NBER WORKING PAPER SERIES

SELF-PREFERENCING AT AMAZON: EVIDENCE FROM SEARCH RANKINGS

Chiara Farronato
Andrey Fradkin
Alexander MacKay

Working Paper 30894
http://www.nber.org/papers/w30894

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2023

We thank James Dana, Yutao Chen, Yejia Xu, and Hannah Huihan Zhang for outstanding research assistance. We also thank Shane Greenstein, Jesse Shapiro, Raluca Ursu, and seminar and conference participants at the NUS Economics of Platforms Workshop, the 2023 ASSA Meeting, and Harvard Business School for valuable comments. The authors have no material or relevant financial interests to disclose. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Self-Preferencing at Amazon: Evidence from Search Rankings
Chiara Farronato, Andrey Fradkin, and Alexander MacKay
NBER Working Paper No. 30894
January 2023
JEL No. D12,D83,L13,L15,L81

ABSTRACT

We study whether Amazon engages in self-preferencing on its marketplace by favoring its own brands (e.g., Amazon Basics) in search. To address this question, we collect new micro-level consumer search data using a custom browser extension installed by a panel of study participants. Using this methodology, we observe search positions, search behavior, and product characteristics. We find that Amazon branded products are indeed ranked higher than observably similar products in consumer search results. The prominence given to Amazon brands is 30% to 60% of the prominence granted to sponsored products.

Chiara Farronato
Harvard Business School
Morgan Hall 427
Soldiers Field
Boston, MA 02163
and NBER
cfarronato@hbs.edu

Andrey Fradkin
Questrom School of Business
Boston University
Rafik B. Hariri Building
595 Commonwealth Avenue
Boston, MA 02215
fradkin@bu.edu

Alexander MacKay
Harvard Business School
Morgan Hall 221
Soldiers Field
Boston, MA 02163
amackay@hbs.edu

1 Introduction

Many firms, from retailers to investment management companies, offer their own products alongside products sold by competitors. While this practice has largely been accepted in many settings (see, e.g., Dubé, 2022), it has faced substantial scrutiny in online markets, especially for large digital platforms such as Google and Amazon.[1] Regulators are especially concerned that digital platforms may give preferential treatment to their own products—e.g., Google Maps or Amazon Basics—over those of third-party sellers, a practice referred to as *self-preferencing*.

Self-preferencing can be good or bad for consumers. When digital platforms introduce and promote new products, this can increase variety and generate competition that lowers prices and increases quality. However, if the favorable treatment of vertically owned products makes it difficult for consumers to find their preferred options, consumers could be harmed. In addition, self-preferencing could discourage product innovation by other firms and could cause some competitors to leave the market altogether.

In this paper, we explore whether Amazon engages in self-preferencing on its marketplace. We find that Amazon branded products are indeed ranked higher than observably similar products in consumer search results. To show this, we collect new micro-level consumer search data using a custom browser extension installed by a panel of study participants. Using this methodology, we observe search positions, search behavior, and product characteristics. This allows us to evaluate whether Amazon brands are ranked higher in search results, holding other observable factors constant.

Our work contributes to a recent literature on vertical integration on Amazon (Jeffries and Yin (2021), Lee and Musolff (2022), Gutierrez (2022), Lam (2022), Chen and Tsai (2022), Raval (2022)). A key advantage of our approach is that our data reflect real consumer searches, for which results, including delivery times and targeted ads, can be personalized. Our focus on search results as a venue for self-preferencing is justified by the fact that, in our data, half of product pages are reached through a search conducted on Amazon. Our evidence thus points to the pivotal role of search results in the purchase decision.

2 Data

To explore Amazon's gatekeeping role in search results, we use data from two pilot studies conducted in the Summer and Fall of 2022.[2] We recruited participants residing in the US from CloudResearch and Facebook (via ads) for a study to understand the costs and benefits of vertical integration in online platforms. What we describe here about self-preferencing in search

[1]For example, the European Union recently passed the Digital Markets Act with the goal of limiting the market power of large platforms. In the US, President Biden's Executive Order on Promoting Competition in the American Economy and the House Antitrust Subcommittee Investigation of Competition in Digital Markets discuss similar interventions.

[2]The pilot studies were both approved under Harvard IRB21-1677.

results constitutes a first step towards answering the more relevant question about the effects of vertical integration for consumer welfare, an objective of our larger project.

In order to be eligible for the pilots, participants had to be frequent Amazon shoppers—i.e., they purchased from Amazon at least twice a month—and mostly used the Chrome browser on their desktop computer for online shopping. These device criteria are required because we tracked users' online behavior with Webmunk, a desktop web browser extension developed for studies of this type (Farronato, Fradkin and Karr (2023)). Consenting participants installed Webmunk on their Chrome browser for six weeks,[3] allowing us to track their browsing activity.

We have data on users' searches on Amazon, which we use to study how Amazon ranks products. Search results are the most important channel for product discovery. In our sample, 46.5% of product pages are reached from a search result page, which is the largest referral source (the next largest referral source is links from other web domains at 11.2%). Further, the order in which products are displayed appears to be important because users do not see all search results. In 72.1% of searches consumers do not click past the first results page, and, based on scroll position data, only half of the products on a full page of results are actually seen by consumers.

Our data contain 228,281 search results– including both sponsored and organic results—in 3,019 unique searches conducted by 184 users. On average, Amazon returns 76 results per search, with large variation across searches (standard deviation of 70). The number of items returned for a given search is both a function of product availability and the consumer's decision whether to progress across multiple results pages.

The search terms are very idiosyncratic. The most common search is for gift cards, which we do not include in our sample. We see some identical search terms across two distinct participants (e.g., "paper towels", "trail mix", and "dayquil"), and at most we observe the same search term across three participants (for "cat food"). The vast majority of searches are unique to each user.

We generate an indicator for whether a search result is for an Amazon-branded product. We do this in two steps. First, in real time, the browser extension identifies Amazon brands by comparing the content of the product's HTML with a list of pre-determined character strings that include the most popular Amazon brands[4] as well as whether the item is flagged by Amazon itself as an Amazon brand.[5] Second, after the data collection, we check whether the product

[3] In case participants took longer than 6 weeks to complete the study, which is possible, we cap the tracking period to 60 days. For some participants, we have less than 6 weeks because the second pilot is still ongoing at the time of this writing.

[4] We search for the following Amazon brands: 'Amazon Basic Care', 'Amazon Basics', 'Amazon Collection', 'Amazon Commercial', 'Amazon Elements', 'Amazon Essentials', '206 Collective', 'Amazing Baby', 'Buttoned Down', 'Cable Stitch', 'Core 10', 'Daily Ritual', 'Goodthreads', 'Isle Bay', 'Lark & Ro', 'Moon and Back by Hanna Andersson', 'Mountain Falls', 'P2N Peak Performance', 'Pinzon', 'Presto!', 'Simple Joys by Carter's', 'Solimo', and 'Spotted Zebra'.

[5] Amazon started including the Amazon brand badge to search results before the start of our pilots. Whenever an Amazon brand is advertised, Amazon shows the Featured from our brands flag below the product image rather than the Sponsored flag. The browser extension identifies both phrases 'Amazon Brand' and 'Featured from Our Brands'

Table 1: Amazon Brands versus Other Products

Variable	Other Products	Amazon Brand
Share Sponsored	0.232	0.248
Share Prime	0.689	0.695
Share Same-Day Delivery	0.044	0.056*
Share Overnight Delivery	0.018	0.033*
Share with No Ratings	0.044	0.004*
Average Stars	4.484	4.521*
Num. Ratings	7,644	20,134*
Average Price ($)	37.83	25.77*
Average Rank	43.02	33.12*
Num. Products	54,617	2,920

Notes: The table presents descriptive characteristics of Amazon-branded products (right column) compared to other products (left column). For this table, we consider only the products appearing in searches where at least one Amazon brand also appears. 594 out of the 3,019 searches return at least one Amazon-branded product. The star denotes statistically significant differences between the two columns at the 1% confidence level.

title contains any of the same pre-determined strings. Our browser extension is able to identify 98.6% of all Amazon-branded products in real time.

Our data show that, on average, 1.3% of search results are Amazon-branded products. However, there is large heterogeneity across searches. Only 19.7% of searches return at least one Amazon branded product. Among those searches, on average 5.9% of results are Amazon brands (standard deviation of 6.4%).

We observe meaningful differences in product characteristics for searches that return Amazon brands compared to those that do not. On average, products in searches with Amazon brands tend to have more consumer ratings and lower prices compared to products in other searches. They are also more likely to be eligible for Amazon Prime benefits, which include faster delivery and free shipping.

Table 1 presents product-level characteristics for the subsample of searches that return at least one Amazon brand. The right column focuses on Amazon brands, while the left column includes other products. On average, Amazon brands and other products are similar in Prime eligibility and the rate at which they are sponsored, but are very different across other dimensions. Amazon brands are more likely to have faster shipping and more likely to have at least one rating. Conditional on being rated, they have more than twice as many customer reviews. Amazon-branded products also tend to be cheaper, with an average price of $26 compared to $38 for other products. After controlling for many observable characteristics, Amazon brands remain about 30% cheaper and have 68% more reviews than other similar products.

Finally, Table 1 shows that on average, Amazon-branded products appear more prominently in search results. The average rank for Amazon brands is 33, compared to 43 for other products. The next section focuses on product prominence in more detail.

as denoting Amazon brands.

3 Prominence in Search Results

In this section, we focus on the prominence given to products in Amazon search results, with specific interest in identifying the features correlated with prominence. There are a number of seller blogs describing the factors entering Amazon A10's algorithm,[6] although the actual weights are, of course, proprietary.

We obtain all product features appearing in the search results. The features include prices (including any discounts), quality metrics from reviews (number and average star rating), delivery speed (including Prime eligibility), stock availability, and the number of new and used product options. Because we have the search terms and the product titles, we also create measures of product relevance by measuring the cosine similarity and Levenshtein distance between search terms and product titles. Finally, we capture whether an item is sponsored, albeit with some noise.[7] To ensure that our results are not affected by such noise, in robustness checks we exclude products for which the sponsored flag is imprecisely measured. Those products are contained in special carousels (e.g., "Highly Rated" or "Amazon's Choice" carousels) and constitute 13.2% of all search results.

We run OLS regressions of the following form:

$$y_{ij} = \alpha \; amazon_{ij} + \beta \; sponsored_{ij} + \gamma X_{ij} + \epsilon_{ij},$$

where i denotes a search result for search j. We use the rank of the product in the search results page as main outcome y_{ij}. To compute the rank, we assign rank 1 to the first product shown on the upper left side of the page, and we then sequentially allocate rank from left to right and top to bottom, like one would read a book. We use position data of each product on the web page to construct this rank. For the majority of products, a rank value can be extracted from HTML tags, which we use to validate our outcome variable.[8]

We are interested in whether the dummy for Amazon brands ($amazon_{ij}$) predicts prominence in search results. To make the size of the coefficient estimate interpretable, we compare it to the size of the coefficient for the sponsored dummy ($sponsored_{ij}$), which we expect should increase prominence.

The vector X_{ij} contains additional controls. First, we include search spell fixed effects to account for differences in the types of products that appear across different searches. Second, we add a large set of covariates that can be broadly grouped into five groups: product relevance with the search performed, product availability (both in terms of stock and used/new options

[6] For example, `https://eva.guru/blog/amazon-a10-algorithm/`.

[7] The sponsored flag on Amazon is usually right next to the product. Sometimes, however, when the product is nested in a carousel of multiple results, the sponsored flag is outside of the HTML that we capture. In those cases, we check for other HTML tags that indicate sponsoring.

[8] We remove fewer than 1% of search results for which the HTML does not contain the rank set by Amazon (because, for example, it is in a special carousel) and for which the browser extension fails to record the position on the webpage.

Table 2: Features Correlated with Product Prominence

	Rank (1)	Rank (2)	Rank (3)	Rank (4)	Rank (5)	In Top 10 (6)
Amazon Brand	-6.16***	-3.84***	-3.87***	-3.17***	-2.90**	0.028**
	(1.02)	(0.975)	(0.975)	(0.758)	(0.992)	(0.010)
Sponsored	-7.09***	-6.55***	-6.55***	-5.97***	-8.32***	0.098***
	(0.256)	(0.282)	(0.282)	(0.294)	(0.476)	(0.003)
Major Brand			-0.662			
			(0.602)			
Search Spell FE	Yes	Yes	Yes	Yes	Yes	Yes
Additional Controls		Yes	Yes	Yes	Yes	Yes
R^2	0.489	0.516	0.516	0.520	0.468	0.148
Observations	228,281	228,281	228,281	198,116	57,537	228,281
Mean of Y	38.9	38.9	38.9	34.5	42.5	0.179
Sample	All	All	All	No Carousels	Search Has Amazon Brand	All

Notes: The table presents estimates of OLS regressions of two measures of product prominence on product characteristics. Columns (1) through (5) show results where the outcome is rank (rank 1 corresponds to the top-left result); column (6) shows estimates for a dummy for whether the product is in the top 10 results. Column (1) through (5) apply various checks to the rank regression: column (1) only has search spell fixed effects; column (2) adds observable product characteristics derived from the search results page; column (3) adds the major brand dummy; column (4) excludes products in special carousels (such as "Highly Rated" and "Amazon's Choice"); and column (5) restricts attention to searches returning at least one Amazon brand. Throughout, standard errors are clustered at the search level. ***=0.001, **=0.01, *=0.05, .=0.10.

available), product quality (as proxied by consumer reviews), price, and delivery speed and fees.

Table 2 presents the coefficient estimates for the Amazon brand dummy and the sponsored dummy. Column (1) confirms, that within a search spell, sponsored products have lower rank, i.e., they are closer to the top of the page. In particular, being sponsored pushes a product up by 7 positions on average, an 18% increase in prominence. Amazon brands are also given more prominence. Without controlling for observable characteristics, the Amazon brand coefficient is nearly as large as the effect of sponsoring, and the two coefficients cannot statistically be distinguished from one another.

Column (2) reports the results including all characteristics observable to us as controls. The sponsored dummy coefficient only slightly changes, whereas the Amazon brand coefficient gets closer to zero, changing from -6.16 to -3.84. This change suggests that Amazon branded products tend to have characteristics, such as high ratings and low prices, that organically push them to the top of the page. Nonetheless, the Amazon brand dummy remains a strong predictor of rank, roughly 60% as large in magnitude as the sponsored coefficient. Estimates for other features are not presented, but typically go in the expected direction: low prices, fast delivery, and high ratings tend to increase product prominence.

The results so far suggest that Amazon brands are given additional prominence in search results that cannot be explained by other observable features such as prices, ratings, or delivery times. However, this may simply be due to other characteristics that we cannot observe. One possibility is that consumers like recognizable brands. To control for that, we generate an indicator for whether a product is a "major brand." We use a simple procedure to manually code recognizable brands that appear frequently in our search data,[9] which detects 3.9% of search results as carrying a major brand. We end up with 86 major brands, which include well-known brands like Adidas, Band-Aid, Duracell, Heinz, Oral-B, Pampers, and Ziploc. Column (3) in Table 2 adds the major brand dummy to our baseline regression. The major brand coefficient is of the expected sign, but only a fraction of the size of the Amazon brand coefficient and statistically indistinguishable from zero.

We also consider specifications that exclude products in special carousels (column (4)) and constrain searches to those returning at least one Amazon brand (column (5)). Finally in column (6), we use a dummy variable for whether the product is shown in the top 10 positions as the outcome of interest. We choose the top 10 positions as an outcome of interest since it roughly corresponds to the sponsored banner at the top of the page and the first two rows of standard search results. All specifications shown, as well as a number of additional checks, including specifications with interaction terms and machine learning approaches, indicate that carrying an Amazon brand is a meaningful predictor of greater prominence in search. The effect of Amazon brands tends to be 30% to 60% as large as the effect of sponsoring.

4 Conclusion

Our results, based on actual consumer searches, confirm existing (anecdotal or audit-based) evidence that Amazon brands are more prominently displayed in search results, above and beyond observable characteristics such as delivery speed and ratings; and above and beyond other major brands sold on Amazon.[10] Although we observe a rich set of product characteristics, there are other factors we do not observe, such as click and purchase rates, that may justify the higher ranking Amazon brands receive. Finally, our findings do not necessarily imply that consumers are hurt by Amazon brands' position in search results. We are currently conducting further research to study the effects of Amazon brands and its ranking policies on consumer welfare.

[9]Specifically, we focus attention on the first word of product names that appear in at least 25 search results or at least 10 searches. Three research assistants were asked to flag each word as a major brand that they recognize. If two out of the three research assistants flag a word, we consider it a major brand. We manually reviewed the list to validate the brands.

[10]Our findings of self-preferencing are conditional on the endogenous set of search results that are obtained by users. We do not investigate other channels of self-preferencing that may occur when the platform chooses whether products are presented in a given search.

References

Chen, Nan, and Hsin-Tien Tsai. 2022. "Steering via Algorithmic Recommendations." Working paper.

Dubé, Jean-Pierre. 2022. "Amazon Private Brands: Self-Preferencing vs Traditional Retailing." Working paper.

Farronato, Chiara, Andrey Fradkin, and Chris J. Karr. 2023. "Webmunk: A New Tool for Studying Online Consumer Behavior." White paper.

Gutierrez, German. 2022. "The Welfare Consequences of Regulating Amazon." Working paper.

Jeffries, Adrianne, and Leon Yin. 2021. "Amazon Puts Its Own 'Brands' First Above Better-Rated Products – The Markup." `https://themarkup.org/amazons-advantage/2021/10/14/amazon-puts-its-own-brands-first-above-better-rated-products`.

Lam, H Tai. 2022. "Platform Search Design and Market Power." Working paper.

Lee, Kwok Hao, and Leon Musolff. 2022. "Entry into Two-Sided Markets Shaped by Platform-Guided Search." Working paper.

Raval, Devesh. 2022. "Steering in One Click: Platform Self-Preferencing in the Amazon Buy Box." Working paper.