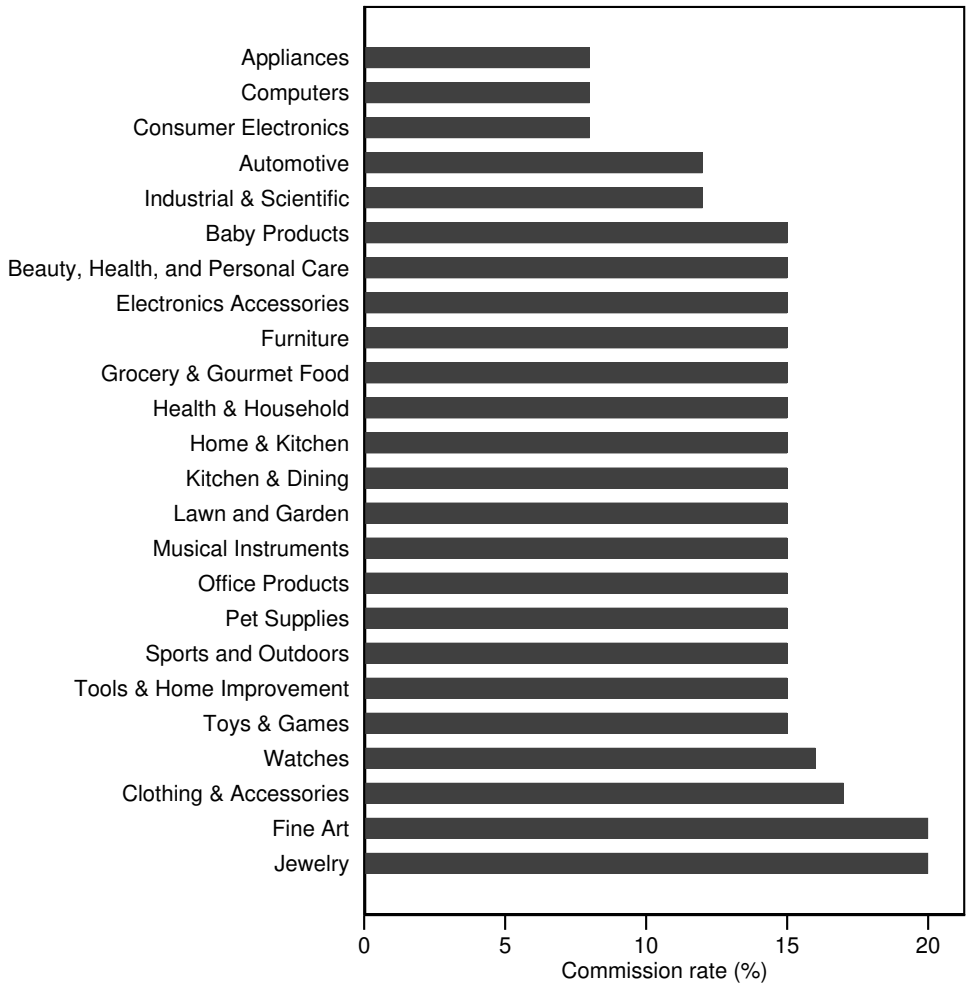
Notes: This figure provides an example of the search results on Amazon. Sponsored products are determined through ad auctions, where sellers are ranked according to their submitted bids. Sellers need to pay the platform to have their products displayed as sponsored products. Organic products are arranged by Amazon's proprietary ranking algorithm, which takes into account various factors such as relevance to the searched keyword, sales performance, and product availability.

Figure A.2: Distribution of Sponsored Positions in Search Results



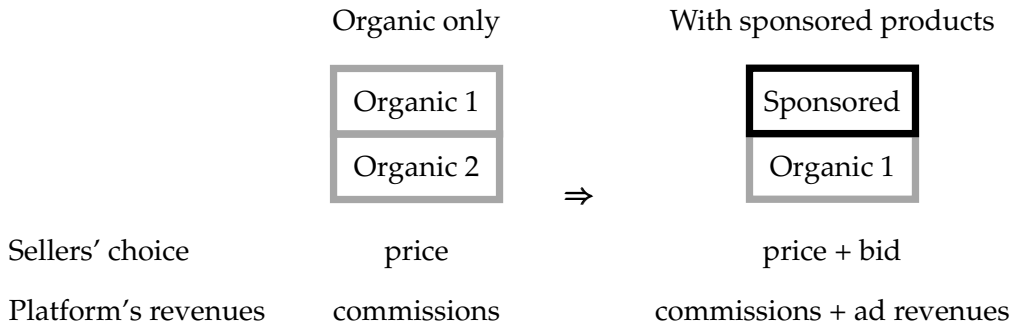
Notes: This figure visually represents the distribution of sponsored positions in search results that encompass a total of 60 positions on the first page, typically organized in 15 rows and 4 columns when viewed on a desktop or laptop browser. Each block corresponds to one position. For each of these positions, I calculate the fraction of searches where a sponsored product appears in that position. A fraction of 1 indicates that the position always displays a sponsored product, while a fraction of 0 indicates that the position always displays an organic product.

Figure A.3: Amazon's Commission Rates Across Product Categories



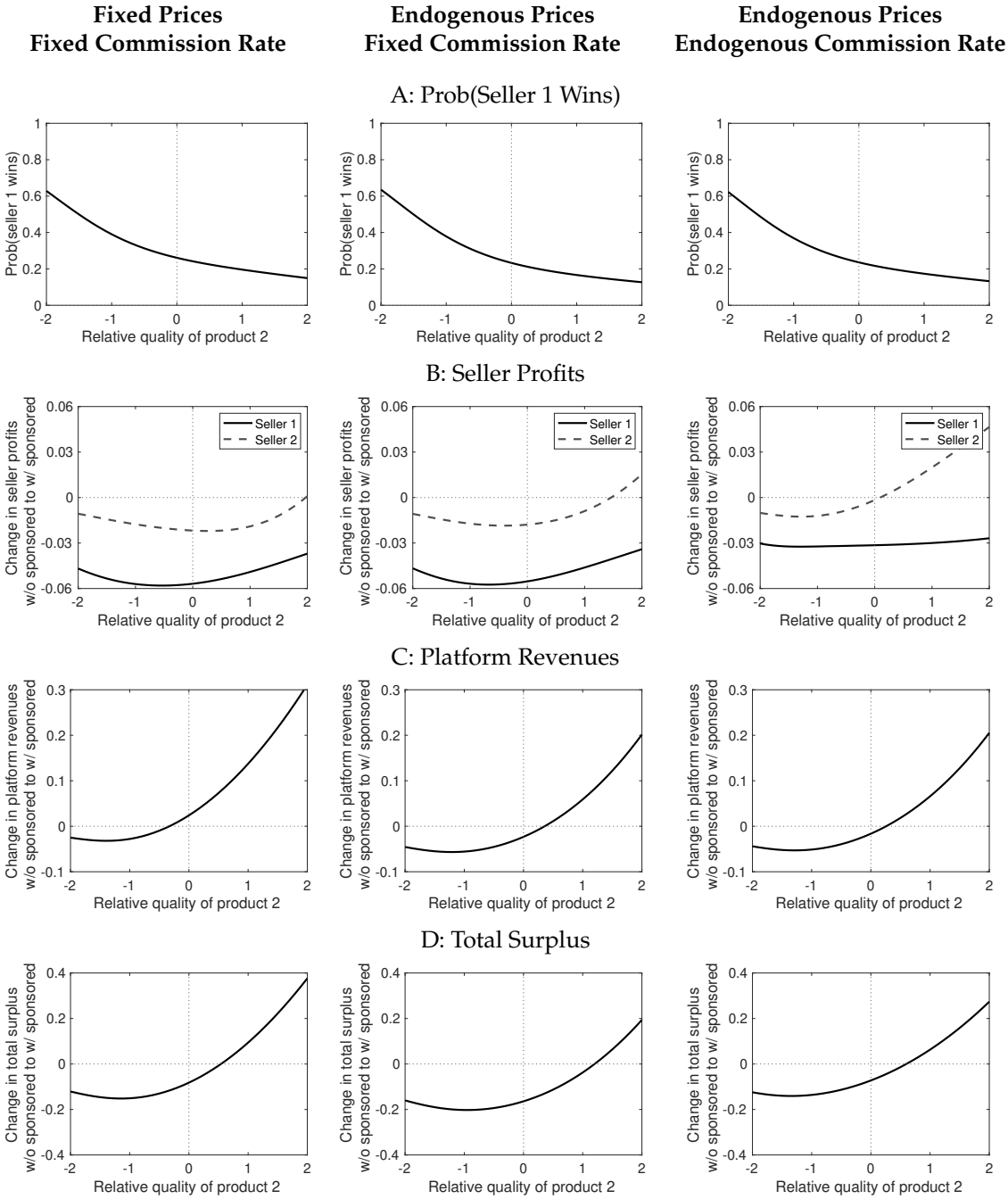
Notes: This figure presents the commission rates set by Amazon across various product categories. The data was collected from Amazon's disclosure, accessed in May 2022. For each unit sold on Amazon's marketplace, Amazon collects a commission fee from the seller equal to a percentage of the selling price.

Figure A.4: Setup of the Stylized Example



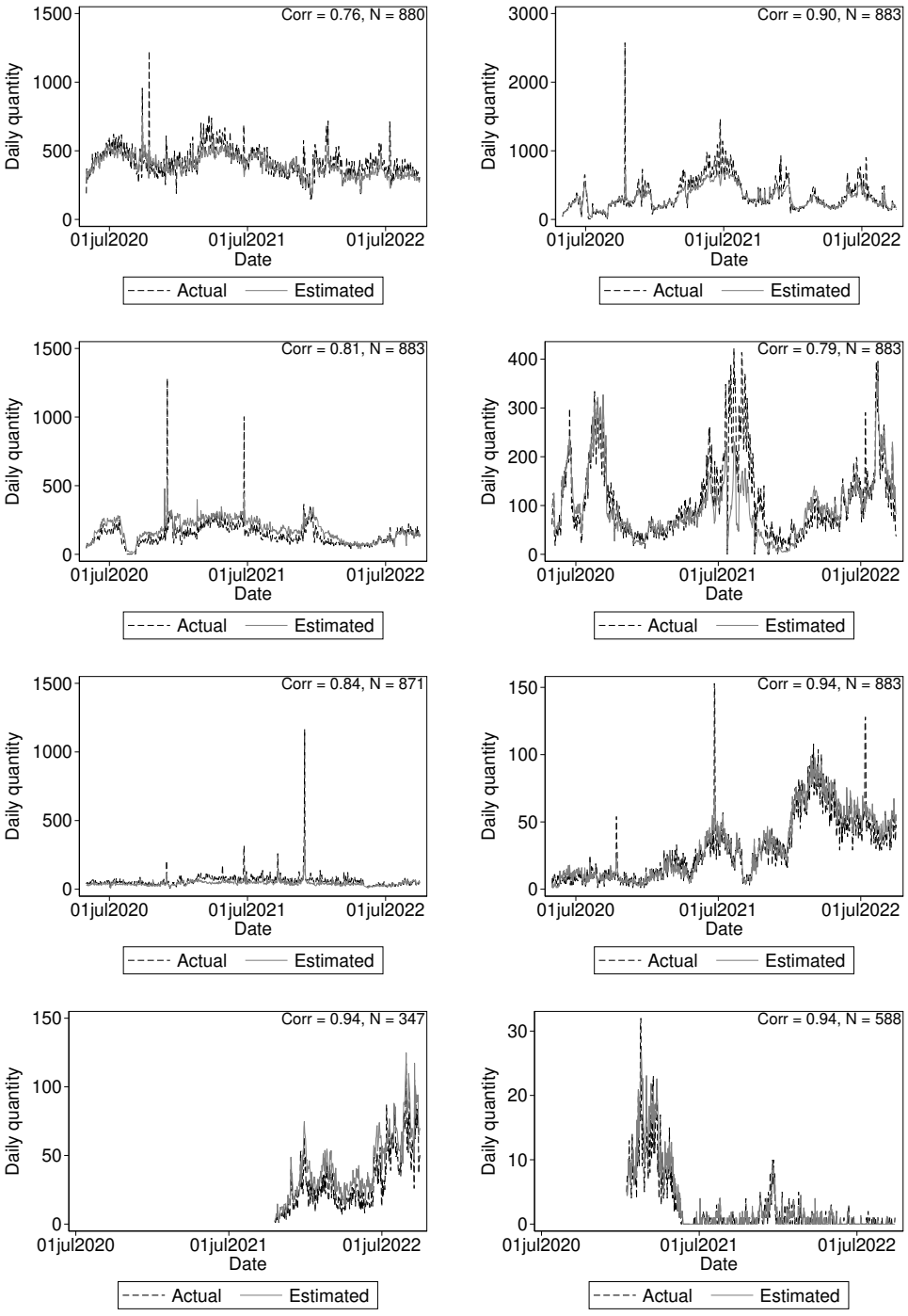
Notes: This figure provides a visual representation of the stylized example's setup in Section 3.1. In the scenario with only organic results, both positions in search results are determined based on organic ranks. Each seller sets a price to maximize its profits. The platform chooses a commission rate to maximize its commissions. In the scenario with sponsored products, the first position is allocated through an ad auction, and the second position displays the top-ranked organic product. Each seller sets a price and submits a bid to maximize its profits. The platform chooses a commission rate to maximize the sum of commissions and advertising revenues.

Figure A.5: Additional Simulation Results of the Stylized Example



Notes: This figure presents additional simulation results of the stylized example. For each panel, the y-axis (except for Panel A) shows the change in outcomes from the equilibrium with two ranked organic positions and no sponsored positions, to the equilibrium with a sponsored position at the top and an organic position at the bottom. The x-axis represents the relative quality of product 2 compared to product 1. Each column corresponds to a different assumption regarding how sellers and the platform can respond to the introduction of a sponsored position. See the notes of Figure 1 for these assumptions. Panel A displays the probability that seller 2 wins the sponsored position. Panels B to D display the changes in the profits of both sellers, the platform's total revenues, and total surplus, respectively. See Section 3 for the setup of the stylized example.

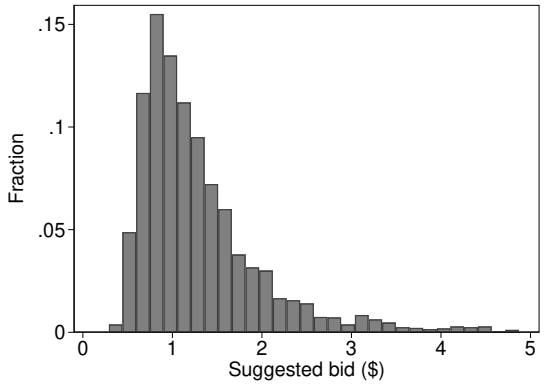
Figure A.6: Comparison of Estimated and Actual Daily Sales Quantity



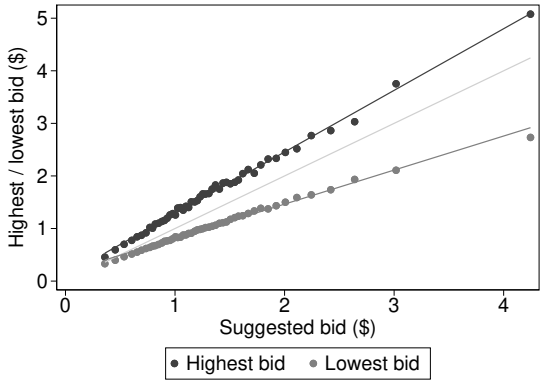
Notes: This figure compares estimated and actual daily sales quantities for eight products sold on Amazon. I obtained actual daily sales quantities over two years for more than one hundred products sold on Amazon in collaboration with a large Amazon seller. The figure displays the actual and estimated daily sales quantities for eight randomly chosen products with varying sales volumes. The estimation is based on Best Seller Ranks using the method detailed in Appendix B.2. Each panel includes the correlation coefficient between actual and estimated daily sales quantities. Appendix Figure A.19 presents the distribution of correlation coefficients across these products.

Figure A.7: Aggregate Statistics of Winning Bids on Amazon

A: Distribution of the Median Winning Bids

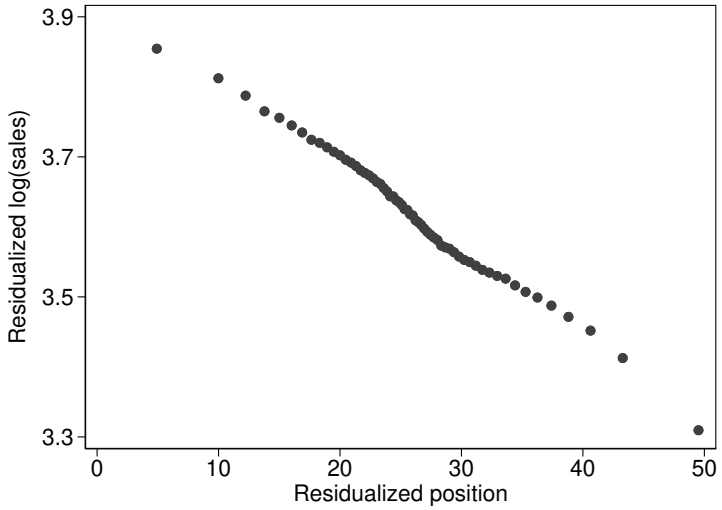


B: Relationships Between Median, Lowest, and Highest Winning Bids



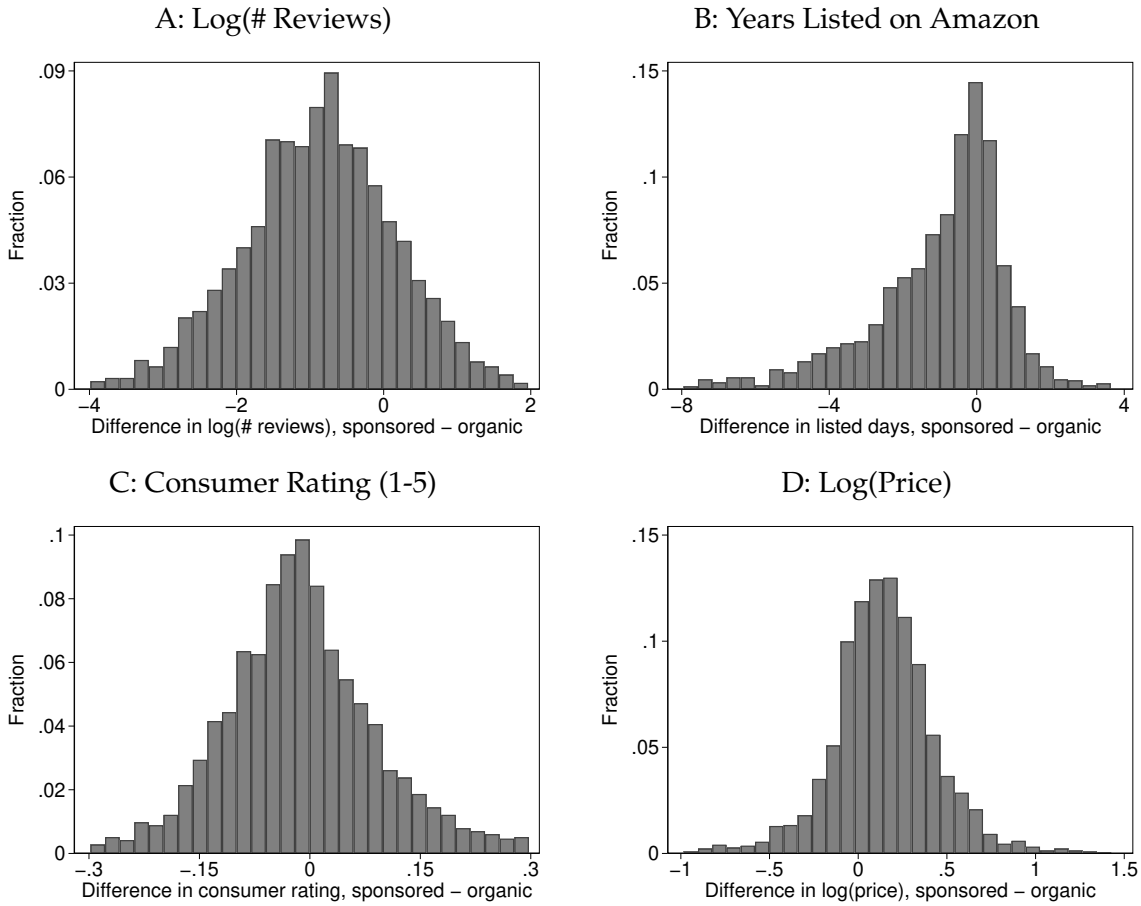
Notes: This figure summarizes the median, lowest, and highest bids of recent auction winners reported by Amazon for each keyword. Panel A plots the distribution of median winning bids across all the keywords in the sample. Panel B presents binned scatter plots between the highest or lowest winning bids and the median winning bid.

Figure A.8: Relationship Between Positions and Sales



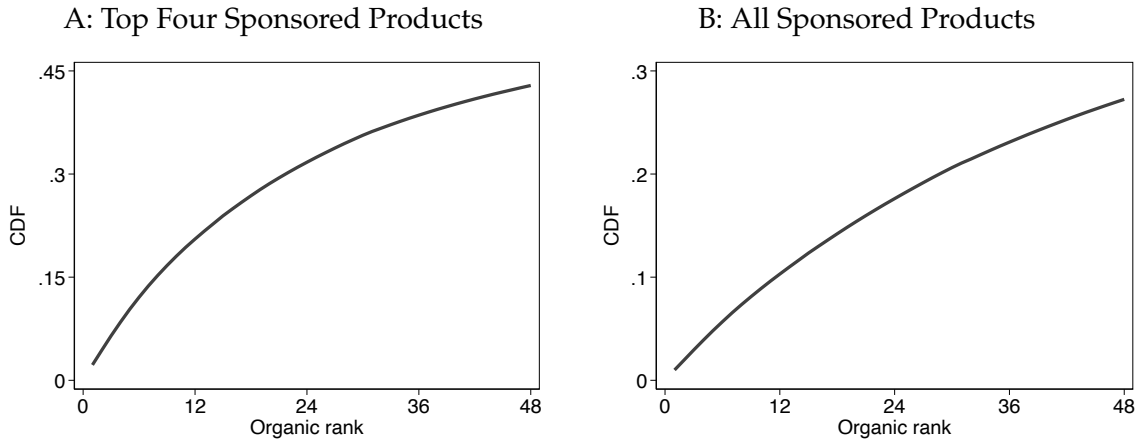
Notes: This figure illustrates the relationship between a product's position in the search results and its sales. Each observation corresponds to a product on a specific day. I first calculate the average position of the product in the search results for a given keyword on a particular day, where lower values indicate higher positions in search results. In situations where a product appears in the search results for multiple keywords, I only retain the keyword where this product appears with the highest frequency, which ensures that each product is associated with a unique keyword. The sample includes products that appear in the search results for a keyword in more than 50% of the searches. I then residualize the average position and the log of the product's total sales quantity by product fixed effects and keyword-day fixed effects, recenter them to their respective sample means, and plot a binned scatter plot of the two residualized variables.

Figure A.9: Distribution of Differences Between Top Sponsored and Organic Products



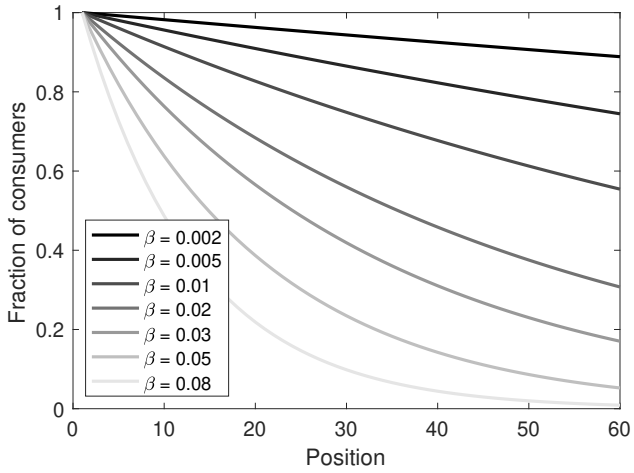
Notes: This figure displays the distribution of the differences between the top four sponsored and four organic products across all the keywords in the sample. I estimate the coefficients from equation (2) separately for each keyword following the same procedure used to construct Figure 2. In Panels A to D, the variables of interest are the log of the number of reviews, total years listed on Amazon, the average consumer rating on a scale of 1 to 5, and the log of the product price, respectively.

Figure A.10: Overlap Between Sponsored and Organic Products



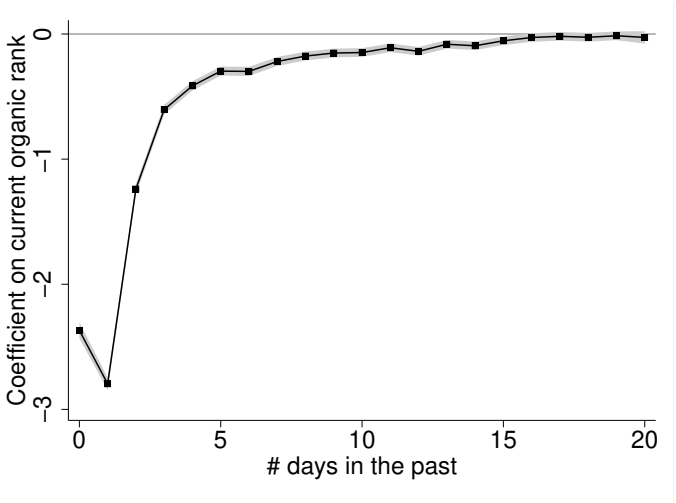
Notes: This figure examines the degree of overlap between sponsored and organic products in Amazon's search results. For each sponsored product, I identify whether it appears in an organic position on the same result page and, if so, its organic rank. This figure plots the cumulative distribution function of organic ranks for sponsored products. Panel A focuses on the top four sponsored products in each search. Panel B examines all sponsored products on the first page.

Figure A.11: Examples of Parametrization in Consumers' Consideration Sets



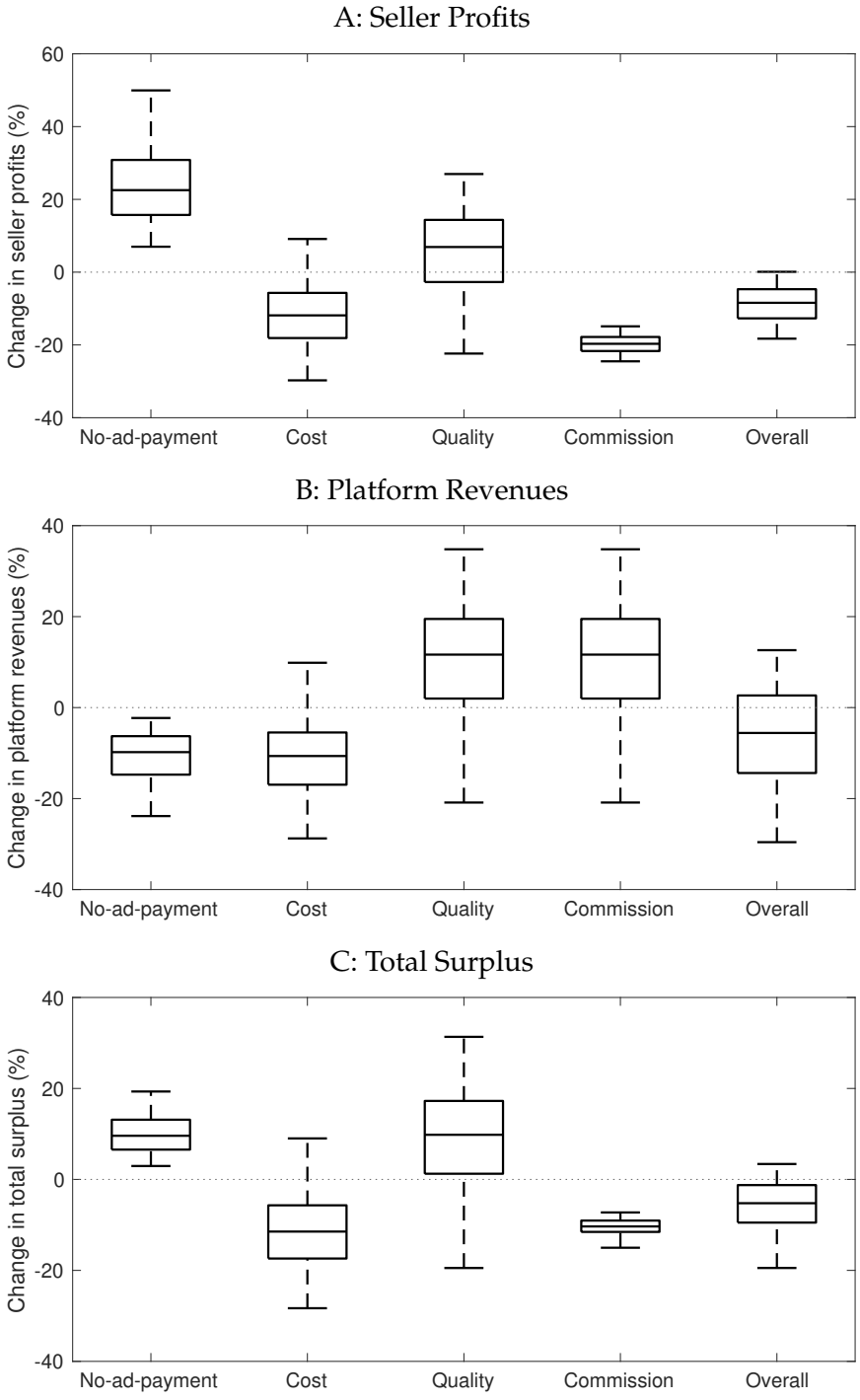
Notes: This figure illustrates the parametrization used to define the fraction of consumers whose consideration sets contain each position in the search results, as discussed in Section 5.1.2. This parametrization is denoted as $\lambda_n = 1/\exp(\beta(n-1))$ for $1 \leq n \leq N$. Different lines in this figure represent this function under different values of the parameter β.

Figure A.12: Effects of Past Sales Performance on Current Organic Rank



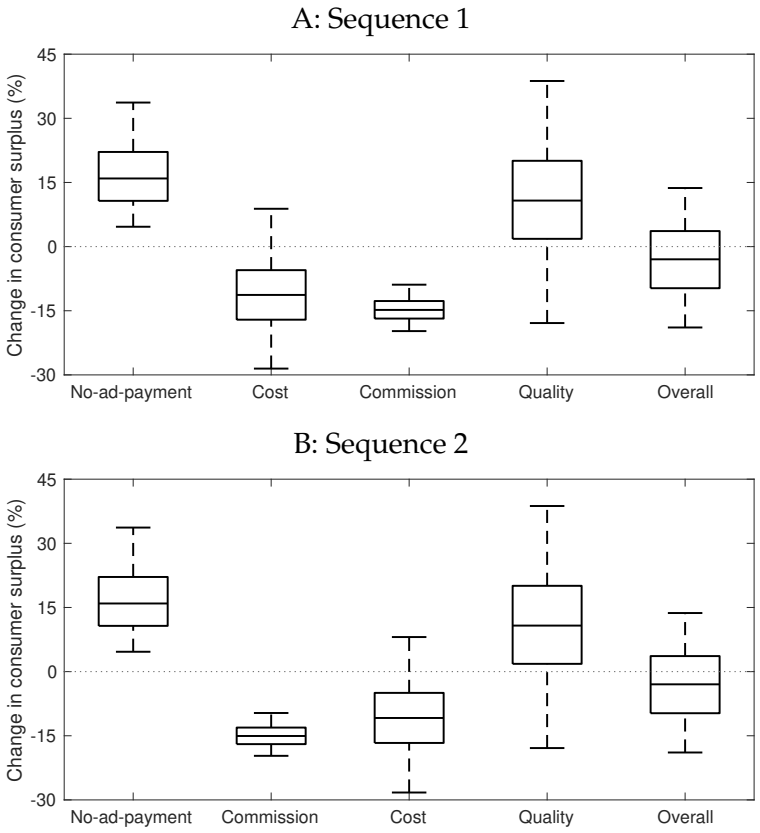
Notes: This figure illustrates the effects of a product's current and past sales performance on its organic rank. Each observation corresponds to a product on a specific day. I calculate the average organic rank of the product in the search results for a given keyword on a particular day, where lower values indicate higher positions in search results. In situations where a product appears in the search results for multiple keywords, I only retain the keyword where this product appears with the highest frequency, which ensures that each product is associated with a unique keyword. The sample includes products that appear in the search results for a keyword in more than 50% of the searches. I regress the average organic rank on the current and lagged log of sales quantities, while controlling for product fixed effects and keyword-day fixed effects. This figure plots the coefficients on the current and lagged sales quantities along with 95% confidence intervals. Standard errors are clustered at the product level.

Figure A.13: Decomposition of Effects of Removing Sponsored Product Advertising



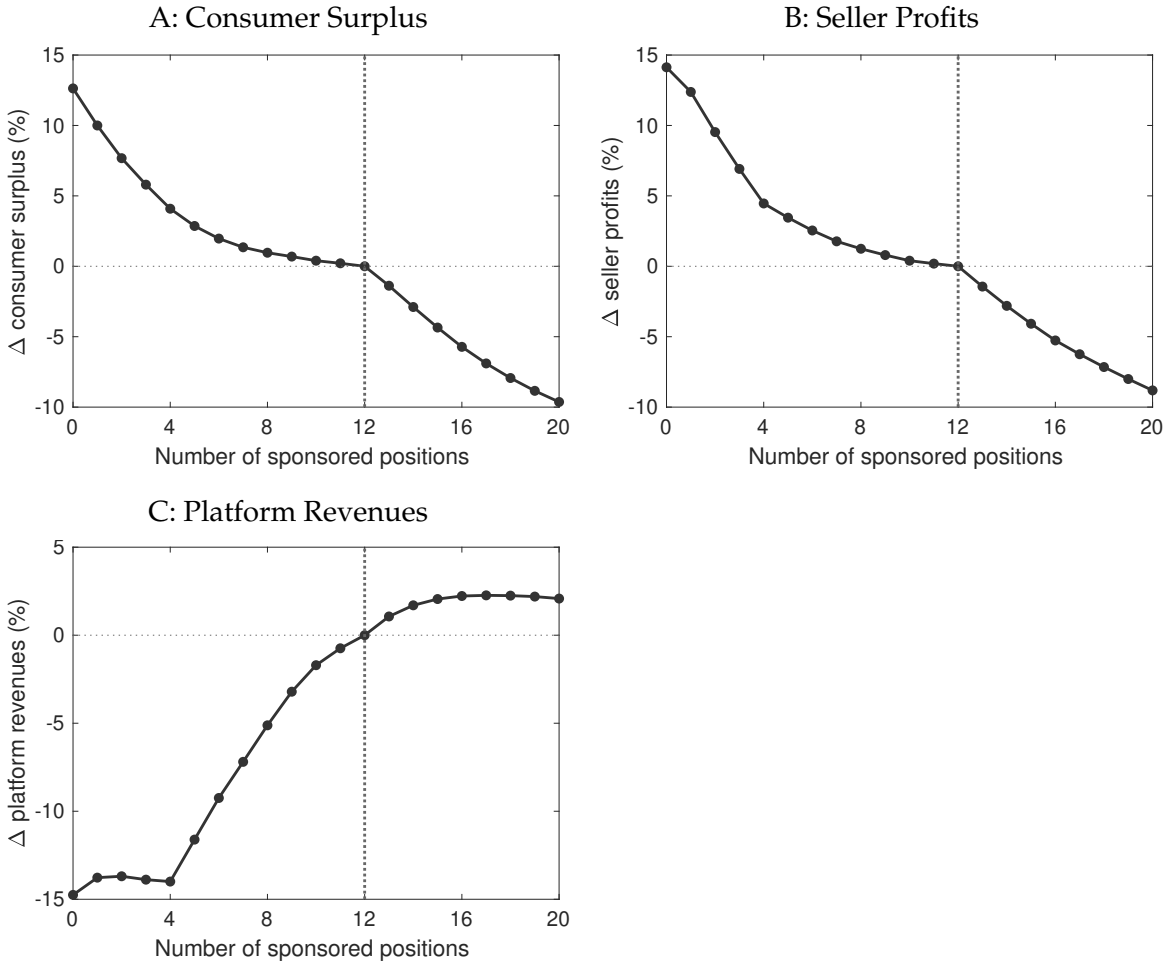
Notes: This figure decomposes the aggregate effects of removing sponsored product advertising on seller profits (Panel A), platform revenues (Panel B), and total surplus (Panel C). See Section 7.2 and the notes of Figure 7 for the definitions of these components. Each box presents the distribution of one of these four effects and the overall effect across market-week pairs, with the middle line indicating the median, the edges of a box indicating the 25th and 75th percentiles, and the outer lines indicating the 5th and 95th percentiles. Figure 7 plots the decomposition of the aggregate effect of removing sponsored product advertising on consumer surplus.

Figure A.14: Alternative Sequences for Decomposing the Effects on Consumer Surplus



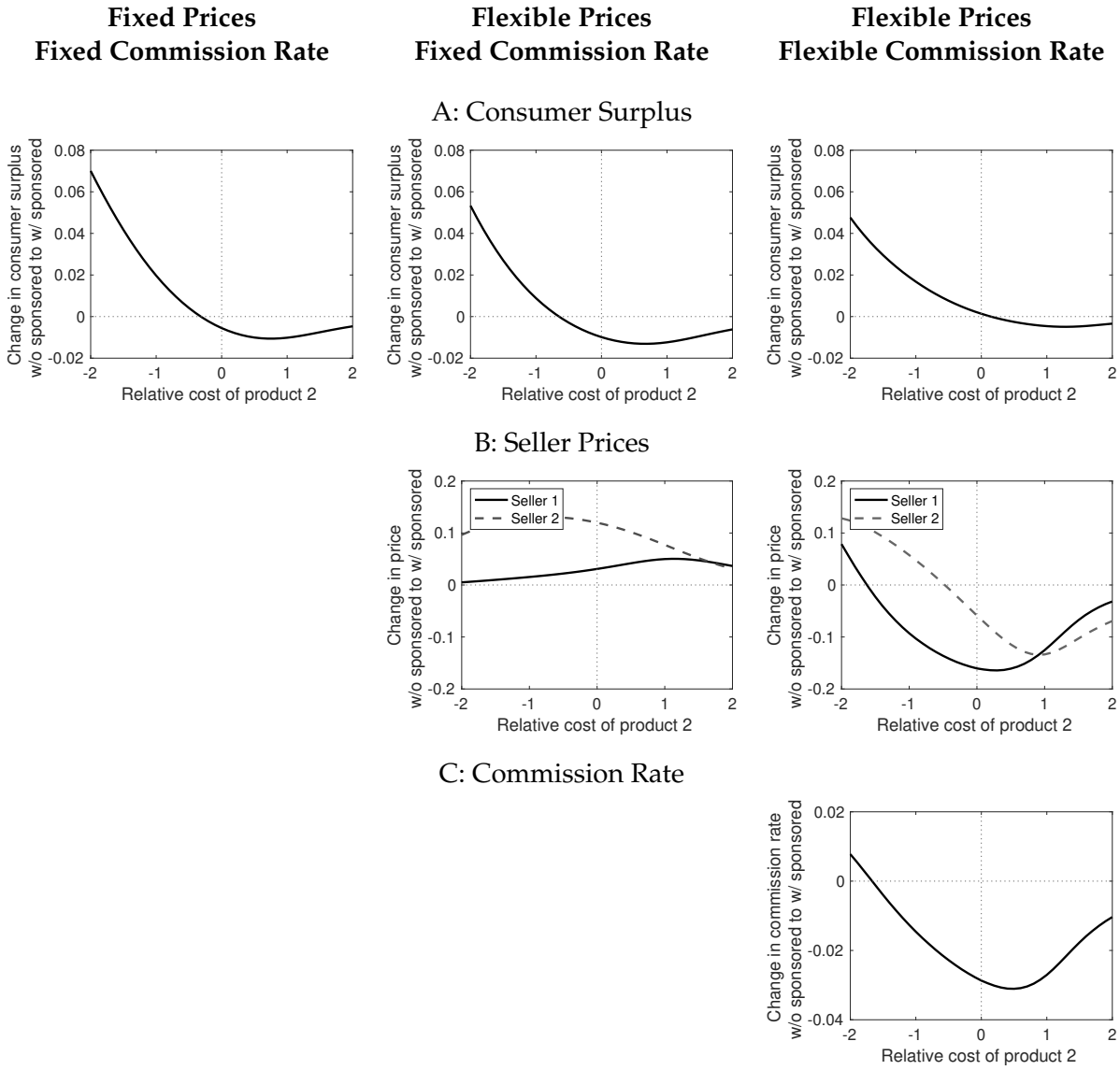
Notes: This figure illustrates alternative sequences for decomposing the aggregate effect of removing sponsored product advertising on consumer surplus, in addition to the baseline sequence presented in Figure 7. Each panel presents a different sequence for the decomposition. Please refer to Section 7.2 for definitions of these components. In each box plot, the middle line represents the median, the edges of the box represent the 25th and 75th percentiles, and the outer lines represent the 5th and 95th percentiles.

Figure A.15: Welfare Effects Under Varying # of Sponsored Positions, Fixed Commission



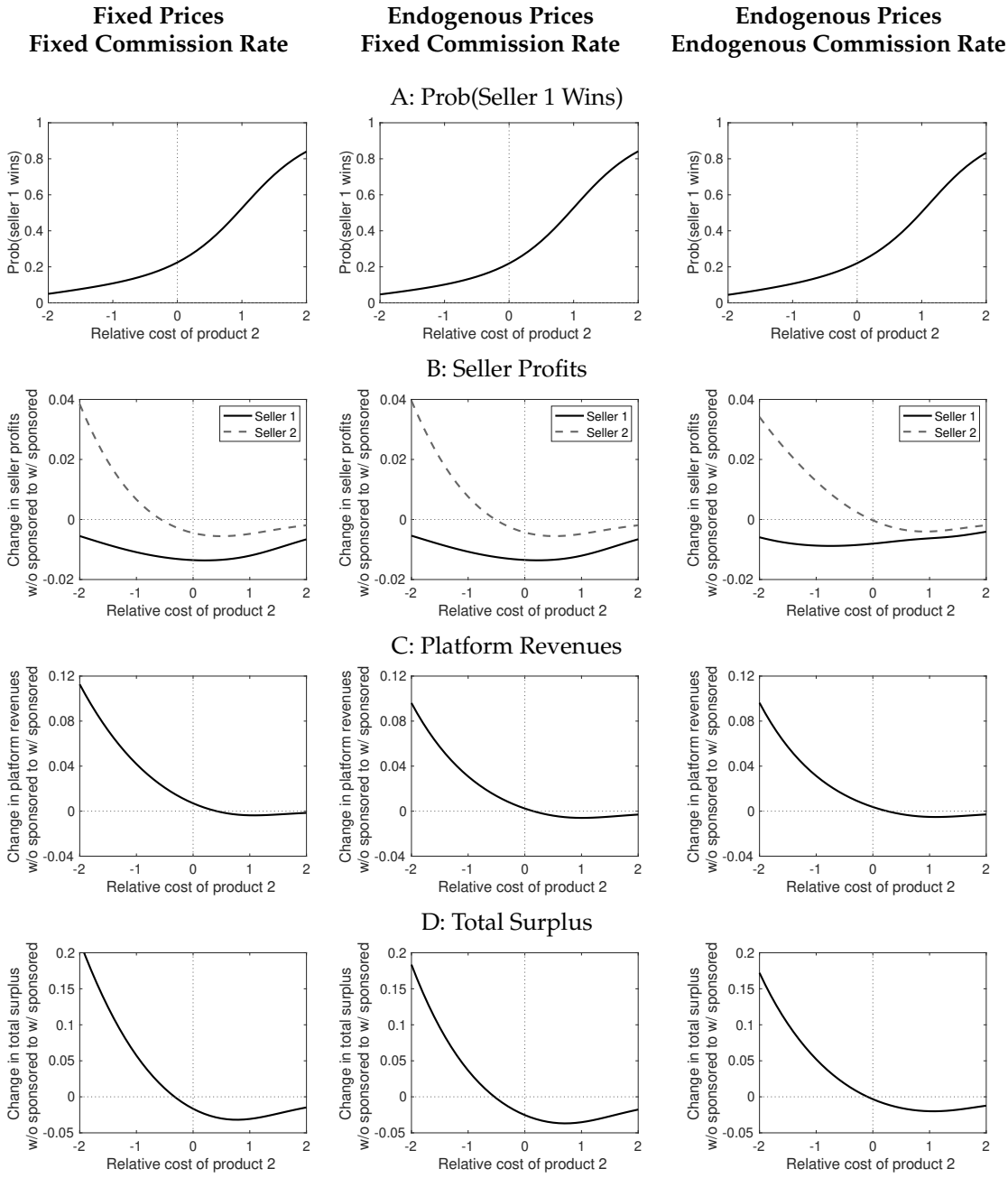
Notes: This figure examines the welfare effects of altering the number of sponsored positions in search results under a fixed commission rate. I change the set of sponsored positions in two directions: (i) a gradual removal of sponsored positions, starting from the bottom and progressing to the top; and (ii) a gradual addition of more sponsored positions in the middle, starting by replacing the ninth organic position and moving downward. Panels A to C depict the aggregate changes in consumer surplus, seller profits, and platform revenues relative to the status quo with 12 sponsored positions. Figure 9 presents the results under endogenous commission rates.

Figure A.16: Simulation Results of the Stylized Example, Varying Cost



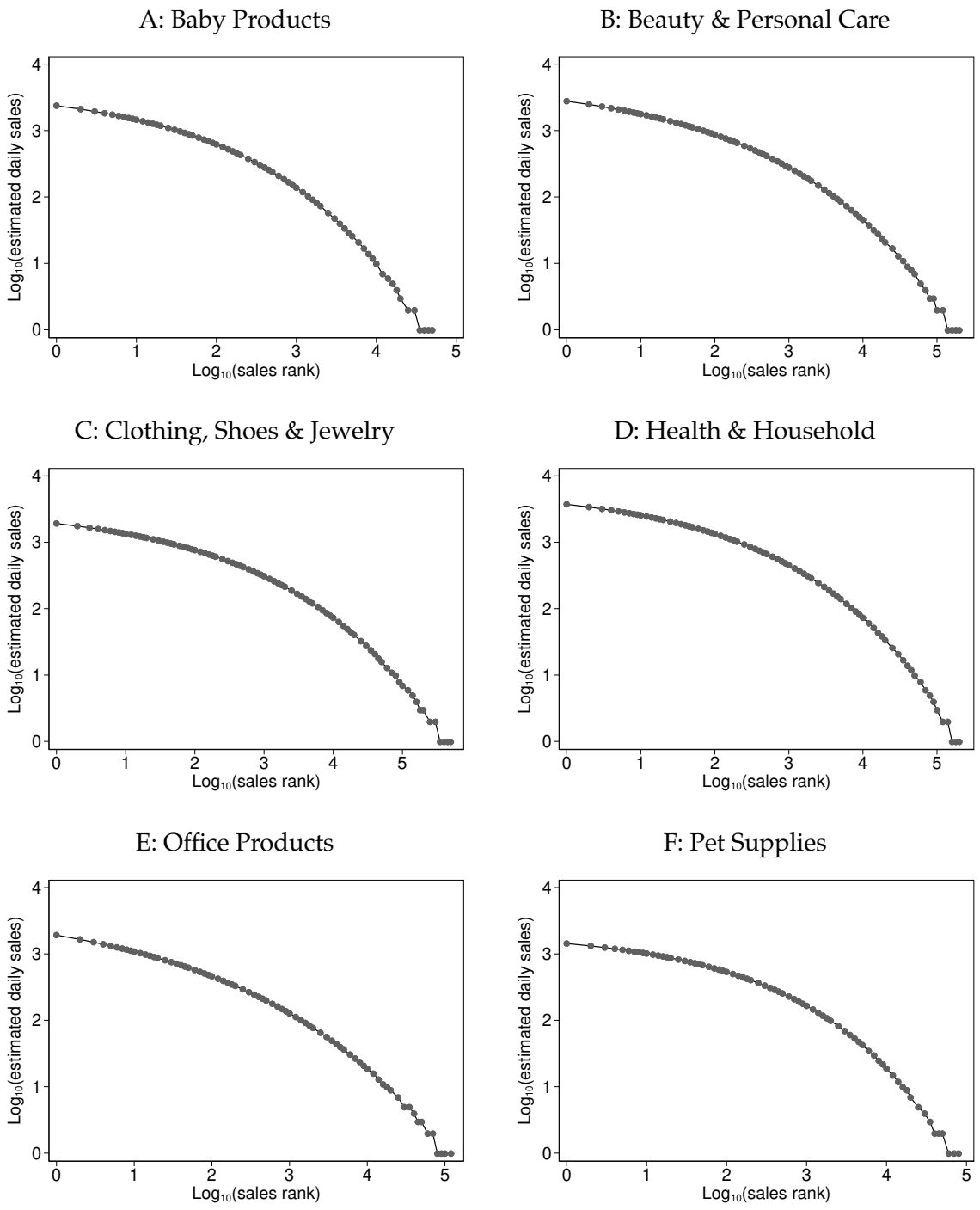
Notes: This figure presents the simulation results of the stylized example. In contrast to the baseline simulations, both products have the same quality, and I hold the marginal cost of product 1 constant while varying the marginal cost of product 2. For each panel, the y-axis shows the change in outcomes from the equilibrium with two ranked organic positions and no sponsored positions, to the equilibrium with a sponsored position at the top and an organic position at the bottom. The x-axis represents the relative marginal cost of product 2 compared to product 1. Each column corresponds to a different assumption regarding how sellers and the platform can respond to the introduction of a sponsored position. See the notes of Figure 1 for these assumptions. Panels A to C display the changes in consumer surplus, the prices of both products, and the commission rate, respectively. See Section 3 for the setup of the stylized example and Appendix Figure A.17 for additional simulation results.

Figure A.17: Additional Simulation Results of the Stylized Example, Varying Cost



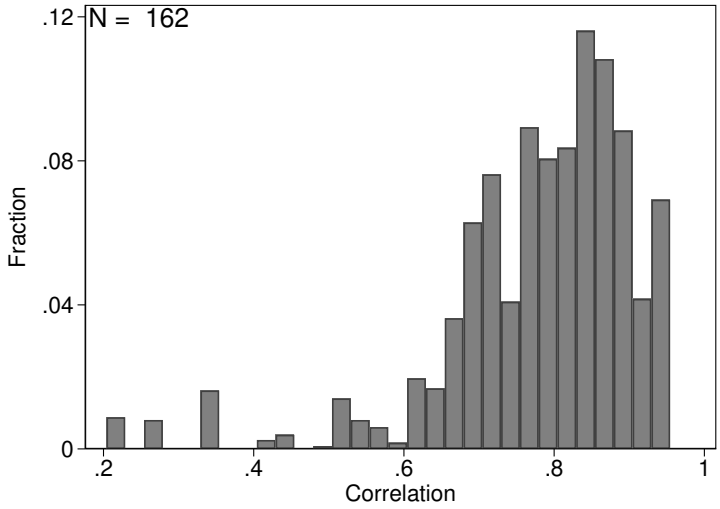
Notes: This figure presents additional simulation results of the stylized example. In contrast to the baseline simulations, both products have the same quality, and I hold the marginal cost of product 1 constant while varying the marginal cost of product 2. For each panel, the y-axis (except for Panel A) shows the change in outcomes from the equilibrium with two ranked organic positions and no sponsored positions, to the equilibrium with a sponsored position at the top and an organic position at the bottom. The x-axis represents the relative marginal cost of product 2 compared to product 1. Each column corresponds to a different assumption regarding how sellers and the platform can respond to the introduction of a sponsored position. See the notes of Figure 1 for these assumptions. Panel A displays the probability that seller 2 wins the sponsored position. Panels B to D display the changes in the profits of both sellers, the platform's total revenues, and total surplus, respectively. See Section 3 for the setup of the stylized example.

Figure A.18: Mapping Between Best Sellers Rank and Sales Quantity



Notes: This figure depicts the mappings from Best Sellers Rank (BSR) to sales quantity for six product categories on Amazon. These mappings were constructed by Jungle Scout for each product category using actual sales data obtained from Amazon sellers. For each category, I selected a sample of BSR values with roughly equal spacing on a logarithmic scale (represented by the dots), up to the point where the estimated sales quantity is 0. I then collected the estimated quantities corresponding to these BSR values and utilized a log-log interpolation to impute sales quantities for all possible BSR values within each category (represented by the connecting line). See Appendix B.2 for additional details.

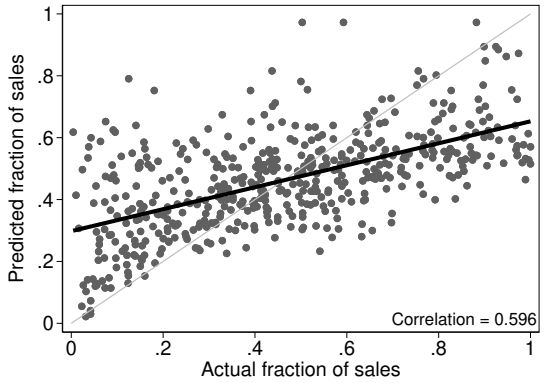
Figure A.19: Correlation Between Estimated and Actual Daily Sales Quantity



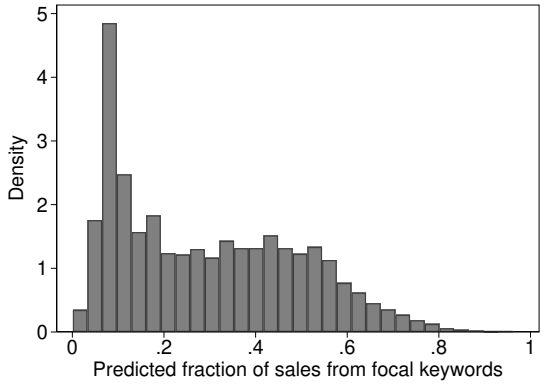
Notes: This figure displays the distribution of correlation coefficients between actual and estimated daily sales quantities for a small number of products sold on Amazon. I obtained actual daily sales quantities over two years for more than one hundred products sold on Amazon in collaboration with a large Amazon seller. Estimates of daily sales quantities were derived based on Best Seller Ranks using the method detailed in Appendix B.2.

Figure A.20: Calibrated Fraction of Sales from Focal Keywords

A: Actual and Predicted Fractions in the Prediction Sample

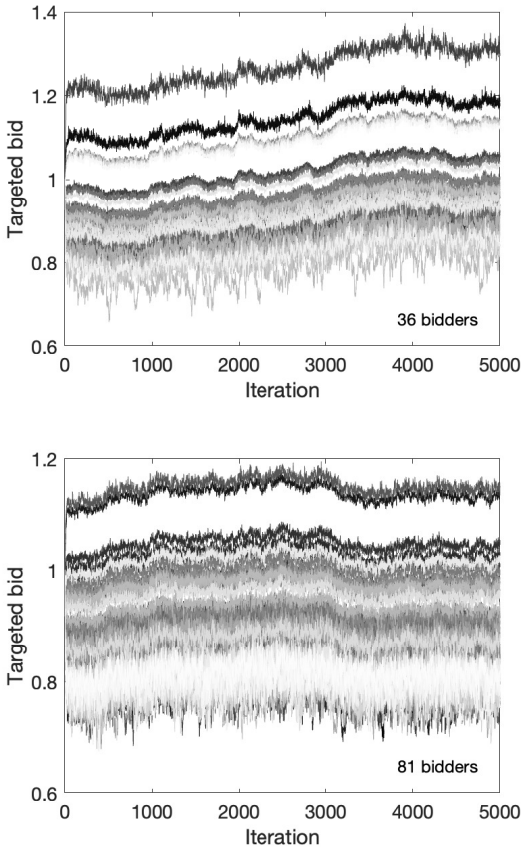


B: Distribution of Predicted Fractions in the Main Sample



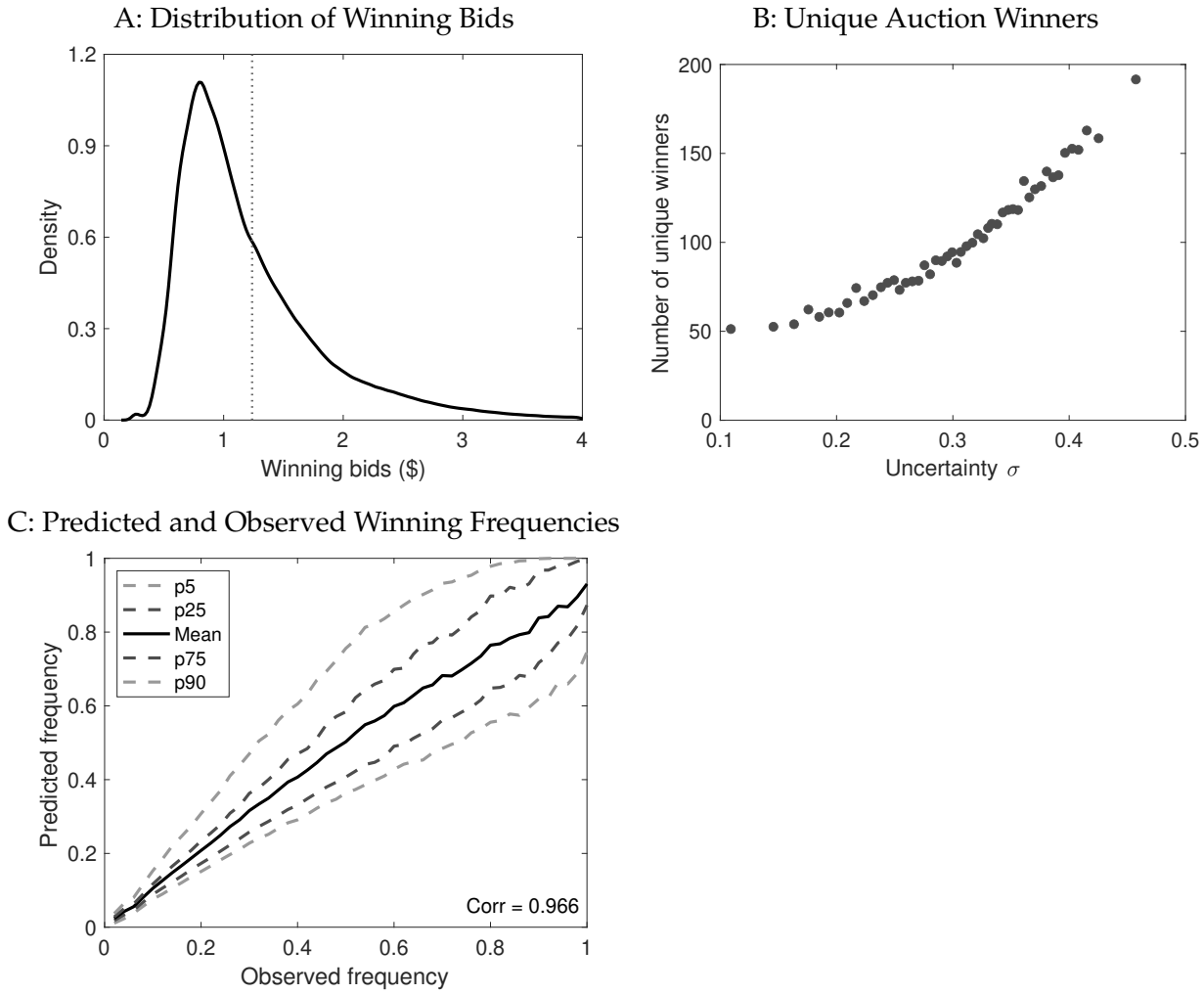
Notes: This figure illustrates the calibrated fraction of a product's total sales attributed to focal keywords. Panel A compares the actual and predicted fractions for a small group of products, where I can observe a comprehensive list of keywords whose search results contain these products. The actual fraction is constructed in equation (C.11). The predicted fraction is constructed by regressing the actual fraction on five predictors defined in Appendix C.1 and forming a linear prediction. Panel B plots the distribution of the calibrated fractions for all products in the estimation sample, which are bounded within $[0.001, 1]$.

Figure A.21: Examples of Gibbs Sampler Sequences



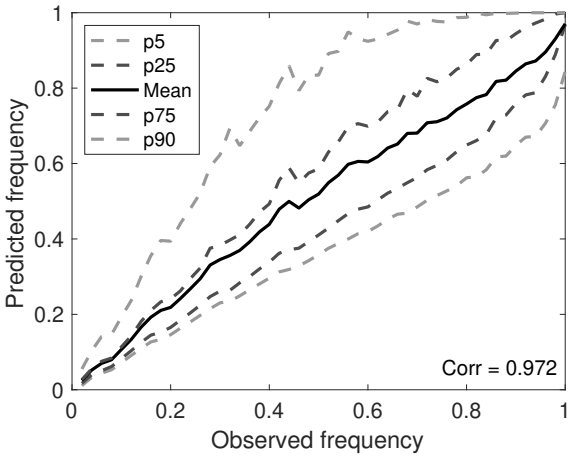
Notes: This figure presents two examples of the posterior draws for sellers' bids using a Gibbs sampler with 5,000 iterations. Each line represents a seller's bid. See Appendix D.2 for additional details on this algorithm.

Figure A.22: Estimates of the Auction Model



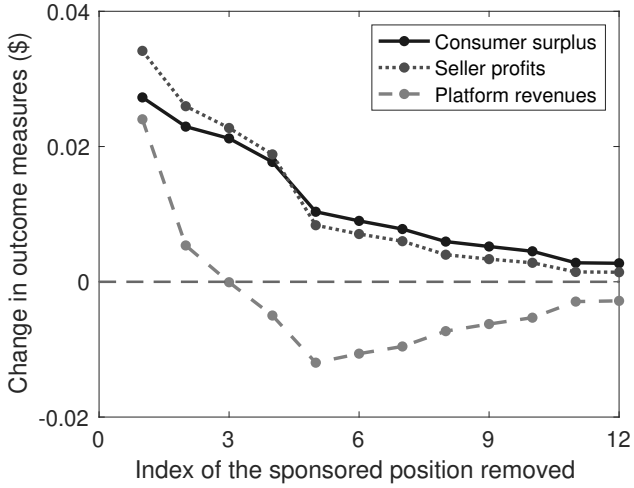
Notes: This figure summarizes the estimates of the auction model. Panel A depicts the distribution of the winning bids, pooling data from all market-week pairs. The dashed line indicates the mean. Panel B illustrates the relationship between the parameter that measures uncertainty, σ_b, and the number of unique auction winners in each market. Panel C compares the predicted frequencies of products winning the auction and observed winning frequencies. For each percentile value in the observed winning frequency, the figure displays the corresponding average predicted winning frequency, along with the 5th, 25th, 75th, and 90th percentiles.

Figure A.23: Estimates of Organic Ranks



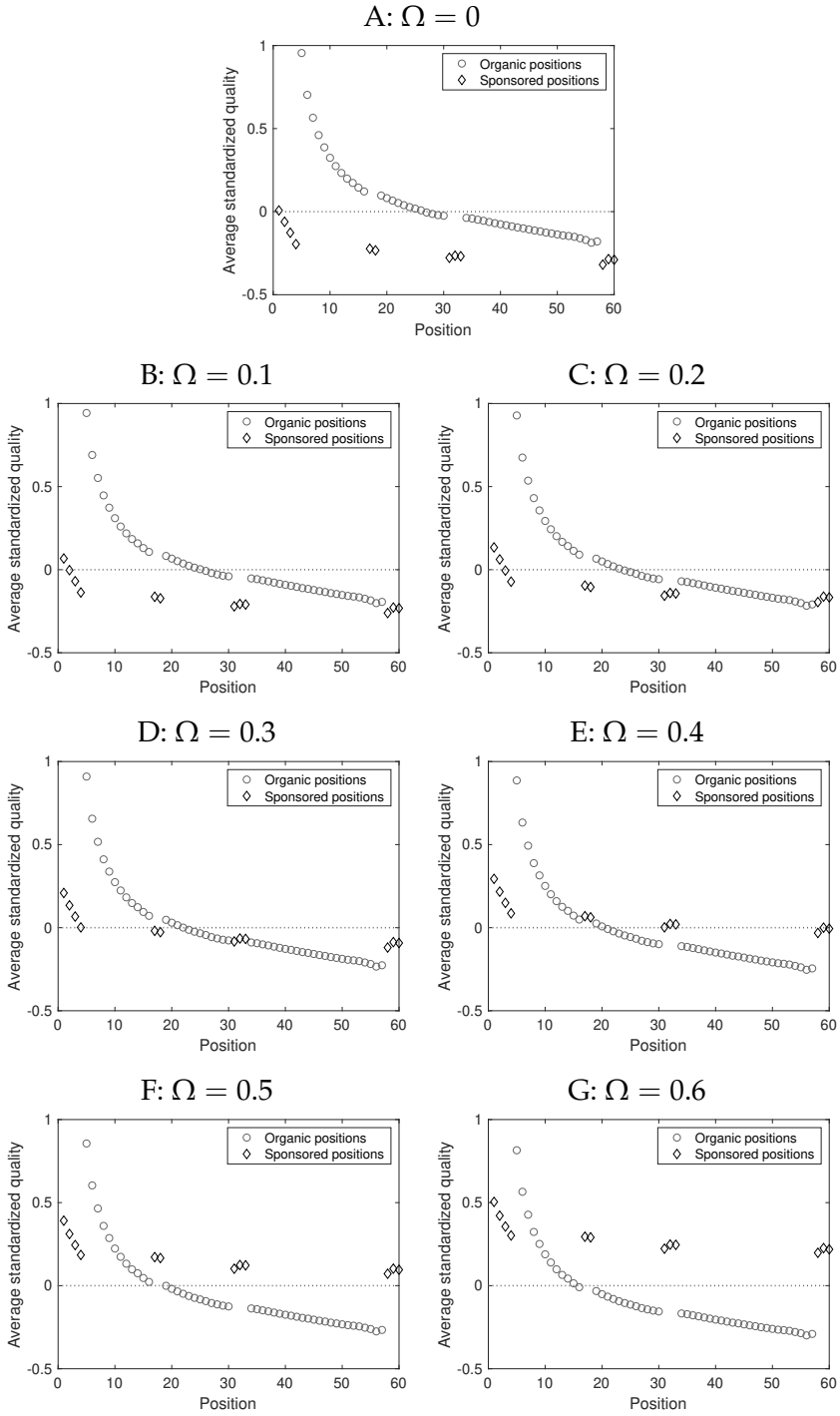
Notes: This figure summarizes the estimates of organic ranks. It compares the predicted frequencies of products appearing in organic positions with the observed frequencies. For each percentile value in the observed frequency, the figure displays the corresponding average predicted frequency, along with the 5th, 25th, 75th, and 90th percentiles.

Figure A.24: Impact of Replacing One Sponsored Position with An Organic Position



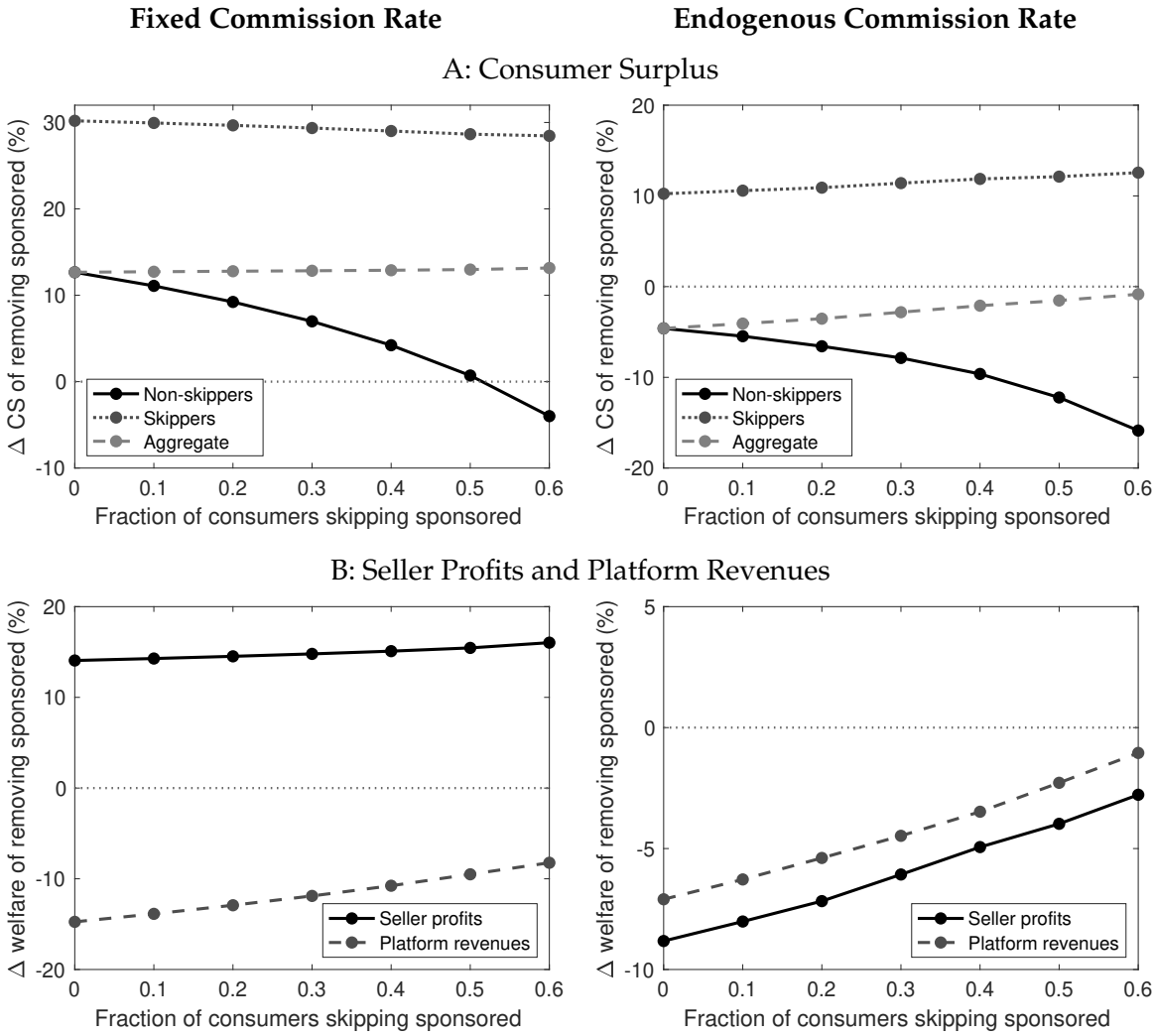
Notes: This figure demonstrates the effects of replacing one sponsored position with an organic position. The x-axis represents the index of the sponsored position being replaced. The y-axis shows the change in consumer surplus, seller profits, and platform revenues.

Figure A.25: Quality Estimates When Consumers Skip Sponsored Products



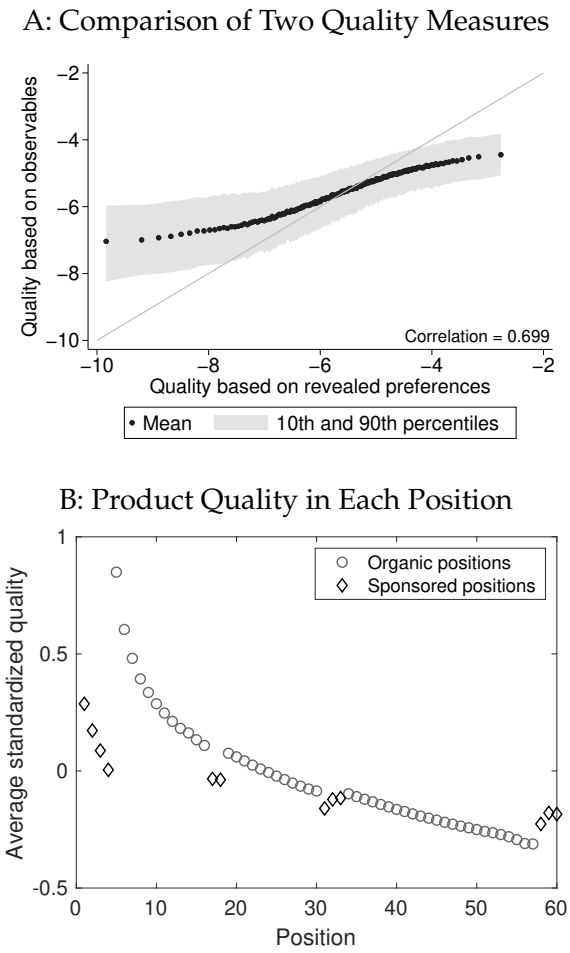
Notes: This figure presents quality estimates for products appearing in each position under different assumptions regarding consumer behaviors. Each panel corresponds to a different fraction of consumers who consistently skip sponsored positions, denoted by Ω. When a fraction λ_n of consumers considering a product in the n-th *organic* position, only a fraction $(1-\Omega)\lambda_n$ of consumers consider a product in the n-th *sponsored* position. The first panel with $\Omega = 0$ corresponds to the baseline results and replicates Panel B of Figure 3. See Appendix F.1.1 for more details on the model and Figure 3 on the construction of the quality estimates.

Figure A.26: Welfare Effects When Consumers Skip Sponsored Products



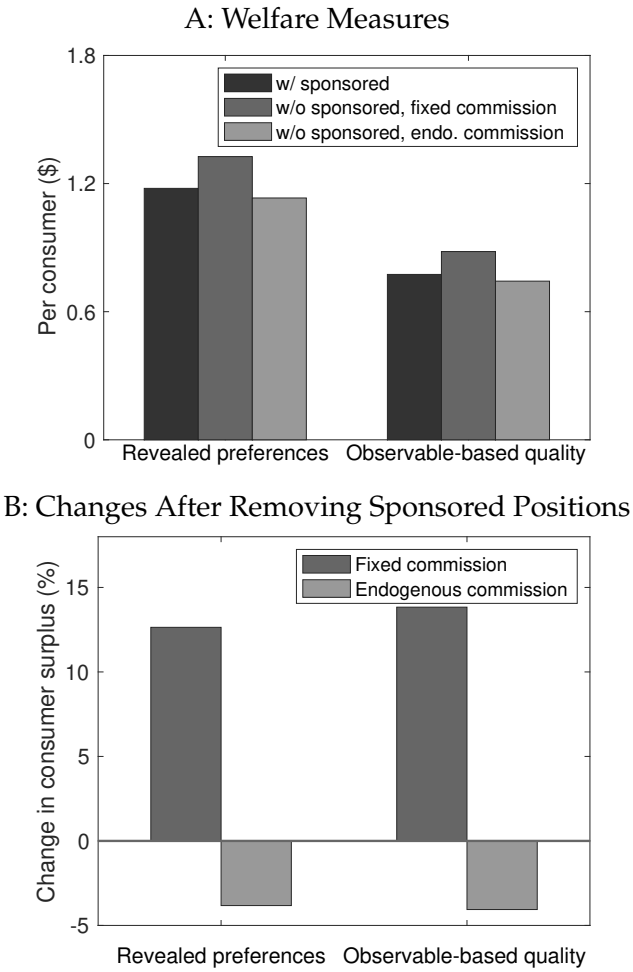
Notes: This figure displays the welfare effects of removing sponsored positions under different assumptions regarding consumer behaviors. In each panel, the x-axis represents the fraction of consumers who consistently skip sponsored positions, ranging from 0 to 0.6 with increments of 0.1. Panel A illustrates the changes in consumer surplus for skippers (those who always skip sponsored products), non-skippers (those who consider all products), and all consumers. Panel B shows the changes in seller profits and platform revenues. The left column represents scenarios with a fixed commission rate, while the right column considers endogenous commission rates. The very left point in each panel corresponds to the baseline results reported in Table 3. See Appendix F.1.1 for more details on the model.

Figure A.27: Quality Measure Based on Observables



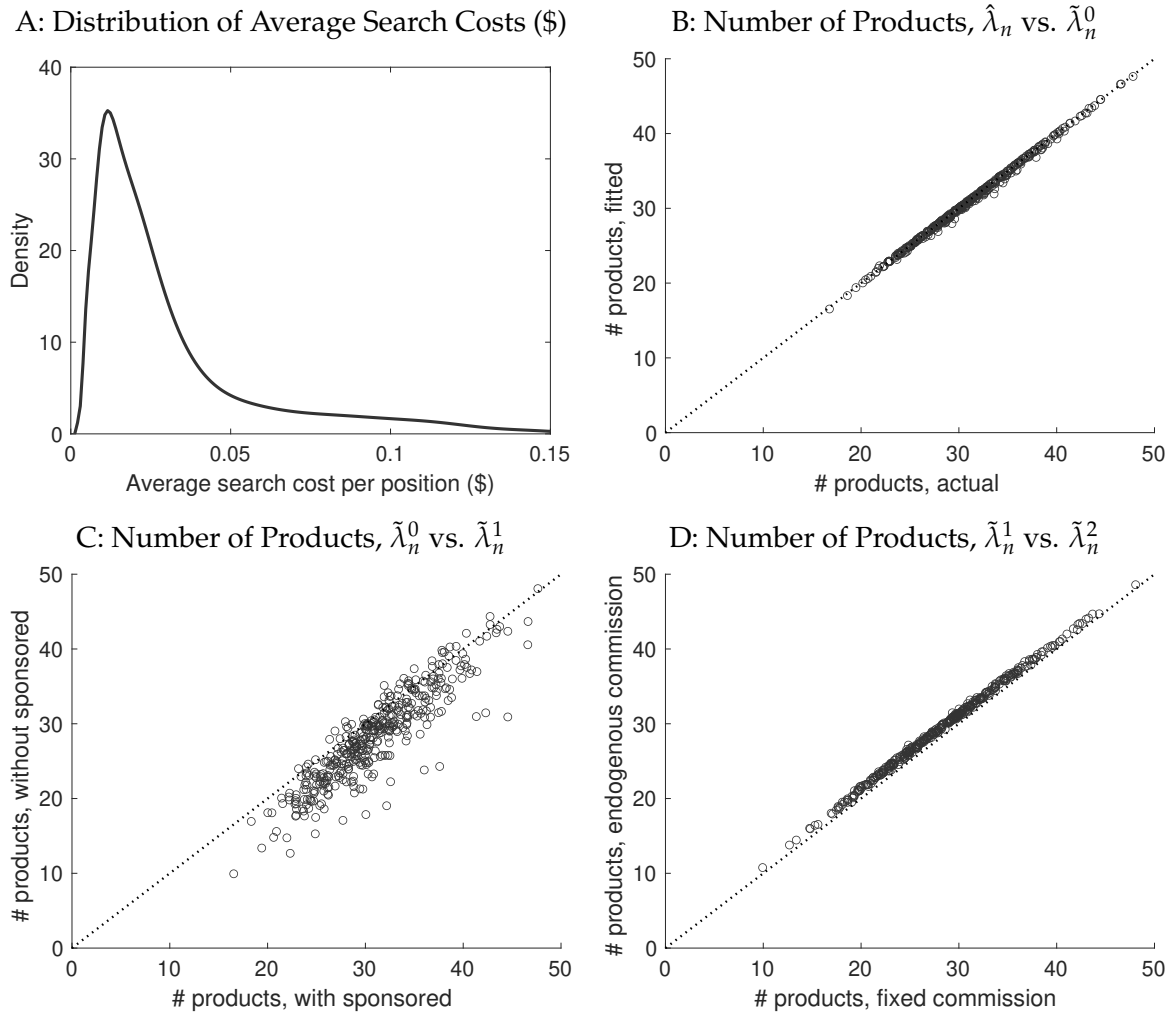
Notes: This figure presents a quality measure based on observable product characteristics. I construct this measure by regressing estimated product quality based on revealed preferences obtained in Section 6.1 on market fixed effects and four observables, restricting to products that never appear in a sponsored position within a week. These observables include the log of the number of reviews, the average consumer rating, an indicator for Prime eligibility, and an indicator for Amazon's Choice. Panel A compares product quality based on observables with product quality based on revealed preferences. For each value on the x-axis, the dot represents the mean, and the shaded area represents the 10th and 90th percentiles of the quality distribution. Panel B displays the quality estimates for products appearing in each position using the observable-based measure. See Appendix F.1.2 for more details on the construction of this quality measure.

Figure A.28: Welfare Effects Using Observable-Based Quality



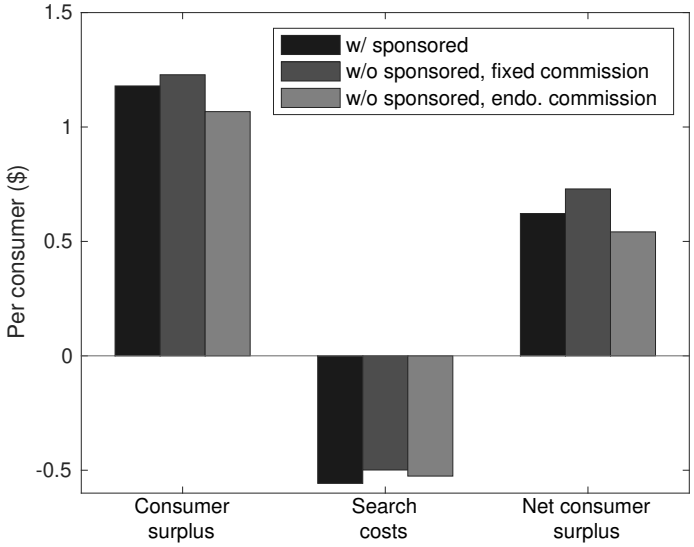
Notes: This figure displays the welfare effects of removing sponsored positions using the observable-based quality measure. Panel A presents consumer surplus under two quality measures: one based on revealed preferences (left block) and one based on observables (right block). Each block includes three scenarios: the status quo with sponsored positions, a scenario removing sponsored positions under a fixed commission rate, and a scenario removing sponsored positions under endogenous commission rates. Panel B illustrates the change in consumer surplus under the two quality measures, considering both fixed and endogenous commission rates. In both panels, the left block corresponds to the baseline results reported in Table 3. See Appendix F.1.2 for more details on the construction of this quality measure and the calculation of consumer surplus.

Figure A.29: Distribution of Search Costs and Number of Products Considered



Notes: This figure illustrates the distribution of the average search cost and the average number of products in consumers' consideration sets. Appendix F.2 outlines a sequential search model, where consumers' per-step search cost follows a log-normal distribution. The model can predict the fraction of consumers considering each position in the search results. Panel A displays the distribution of the average per-step search cost across markets, converted into dollars. I use the estimated distribution of consumers' per-step search cost in each market to predict the proportion of consumers considering each position in four scenarios, λ_n, and calculate the average number of products in consumers' consideration sets as $\sum_{n=1}^{N} n(\lambda_n - \lambda_{n+1})$. These four scenarios are: (i) the actual estimates in the status quo $\hat{\lambda}_n$; (ii) the search process in the status quo implied by the estimated distribution of search costs $\tilde{\lambda}_n^0$; (iii) the scenario with sponsored positions removed under a fixed commission rate $\tilde{\lambda}_n^1$; and (iv) the scenario with sponsored positions removed under endogenous commission rates $\tilde{\lambda}_n^2$. Panels B to D compare the average number of products in consumers' consideration sets for (i) vs. (ii) (as an evaluation of the model fit), (ii) vs. (iii), and (iii) vs. (iv), respectively. See Appendix F.2 for more details on the model.

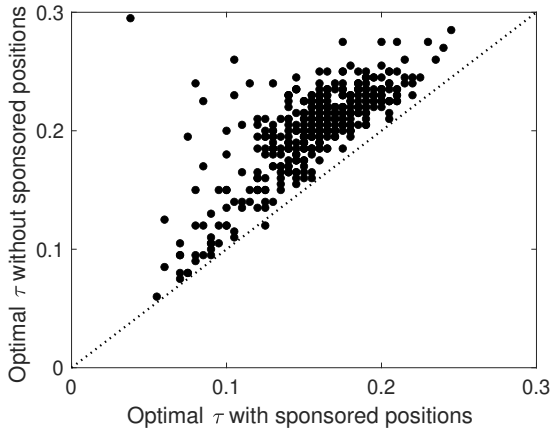
Figure A.30: Effects on Consumer Surplus with Endogenous Consideration Sets



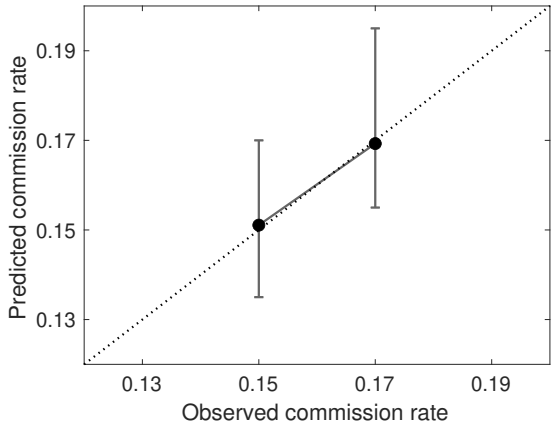
Notes: This figure displays the overall effect on consumer surplus with consideration sets endogenized and search costs incorporated. The first block illustrates consumer surplus without accounting for search costs, the second block quantifies the total search costs incurred, and the third block presents the net consumer surplus as the difference between consumer surplus and search costs. Each block includes three scenarios: the status quo with sponsored positions, a scenario with sponsored positions removed under a fixed commission rate, and a scenario with sponsored positions removed under endogenous commission rates. See Appendix F.2 for more details on the model.

Figure A.31: Market-Specific Commission Rates

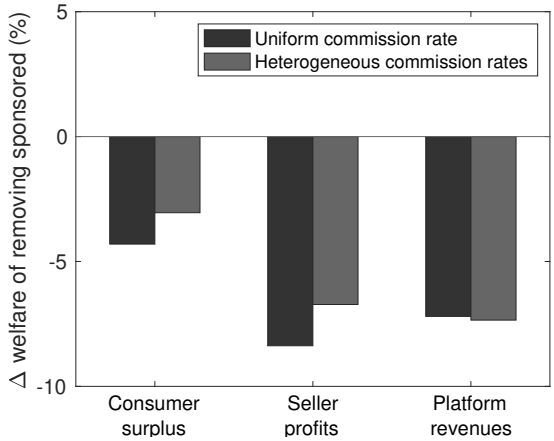
A: Optimal Commission Rates With and Without Sponsored Positions



B: Comparison of Predicted and Observed Commission Rates

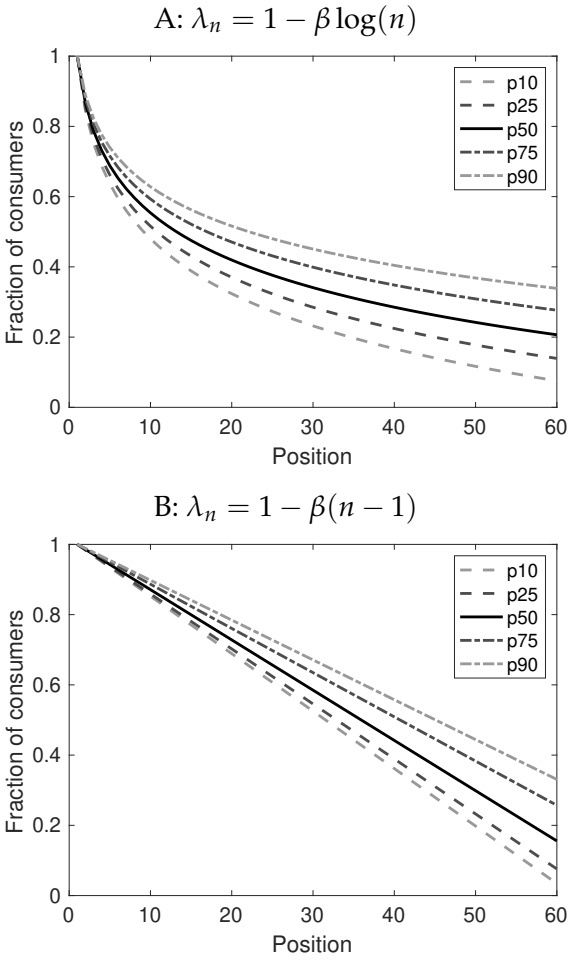


C: Welfare Effects with Market-Specific Commission Rates



Notes: This figure explores the implication of market-specific optimal commission rates. Panel A illustrates the optimal commission rates with and without sponsored positions in each market that maximize the platform's objective function. Panel B compares the predicted commission rates for markets with an observed commission rate of 15% and 17%. The dot represents the mean, and the error bar represents the 25th and 75th percentiles. Panel C displays the welfare effects of removing sponsored positions under uniform or market-specific commission rates.

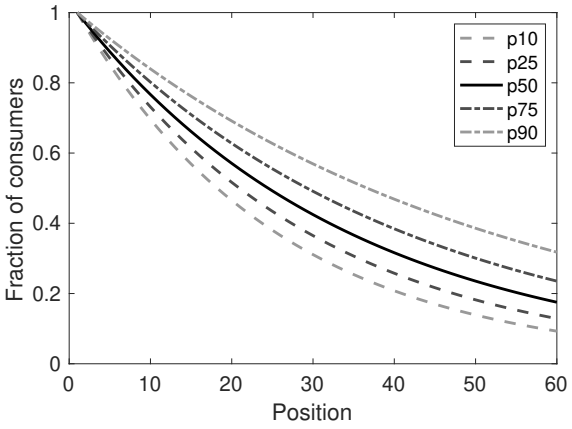
Figure A.32: Demand Estimates Under Alternative Parametrization of Search Frictions



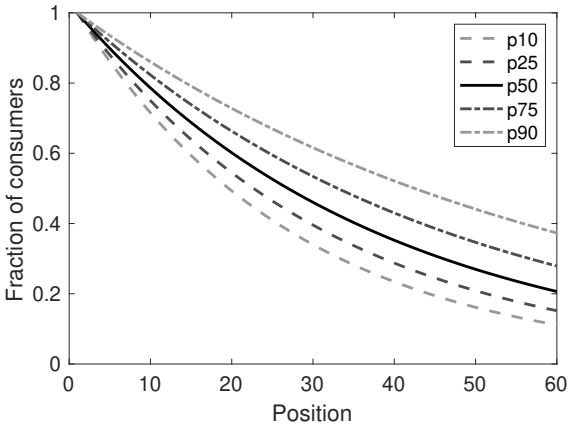
Notes: This figure presents demand estimates using different functional forms of consideration sets. The fraction of consumers whose consideration sets contain the n-th product in the search results is denoted as λ_n. Panel A presents the estimates assuming a functional form $\lambda_n = 1 - \beta \log(n)$. Panel B presents the estimates assuming a functional form $\lambda_n = 1 - \beta(n-1)$. Different lines correspond to different percentiles of λ_n across markets. Panel A of Figure 3 presents the estimates under the baseline specification, $\lambda_n = 1/\exp(\beta(n-1))$.

Figure A.33: Demand Estimates Under Alternative Parametrization of Utility Function

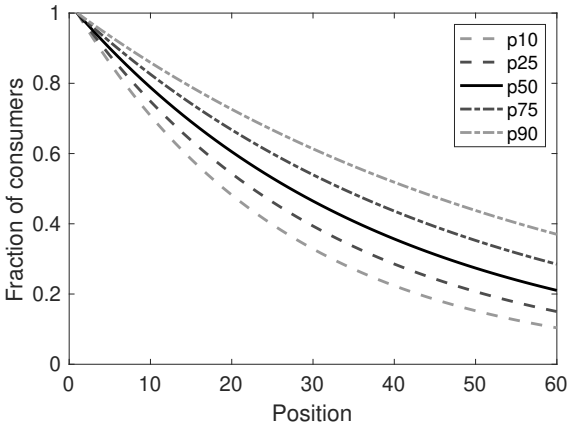
A: Dropping Linear Time Trend



B: Including Product-Month Fixed Effects

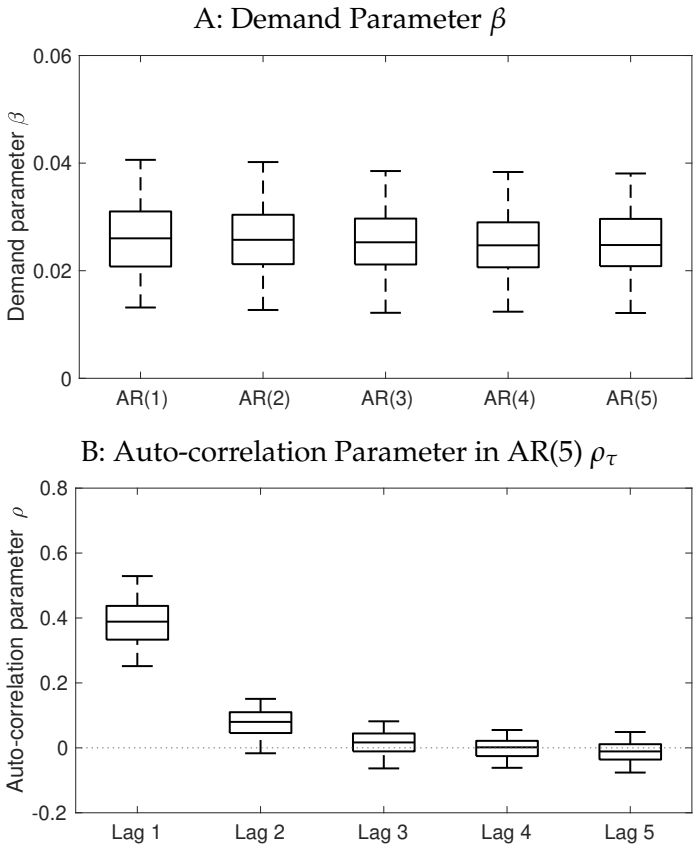


C: Including Product-Specific Quadratic Time Trend



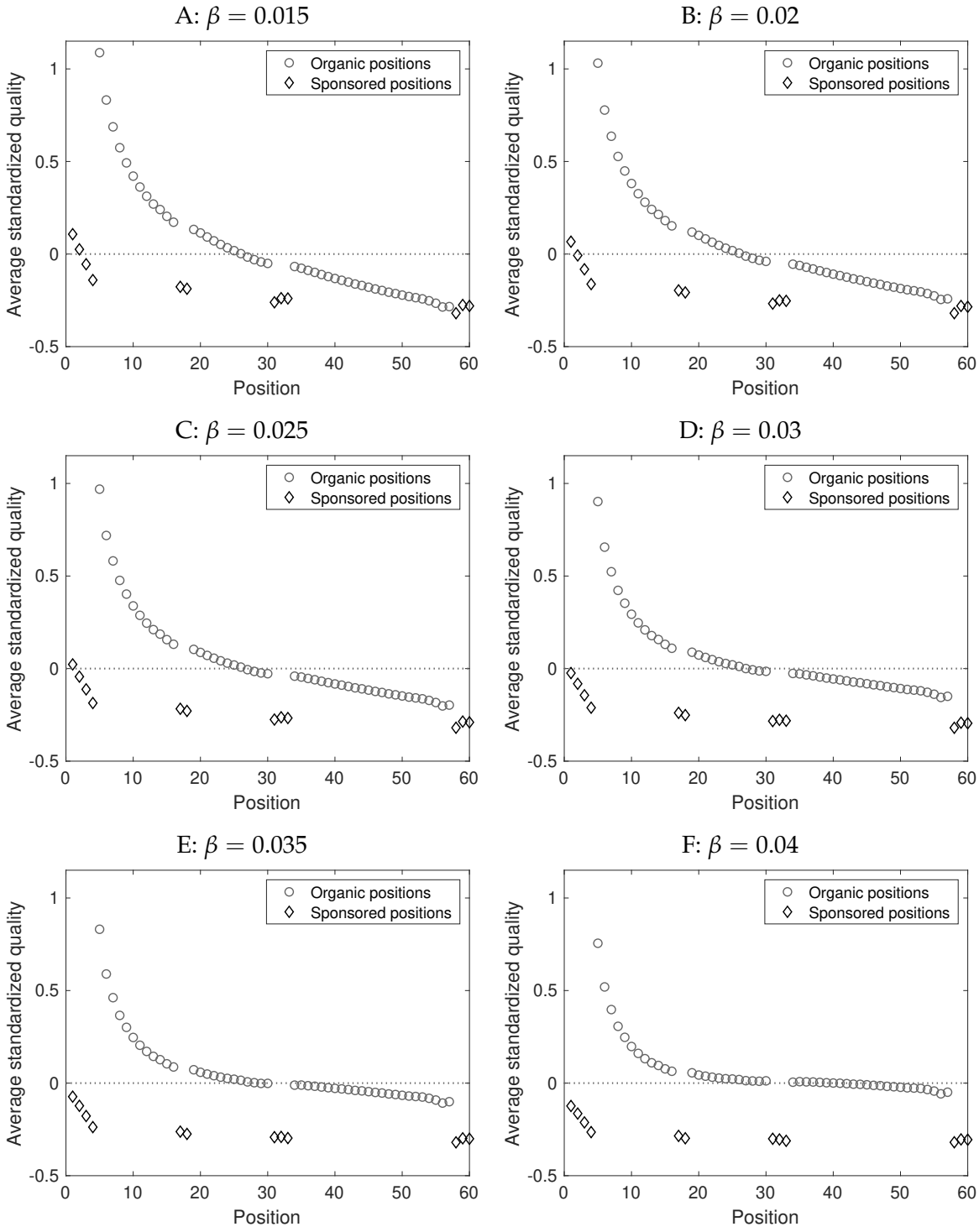
Notes: This figure presents demand estimates employing alternative parametrizations of consumers' utility function in equation (3). Panel A of Figure 3 presents the estimates under the baseline specification, which includes a product-specific linear time trend. In this figure, Panels A to C show the results when I drop the linear time trend, replace it with product-month fixed effects, or introduce a product-specific quadratic time trend. Different lines correspond to different percentiles of λ_n across markets.

Figure A.34: Demand Estimates Under Alternative Autoregressive Models



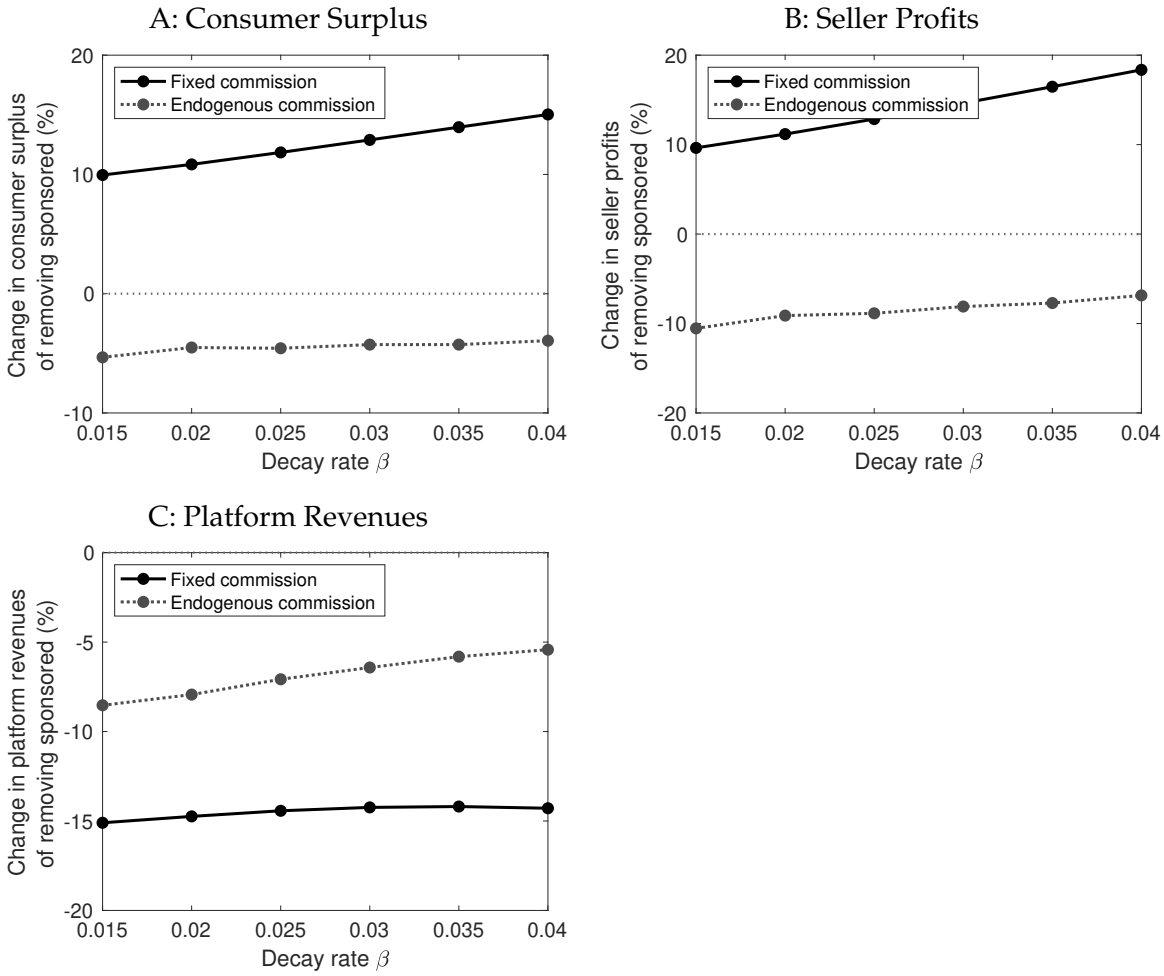
Notes: This figure presents demand estimates under a generalized autoregressive model where the unobserved demand shock ξ_{jt} follows an AR(p) process given by $\xi_{jt} = \sum_{\tau=1}^{p} \rho_\tau \xi_{j,t-\tau} + \eta_{jt}$, with $\eta_{jt} \perp \xi_{j,t-\tau}, \forall \tau \geq 1$. The baseline analysis in Section 5.1.2 sets $p = 1$. Panel A displays the estimates of the demand parameter β under different assumptions for p. Panel B shows the estimates of the autoregressive parameters ρ_τ for $\tau = 1, ..., 5$ when $p = 5$.

Figure A.35: Quality Estimates Under Different Extents of Search Frictions



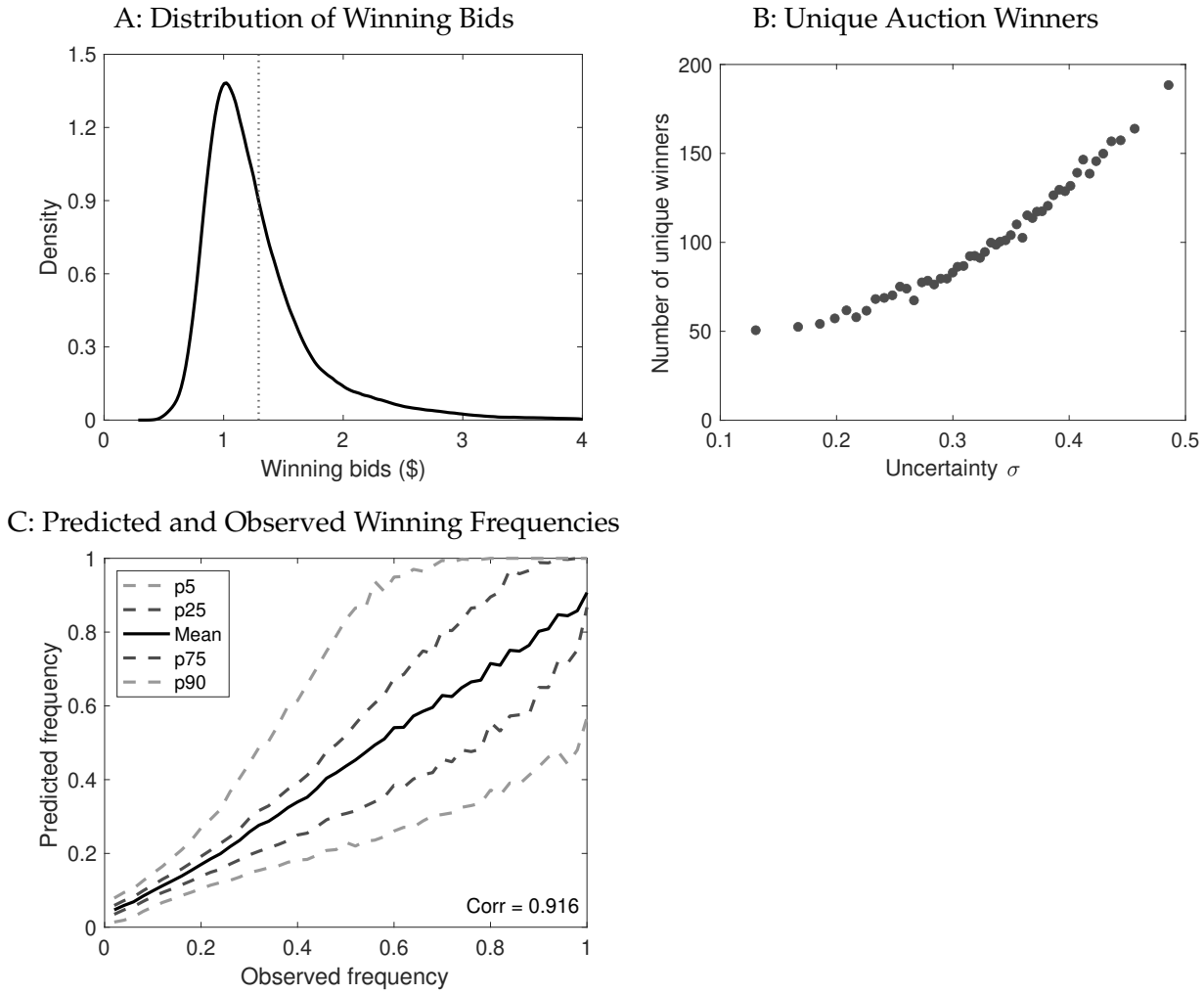
Notes: This figure presents quality estimates for products appearing in each position under different assumptions regarding the extent of search frictions. The fraction of consumers whose consideration sets contain the n-th product in the search results is parametrized as $\lambda_n = 1/\exp(\beta(n-1))$. Each panel presents the quality estimates under a different value of the parameter β.

Figure A.36: Welfare Effects Under Different Extents of Search Frictions



Notes: This figure displays the welfare effects of removing sponsored positions under different assumptions regarding the extent of search frictions. In each panel, the x-axis represents the parameter β in the parametrization of the fraction of consumers whose consideration sets contain the n-th product in the search results, $\lambda_n = 1/\exp(\beta(n-1))$, ranging from 0.015 to 0.04 with increments of 0.005. Panels A to C illustrate the change in consumer surplus, seller profits, and platform revenues. In each panel, I consider both fixed and endogenous commission rates.

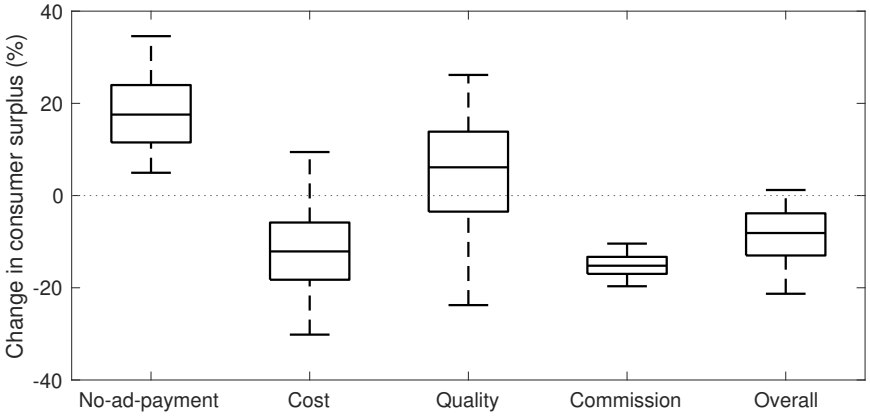
Figure A.37: Estimates of the Auction Model, Alternative Parametrization



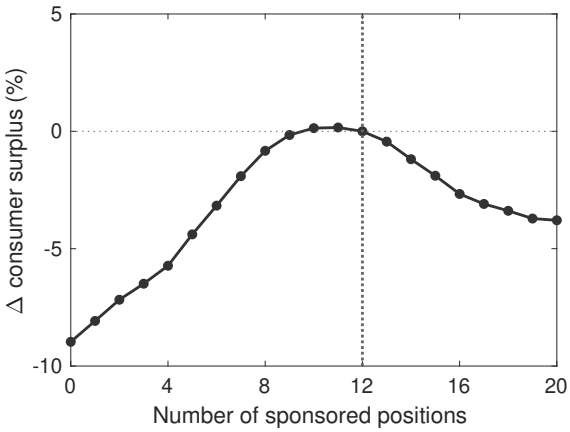
Notes: This figure summarizes the estimates of the auction model under an alternative parametrization. In the baseline analysis, when a seller submits a bid b_j, the realized bid $\tilde{b}_j$ follows a log-normal distribution: $\log(\tilde{b}_j) \overset{\text{i.i.d.}}{\sim} N(\log(b_j), \sigma_b^2)$. I explore an alternative distributional assumption where the realized bid follows a normal distribution: $\tilde{b}_j \overset{\text{i.i.d.}}{\sim} N(b_j, \sigma_b^2)$. Panel A depicts the distribution of the winning bids, pooling data from all market-week pairs. The dashed line indicates the mean. Panel B illustrates the relationship between the parameter that measures uncertainty, σ_b, and the number of unique auction winners in each market. Panel C compares the predicted frequencies of products winning the auction and observed winning frequencies. For each percentile value in the observed winning frequency, the figure displays the corresponding average predicted winning frequency, along with the 5th, 25th, 75th, and 90th percentiles.

Figure A.38: Consumer Surplus Under Alternative Assumption on Duplicated Listings

A: Decomposition of the Change in Consumer Surplus



B: Change in Consumer Surplus Under Varying Numbers of Sponsored Positions



Notes: This figure presents the welfare results under an alternative assumption concerning duplicated listings. The baseline analysis assumes that the second appearance of the same product in search results does not receive an independent idiosyncratic preference shock when calculating consumer surplus. Appendix G.6 considers an alternative assumption where the second appearance also receives its independent preference shock. Panel A decomposes the aggregate change in consumer surplus into four components, corresponding to Figure 7. Panel B presents the change in consumer surplus when varying the number of sponsored positions in the search results relative to the status quo with 12 sponsored positions, corresponding to Panel B of Appendix Figure 9.

Table A.1: Examples of Markets and Keywords

Market	Keyword	30-Day Search Volume
baby walker	baby walker	137,093
	baby walkers for boys	25,303
	walker for baby boy	24,834
	baby walkers for girls	17,125
	walker for baby girl	15,509
pedicure kit	pedicure kit	118,896
	manicure kit	26,475
	pedicure tools	23,148
	pedicure	12,891
	pedicure supplies	12,396
paper towels	paper towels	513,567
	bounty paper towels	52,401
	paper towel	49,573
	paper towels bulk	32,874
	viva paper towels	14,100
coffee maker	coffee maker	520,363
	coffee pot	53,636
	coffee machine	45,263
	coffee makers	30,407
	drip coffee maker	12,390
sticky notes	sticky notes	113,604
	post it notes	106,387
	notes	15,766
	post it	11,244
	post its	10,918
dog bed	dog bed	488,901
	dog beds	75,156
	large dog bed	59,034
wedding dress	wedding dress	253,725
	wedding dresses for bride	103,461
	wedding dresses	40,288
outdoor toys	outdoor toys	104,029
	kids outdoor toys	45,879
	toddler outdoor toys	43,277
	outside toys	27,196
	outdoor toys for kids	10,840
stapler	stapler	113,008
	heavy duty stapler	10,246

Notes: This table provides several examples of markets and keywords in the sample. Keywords whose search results contain a substantial overlap of products are grouped into the same market using the method described in Appendix B.1. The sample contains 3,237 keywords in 546 markets. Data on search volumes are collected from Jungle Scout.

Table A.2: Prediction of the Fraction of Sales From Focal Keyword

	Dep. Var. = Frac. of Sales From Focal Keyword				
	(1)	(2)	(3)	(4)	(5)
	$\beta = 0.01$	$\beta = 0.02$	$\beta = 0.03$	$\beta = 0.04$	$\beta = 0.05$
Log(Rank)	-0.063	-0.091	-0.113	-0.132	-0.147
	(0.010)	(0.011)	(0.011)	(0.012)	(0.012)
Log(# Other Keywords + 1)	-0.018	-0.025	-0.031	-0.036	-0.039
	(0.024)	(0.025)	(0.026)	(0.027)	(0.028)
Fraction of Search Volume	0.608	0.566	0.525	0.486	0.449
	(0.058)	(0.060)	(0.062)	(0.064)	(0.066)
Log(Avg. Rank in Other Keywords)	0.072	0.080	0.085	0.088	0.089
	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)
Appear in Other Keywords	-0.224	-0.260	-0.284	-0.299	-0.308
	(0.053)	(0.055)	(0.057)	(0.059)	(0.060)
Constant	0.256	0.361	0.447	0.519	0.580
	(0.066)	(0.069)	(0.072)	(0.074)	(0.076)
N	496	496	496	496	496
R^2	0.382	0.364	0.355	0.352	0.353

Notes: This table presents the regression results of predicting the fraction of a product's total sales attributed to focal keywords, as outlined in Appendix C.1. Each observation corresponds to a product. The dependent variable represents the fraction of the product's total sales from the focal keyword, as constructed in equation (C.11). I include five predictors in the regressions, as defined in Appendix C.1. Each column in the table corresponds to different values of the demand parameter β, where the fraction of consumers considering each position in search results is parametrized as $\lambda_n = 1/\exp(\beta(n-1))$ for $1 \leq n \leq N$. Standard errors are reported in parentheses.

Table A.3: Welfare Effects Under Alternative Assumption on Duplicated Listings

	(1)	(2)	(3)	(4)	(5)
	Status Quo	Counterfactual: Without Sponsored Product Advertising			
		Fixed Commission		Endogenous Commission	
	Level ($)	Level ($)	Change (%)	Level ($)	Change (%)
Consumer Surplus	1.24	1.33	7.2%	1.13	-9.0%
Seller Profits	0.81	0.93	14.1%	0.75	-8.4%
Platform Revenues	0.89	0.76	-14.8%	0.83	-7.2%
Commission Revenues	0.71	0.76	7.4%	0.83	16.9%
Advertising Revenues	0.18	0.00	-100.0%	0.00	-100.0%
Total Surplus	2.94	3.02	2.5%	2.70	-8.3%

Notes: This table presents the welfare effects under an alternative assumption concerning duplicated listings. The baseline analysis assumes that the second appearance of the same product in search results does not receive an independent idiosyncratic preference shock when calculating consumer surplus. Appendix G.6 considers an alternative assumption where the second appearance also receives its independent preference shock. This table replicates Table 3 under this alternative assumption.