A Framework for Detection, Measurement, and Welfare Analysis of Platform Bias*

Imke Reimers
Cornell University and ZEW

Joel Waldfogel
University of Minnesota, NBER, ZEW

March 28, 2024

Abstract

Regulators are responding to growing platform power with curbs on platforms' potentially biased exercise of power, creating urgent needs for both a workable definition of platform bias and ways to detect and measure it. Platform search rankings present an important mechanism for possible self-preferencing; and we develop a simple equilibrium framework in which consumers choose among ranked alternatives, while the platform chooses product display ranks based on product characteristics and prices. We define the platform's ranks to be biased if they deliver outcomes that lie below the frontier that maximizes a weighted sum of seller and consumer surplus. This framework leads to two bias testing approaches, which we compare using Monte Carlo simulations, as well as data from Amazon, Expedia, and Spotify. We then illustrate the use of our structural framework directly, producing estimates of both platform bias and its welfare cost. Policies allowing researchers access to platform data would allow easy implementation of our approach in contexts important to policy makers.

*We are grateful to Senay Sokullu and Chiara Farronato, who served as discussants of the paper, and to audiences at the University of Minnesota, the Toulouse Digital Economics Conference, Indiana University, Queens University, Hong Kong University, the Jornadas de Economia Industrial, Wake Forest University, the Penn State-Cornell conference, the Toulouse online seminar on the economics of platforms, and Cambridge University, who provided helpful comments.

Introduction

Increasingly powerful platforms with interests in the products they sell are facing regulatory scrutiny for giving their own products preferential treatment relative to those of other suppliers, a practice known as "self-preferencing."[1] While self-preferencing can operate through various channels, the Digital Markets Act (DMA) specifically forbids "gatekeeper" platforms from engaging "in any preferential treatment in ranking." Prohibiting self-preferencing sounds simple and appealing, but its definition and measurement are not straightforward since it is not clear what rankings platforms' own products *should* receive in the absence of bias.[2] Opinions vary widely on whether "platform bias" is even a problem meriting regulatory attention. Some observers (e.g., Dubé, 2022) liken it to ubiquitous consumer-friendly store brands. Others, such as Senator Elizabeth Warren and the Indian antitrust authorities, are worried about large market shares and favor outright bans on Amazon's sales of its own products.[3] Either way, the advent of new regulations has created a pressing need for ways to detect, measure, and evaluate the welfare consequences of biased rankings in general and self-preferencing in particular.

We attempt to facilitate the policy evaluation that new regulations will require with a simple equilibrium model that gives rise to a workable definition of platform ranking bias. The model pairs consumer demand for platform-ranked products with the platform's choice of (potentially biased) rankings. We define unbiased rankings as those giving rise to a welfare frontier between maximal consumer and maximal producer surplus. A platform obtaining, say, higher commissions from the sale of some subset of the products on offer

[1] For example, the EU's Digital Markets Act forbids gatekeepers from giving preferential treatment to their own products; and the proposed American Innovation and Choice Online Act would forbid platforms from preferencing "the products, services, or lines of business of the covered platform operator over those of another business user on the covered platform." See `https://www.congress.gov/bill/117th-congress/senate-bill/2992/text`.

[2] Peitz (2022) and Peitz (2023) discuss the challenge of interpreting self-preferencing under the DMA.

[3] See `https://www.nytimes.com/2019/03/08/us/politics/elizabeth-warren-amazon.html` and `https://www.nytimes.com/2020/01/13/technology/amazon-bezos-india-antitrust.html`.

(for example its own products) might rank those products "too high," increasing its own revenue at the expense of overall seller and consumer surplus. Accordingly, we define bias as preferential treatment of platform-favored products which delivers outcomes interior to the welfare frontier.

We put our theoretical framework to three uses. First, we use the framework to produce a workable definition of platform bias: Bias exists when the platform ranks one set of products too high relative to the interests of consumers and sellers. Second, we use the framework to give a theoretical foundation to the ways in which researchers might detect and measure bias. These include both "conditioning on observables" (COO) and "outcome-based" (OB) approaches. In the COO approach, one regresses platform ranks on a platform indicator and controls and measures self-preferencing with the platform rank differential. The OB approach infers bias from differential post-ranking sales outcomes for platform and non-platform products assigned the same rank by the platform. Third, we discuss and illustrate direct estimation of a structural model derived from our framework, which supports not only detection and quantification of bias but also its welfare analysis.

While data requirements for *detecting* platform ranking bias are relatively light, implementing our structural model imposes a heavier burden, requiring data on platform rankings, sales, product characteristics, and the causal purchase consequences of rankings. Although this is a tall data order for researchers lacking cooperation of platforms, these are data that regulators could easily obtain. Moreover, the EU's Digital Services Act (DSA) has provisions allowing "vetted researchers" access to data, and regulators can access these data for enforcement purposes.[4] Moreover, we illustrate our approach with various useful, if imperfect, datasets available to us as researchers without inside access to platform data. This demonstrates the practical promise of our approach.

Our paper proceeds in five sections after the introduction. Section 1 provides background

[4] See `https://www.eu-digital-services-act.com/Digital_Services_Act_Article_40.html`.

on regulatory developments necessitating ways of quantifying self-preferencing, as well as the relevant academic literatures. Section 2 defines bias using a simple equilibrium model of consumer demand and platform ranking choice. This model can be estimated directly and gives rise to tests for, and measures of, platform bias. Section 3 discusses the relationship between our theory and the empirical bias detection approaches. We discuss advantages and challenges of each approach through the lens of our framework; and we present a Monte Carlo simulation demonstrating the possible advantages of the OB over the COO approach. Section 4 describes the data we use to illustrate our approaches, based on Amazon's Kindle Daily Deal pages, Expedia hotel searches, and Spotify's New Music Friday rankings. Section 4 also implements the two bias detection approaches using these platform data. In Section 5, we estimate the structural model using data on Amazon and Expedia (where we observe prices as well as other necessary variables). The approach delivers estimates of rank bias, platform preference for consumer vs seller surplus, and the welfare cost of biased rankings.

The paper offers four takeaways. First, our theoretical model delivers a simple definition of platform bias. Second, both the model and Monte Carlo simulations highlight challenges with the COO approach to detecting bias when platform products have unobserved attributes affecting demand. By contrast, the OB approach is robust to the unobservables problem. Third, we find that the OB and COO tests deliver different, and sometimes conflicting, results across our three empirical contexts. Fourth, the structural model delivers meaningful differences in platform attitudes toward consumers and sellers, bias, and welfare cost across contexts, corresponding intuitively to the descriptive findings.

1 Background

1.1 Policy context

Antitrust authorities around the world are now implementing or contemplating restrictions on retail platforms that would prevent them from giving preference to their own products. For example, under the EU's Digital Markets Act (DMA), which was implemented in 2022 and came fully into effect in March of 2024, "the gatekeeper should not engage in any form of differentiated or preferential treatment in ranking on the core platform service... ...in favour of products or services it offers itself." Moreover, the determinants of its rankings should be "generally fair and transparent."[5] Under the proposed US American Innovation and Choice Online Act (AICOA), it would be unlawful for a platform to "preference the products, services, or lines of business of the covered platform operator over those of another business user on the covered platform in a manner that would materially harm competition."[6] The Federal Trade Commission's 2023 suit against Amazon is motivated in part by Amazon's practice of "biasing [its] search results to preference Amazon's own products over ones that Amazon knows are of better quality."[7] Competition authorities in other countries are also concerned about self-preferencing among online platforms. For example, India forbids Amazon from directly selling its own products.[8]

The canonical problem that these policies seek to address is a platform ranking decision, for example when a platform chooses an ordering of products on a promotional page, or ranked search results. Self-preferencing is present when a platform's own products (or some other group of products the platform is suspected of favoring) obtain a better ranking or

[5] See `https://www.consilium.europa.eu/media/56086/st08722-xx22.pdf`.

[6] See `https://www.congress.gov/bill/117th-congress/senate-bill/2992/text`.

[7] See `https://www.ftc.gov/news-events/news/press-releases/2023/09/ftc-sues-amazon-illegally-maintaining-monopoly-power`.

[8] See `https://www.outlookindia.com/business/explained-why-is-competition-commission-of-india-probing-amazon-news-194362`.

page position than is appropriate for those products. Although researchers have begun to create intuitive tests for bias – see Section 1.2 – the definition of bias is not clear. Given that prohibitions on self-preferencing are, or will be, in force in many places, there is a pressing need for both a definition of bias and a way to measure its consequences. Finally, while platform behavior has in general been difficult for researchers to study, the Digital Services Act has provisions allowing "vetted researchers" access to platform data to conduct studies of the compliance of large platforms with the new regulations.[9]

1.2 Relevant literature

This paper is relevant to three strands of literature. First, it is relevant to theoretical work exploring reasons why platforms might bias their rankings, including Armstrong and Zhou (2011), Hagiu and Jullien (2014), Parker et al. (2020), De Corniere and Taylor (2019), Bourreau and Gaudin (2022), and to the effects of self-preferencing on outcomes (Zou and Zhou, 2023). Our work is also relevant to work on platform decisions about whether to sell their own products (Hagiu and Wright, 2015; Anderson and Bedre-Defolie, 2021; Hagiu et al., 2022).

Second, our paper is closely related to work testing for platform bias. Some research demonstrates bias in specific contexts, such as Amazon's "frequently bought together" recommendations or Amazon's buy box (Chen and Tsai, 2019; Edelman, 2011; Raval, 2022; Cure et al., 2022; Hunold et al., 2020). Other work attempts to measure platform bias in search rankings directly, by regressing platform search rankings on product observables and indicators for platform products (Jürgensmeier and Skiera, 2023; Farronato et al., 2023; Aguiar et al., 2021). We discuss these conditioning on observables (COO) approaches in

[9]Article 40 of the DSA states that "providers of very large online platforms or of very large online search engines shall, within a reasonable period, as specified in the request, provide access to data to vetted researchers." See `https://www.eu-digital-services-act.com/Digital_Services_Act_Article_40.html`. See also `https://www.brookings.edu/articles/platform-data-access-is-a-lynchpin-of-the-eus-digital-services-act/`, as well as Husovec (2023), Edelson et al. (2023), and Leerssen et al. (2021).

some detail in Section 3.

Third, a growing literature uses structural approaches to analyze market power at major platforms such as Amazon. While they do not test for bias per se, they do find results of interest. Lee and Musolff (2021) find that Amazon is likely to favor its own products; but consumers find those products appealing, raising questions about whether Amazon's display rankings reflect bias. Lam (2021) shows that counterfactual random product orderings would be less favorable to the platform than organic orderings, but that also leaves open the question of whether organic search results are biased. While not about rankings, Gutierrez (2021) explores consequences of other Amazon choices for consumer and supplier welfare. Our paper complements these studies by presenting a framework that allows for an explicit definition, and measurement, of biases in rankings.

2 Model

This section introduces a model of consumer choice, platform rankings, and the resulting surplus measures for consumers, sellers, and the platform. The model has two parts. First, consumers confront ranked product lists. They maximize their utility by choosing among the platform's ranked options and, by extension, whether to purchase at all. Second, given consumer preferences, the platform chooses how to rank the products to advance the platform's objectives, which may involve delivering surplus to buyers, sellers, and the platform itself. The combined behaviors of consumers and the platform then give rise to the outcomes of interest, which are the (potentially biased) rankings and their welfare consequences for consumers and sellers.

The model allows us to characterize efficient solutions, i.e., platform rankings that lead to a Pareto frontier running between maximal consumer and seller surplus. The model's supply side allows for deviations from the frontier if the platform is biased and favors one set of products (potentially its own) over others.

2.1 Consumer demand

A consumer who patronizes a platform confronts lists of products ranked by the platform. These lists may arise as responses to search queries or may simply reflect ranked orders in which platforms promote products. We refer to a rank-ordering of products as R.

The consumer's probability of purchasing product j depends on the product's underlying quality (δ_j^0, which itself depends on product characteristics and the product's price) and the ranking the platform assigns to it (r_j). One could use a variety of demand models to get mappings from the product qualities $\{\delta_j^0\}$ and the ranking R to quantities, revenue, and consumer surplus. These include various logit variants, as well as search models (Ursu, 2018; Ursu et al., 2023) and models of choice with limited information (Goeree, 2008; Abaluck and Adams-Prassl, 2021). Given our data and our goal of illustrating the framework in a simple way, we adopt a logit approach.

In the model, consumer i chooses among J products on the ranked list (and the outside option), based on each product's rank-independent quality δ_j^0 and its ranking in the search order, r_j. The consumer's utility for product j when ranked at r_j is given by:

$$u_{ij} = \delta_j^0 + \gamma r_j + \epsilon_{ij},$$

where δ_j^0 is the rank-independent quality of product j, reflecting the consumers' evaluation of product j's characteristics, including its price. The term γr_j embodies the causal impact of product j's promotional rank position on its utility (and purchase probability).[10] The term ϵ_{ij} reflects consumer i's idiosyncratic taste for product j. When ϵ_{ij} follows an extreme value distribution, this is the plain logit; and product j's market share when ranked r^{th} is

[10]Note that δ_j^0 may itself be related to the product's rank r_j, for example if the platform assigns ranks in part based on unobservable product attributes.

given by

$$s_j(r) = \frac{e^{\delta_j^0} e^{\gamma r_j}}{1 + \sum e^{\delta_j^0} e^{\gamma r_j}} = \frac{e^{\delta_j}}{1 + \sum e^{\delta_j}}. \tag{1}$$

2.2 Supply: the platform ranking decision

The platform has J products to present to consumers, so the platform's problem is to choose among $J!$ possible rank orderings. This is a difficult combinatoric problem, given the dozens of products usually under consideration.[11] The ranking that the platform chooses could serve the interests of consumers, sellers, or the platform itself. It is helpful to divide the platform ranking problem into two parts: a) where to locate on a welfare frontier between consumer and seller surplus (how to balance the interests of consumers and sellers), and b) how much to bias the rankings, which would move the solution away from the frontier.

Our setup has three primitives. First, each product has two characteristics, a rank-independent utility δ_j^0 and price p_j. The third primitive is the causal impact of the rank on the purchase propensity and utility: The further the platform ranks a product below the top position, the greater the proportionate reduction in a product's utility and sales.

The welfare frontier requires product rankings that maximize weighted combinations of consumer surplus (CS) and collective seller surplus; and the ranking that maximizes CS is easy to derive. Consumers benefit most from the highest-utility products, and worse platform ranks reduce utility by a monotonically increasing proportion. Hence, consumer surplus is maximized by placing products with the highest rank-independent utility at the best ranks. This is clearly visible in the plain logit, where the surplus that consumers derive from the choice set is given by

$$CS = \frac{M}{\alpha} \ln \left(1 + \sum_{j \in J} e^{\delta_j^0} e^{\gamma r_j} \right). \tag{2}$$

Because $e^{\gamma r_j}$ decreases in ranks when $\gamma < 0$, CS is maximized by ranking products in declin-

[11] This problem is also addressed in Compiani et al. (2021), who develop a model of platform rankings to maximize consumer surplus and profits.

ing order according to rank-independent mean utility δ_j^0. This "consumer-centric" ranking provides one endpoint on the welfare frontier between consumer and seller surplus, where the frontier is the downward-sloping region of the relationship between maximal consumer and seller surplus.

The ranking that maximizes sellers' expected surplus, which delivers the other welfare frontier endpoint, is more complicated to derive. Here, we make two assumptions for simplicity. First, we assume that prices are set by product suppliers prior to the ranking decision.[12] Second, for exposition, we assume zero marginal costs and therefore that the price equals per-unit variable profits.[13] Then, in the plain logit, expected per-consumer seller surplus (or, equivalently, revenue) from product j, when ranked according to R, is given by

$$\pi_j(r_j, R) \equiv \frac{p_j e^{\delta_j^0} e^{\gamma r_j}}{1 + \sum_{j \in J} e^{\delta_j^0} e^{\gamma r_j}}. \tag{3}$$

One intuitive potential strategy for maximizing collective producer surplus is to rank products according to $p_j e^{\delta_j^0}$, the rank-independent part of the numerator. The challenge with this approach is that the ranking R affects both the numerator and the denominator of Equation (3): Any re-ordering that increases the sum of the numerators also can increase the denominator, so that the ratio – revenue – does not necessarily rise. However, in Appendix Section A.1, we show that ranking according to $p_j e^{\delta_j^0}$ produces a very accurate approximation to seller surplus maximization; and we proceed with this approach.[14]

[12] Waldfogel (2024) finds that the Amazon search ranks of Amazon-brand products became 10 rank positions less favorable following Amazon's designation as a "gatekeeper" in September of 2023. Despite the large change in Amazon's ranking algorithm, the prices of Amazon-brand products did not change relative to other products. This provides justification for treating pricing as exogenous in the present exercise.

[13] Two points bear mention. First, if marginal costs are observed, one can simply replace prices with per-unit variable profits throughout the model. Second, many interesting products (e.g., ebooks) have zero marginal costs, so that prices equal per-unit variable profits.

[14] Appendix Section A.1 provides conditions under which this ranking rule maximizes seller surplus, and we explore whether permutations from our proposed ranking (by switching adjacent rank positions) generate higher surplus in one of our empirical contexts. Around our proposed revenue-maximizing rankings, we find higher seller surplus in only 0.16 percent of cases, and in these cases, the surplus from ranking by $p_j e^{\delta_j^0}$ is within 0.001 percent of the surplus we obtain with the permutation. We find no cases in which permutations

Maximizing the individual components provides two endpoints of the welfare frontier. Ranking by $e^{\delta_j^0}$ (or by its logarithmic transformation δ_j^0) maximizes CS, while ranking by $p_j e^{\delta_j^0}$ maximizes seller surplus. Taking a logarithm of $p_j e^{\delta_j^0}$, we obtain a welfare frontier between consumer and seller surplus by ranking products according to an index that weights these two terms:

$$I_j' = \kappa_1 \ln(p_j) + \kappa_2 \delta_j^0, \tag{4}$$

for varying combinations of κ_1 and κ_2. The index has intuitive special cases. For example, if $\kappa_1 = \kappa_2$, the resulting ranking maximizes revenue. If $\kappa_1 = 0$ and $\kappa_2 > 0$, then the resulting ranking maximizes CS. The relative sizes of κ_1 and κ_2 indicate the relative value that the platform attaches to buyers and sellers.

In addition to balancing the interests of consumers and sellers, a platform may have other objectives, for example if it obtains additional benefits from selling some products rather than others. Define $\mathbb{1}_j$ as an indicator for platform-preferred products. Then the index underlying platform rankings can be augmented as

$$I_j = \kappa_1 \ln(p_j) + \kappa_2 \delta_j^0 + \psi \mathbb{1}_j + \varepsilon_j, \tag{5}$$

which we term the platform's supply function. If additional factors (besides $\ln(p_j)$ and δ_j^0) affect the platform's chosen rankings, these rankings will lead to interior departures from the consumer-seller frontier. We propose to measure the welfare effect of biases from the ordinal relationship between ranks and the explanatory variables in Equation (5). If we estimate $\psi \neq 0$, then we have evidence of differential ranking treatment of platform products. Furthermore, we can debias the rankings by solving the model with $\psi = 0$.[15]

deliver both higher revenue and consumer surplus.

[15] We include an error term (ε_j) because the parameterization will not fit the data perfectly. As a result, the model's characterization of actual rankings will deviate from the observed values. In our empirical applications, we show that the model's characterization of actual rankings is very similar to observed rankings.

2.3 Discussion of our bias definition

For an intuitive definition of bias, consider a case in which the platform sells its own products alongside those of suppliers and gets a proportionate share c_j of revenue from each product j. Then the platform will maximize its own revenue by ranking products according to $c_j p_j e^{\delta_j^0}$. If c_j is larger for the platform's own products, this ranking will reduce total revenue (and therefore the total proceeds going collectively to suppliers and the platform) relative to the welfare frontier. When the platform commission rate is not constant across products, platform profit maximization produces bias that gives rise to an inefficiency.[16]

Having said this, we recognize that it is customary for retailers with "store brands" to privilege their own products. Retailers are, after all, interested in their own proceeds and not total seller surplus. For example, the Apple Store sells *only* Apple products, and grocers and drugstores commonly feature their store-brand products prominently alongside third-party suppliers' name-brand products. Few would argue that these practices are objectionable (e.g., Dubé, 2022).

Yet, the extent to which a retailer can engage in self-preferencing may depend on its market share. A large platform without much competition would face little market discipline against self-preferencing. A retailer facing competition, on the other hand, might be limited in its ability to self-preference. And, indeed, critics of platform self-preferencing make a distinction between large platforms and other retailers. For example, under the Digital Markets Act, the EU has designated six services as "gatekeepers" that are forbidden from self-preferencing.[17]

We view our framework as a tool for defining platform bias and measuring its extent in platform ranks. Whether bias is sufficiently harmful to welfare to warrant its prohibition is a separate question, and our framework may provide useful input into such analysis.

[16] Appendix Section A.2 discusses the relationship between commissions and bias when marginal costs are positive.

[17] See `https://ec.europa.eu/commission/presscorner/detail/en/ip_23_4328`.

3 Comparing bias-testing approaches

In our theoretical framework, bias enters through the platform supply function in Equation (5) when $\psi \neq 0$. There are two broad ways to test for bias; and both approaches, along with their strengths and weaknesses, are interpretable through our theoretical framework. One – the conditioning on observables (COO) approach – is to regress display rankings on controls and a platform indicator, and to interpret the coefficient on the platform indicator as bias. A second, outcome-based (OB), approach asks whether platform and non-platform products achieve different levels of ex-post success, conditional on the ranks the platform assigns them. In this section, we motivate both tests, and we compare their effectiveness using Monte Carlo simulations.

3.1 Conditioning on observables tests

The supply function in Equation (5) is an ordinal rank index: By construction, products at better ranks would have higher values of I_j. Although rankings are based only on the ordinal information in I_j, researchers using COO tests implicitly treat the index as cardinal. They estimate the ranking/supply function directly by regressing ranks on factors relevant to product appeal as well as the platform indicator $\mathbb{1}_j$. However, because the terms in the supply function are not all directly observable, regressions may take the following form:

$$r_j = X_j\lambda + \alpha p_j + \psi \mathbb{1}_j + \nu_j, \tag{6}$$

where X_j includes observable product characteristics. Then, researchers interpret the estimated ψ as a measure of platform bias. This, for example, is the approach of Jürgensmeier and Skiera (2023) and Farronato et al. (2023).

Our theoretical framework points out both the appeal and potential shortcomings of this approach. Given cardinality and linearity assumptions, if the control variables X_j contain all relevant determinants of δ_j^0, then the coefficient ψ in the regression in Equation (6) measures bias accurately. In particular, if $\psi < 0$, then ranks of platform products are ψ positions lower (better) than the controls warrant, indicating that the platform is biased in favor of platform products.

However, if one cannot adequately characterize rank-independent mean utility with X_j, then ψ may reflect not only the platform's exercise of bias but also demand-based reasons for a product's ranking. For example, if the platform's own products are appealing to consumers beyond what is captured in X, then a negative coefficient on the platform indicator would reflect a combination of possible bias and desirable, unobserved product characteristics affecting rankings.[18]

3.2 Outcome-based tests

Outcome-based tests for bias may also be viewed through the lens of our supply function I_j. In the absence of bias, $\psi = 0$, so that $\kappa_1 \ln(p_j) + \kappa_2 \delta_j^0$ would be equal for both platform and non-platform products at the same rank r_j. Rearranging terms, this indicates that we could test for bias by estimating

$$\delta_j^0 = \mu_r - ({}^{\kappa_1}/_{\kappa_2}) \ln(p_j) + ({}^{\psi}/_{\kappa_2}) \, \mathbb{1}_j + \nu_j, \tag{7}$$

where μ_r is a rank fixed effect, and the coefficient on $\mathbb{1}_j$ reflects bias. While δ_j^0 is not directly observable, we can proceed if we can observe q_j. This is because, conditional on rank, $\ln(q_j)$

[18] This is the platform analog to a host of familiar social science problems in which researchers seek to measure unwarranted disparity. Examples include measurement of discrimination in labor markets and unwarranted variation in criminal penalties (Klepper et al., 1983). Researchers pursuing those questions have long recognized the challenges of the conditioning on observables approaches, and those challenges are present in platform contexts as well.

is proportional to δ_j^0.[19] Hence, a regression of log-quantities on rank dummies, $\ln(p_j)$, and $\mathbb{1}_j$ delivers a test for rank bias. This test works as long as the causal effect of rank on sales operates the same for platform-preferred and non-preferred products.[20]

The OB approach, as also implemented in Aguiar et al. (2021), frees us from both the need for functional form assumptions on I_j and the need to observe all of the product characteristics that explain δ_j^0. However, the OB approach brings the additional need to observe the outcome affected by the rank (the quantity sold for each product).

3.3 Monte Carlo simulation

A Monte Carlo simulation intuitively illustrates the possible tradeoff in using the OB vs the COO approaches. Suppose that the platform observes variables X and Z, which are predictive of rank-independent sales success q^0:

$$q^0 = \beta X + \tau Z + \epsilon.$$

Assume further that, because of causal effects of ranks on sales, realized sales quantities depend on ranks assigned according to

$$q = e^{\gamma r} q^0$$

For the simulation, we draw X and Z from standard normal distributions. The variable X is observed by the researcher, while Z is not. Moreover, Z is potentially correlated with an indicator $\mathbb{1}$ for platform-owned products: We draw Z and a latent variable D

[19] For example, in logit, $\delta_j = \ln(s_j) - \ln(s_0)$. First, $s_j = q_j/M$ (where M is market size). Second, s_0 (the market share of the outside good) is constant across products j in J, so that $\ln(q_j)$ is proportional to δ_j. Hence, conditional on rank, $\ln(q_j)$ is proportional to δ_j^0.

[20] The OB approach can also produce a measure of bias in terms of rank positions. Suppose that $q_j^0 = A + Br_j + C\mathbb{1}_j$. Then C provides information about bias, and C/B measures the degree of bias in rank terms.

(determining whether a product is platform-owned) from a joint standard normal distribution with correlations ρ varying from -0.75 to 0.75, and we define the platform-owned product indicator $\mathbb{1} = 1$ if $D > 1$. As a result, about 15.9 percent of observations are platform-owned. We choose $\beta = \tau = 50$, and we draw ϵ from a normal distribution with standard deviation 100. Finally, we set $\gamma = -0.02$.

We abstract from seller surplus, and we instead assume that the platform optimizes on the total quantity sold (which maximizes CS) but may do so with bias: The platform may treat its own products as though they would sell more (or less) than they actually do. Hence, the platform ranks products according to an index based on the determinants of expected sales success, plus possible bias:

$$I = \beta X + \tau Z + \psi \mathbb{1},$$

where $\psi \neq 0$ indicates bias. We let ψ vary between -100 and 100. For each ψ and ρ, we simulate 500 iterations for 200 "markets," with 50 ranked products in each market.[21]

We are interested in the abilities of the COO and OB tests to correctly identify the direction of bias for varying levels of both true bias and the correlation between $\mathbb{1}$ and the unobserved rank determinants. To this end, we use our simulated data to perform the two tests for bias, each with linear and logarithmic specifications. First, we implement a COO test, regressing the rank r_j on just the (observable) X_j and the platform indicator:

$$r_j = \beta X_j + \psi' \mathbb{1}_j + \nu_j',$$

and we also employ $\ln(r_j)$ as the dependent variable. Second, we perform an outcome-based

[21] We consider a two-way grid of $\psi = -100, -75, \ldots, 100$ and $\rho = -0.75, -0.5, \ldots, 0.75$.

test with the following regression:

$$q_j = \mu_r + \psi^{''} \mathbb{1}_j + \nu_j^{''},$$

where μ_r denotes rank dummies, and we also use $\ln(q_j)$ as the dependent variable.

Each of these tests delivers one of three results: significant positive bias, significant negative bias, or a result indistinguishable from zero. In Figure 1, we compare the detected presence and direction of bias to the true, simulated bias. The correlation between Z and the latent D underlying the platform indicator varies along the figure's x-axes, and the underlying bias varies along the y-axes. The colors indicate the share of simulations correctly detecting the true bias, ranging from yellow (100 percent correct) to purple (zero percent).

The top panels show the COO tests, using rank and its logarithm as dependent variables. Both level and log specifications find the wrong answer rather frequently (about 22 percent of the time across all chosen ψ and ρ), especially when the correlation is substantial and when the true bias is small.[22] This is not surprising, as it is impossible to distinguish bias from a simple correlation between Z and $\mathbb{1}$. For example, in our setup, the platform-favored products would have an average ranking of 14 both in a simulation with a strong positive bias ($\psi = 75$, which is roughly a standard deviation of q_j) and no correlation between Z and $\mathbb{1}$, as well as with zero bias ($\psi = 0$) and a correlation ρ of 0.75.[23]

The bottom panels of Figure 1 report the OB tests from linear and logarithmic specifications. Compared to the COO test based on observable X, the OB approaches are more accurate. While the COO tests obtain the correct answer in 78 percent of cases, the OB approach is correct in 98 percent of cases in the linear specification (and 96 percent of cases

[22]When there is no bias, both COO tests find the correct answer only 13 percent of the time.

[23]Being able to condition on Z would largely solve the problem. In the linear specification, COO tests including Z deliver the correct answer every time when bias is present, and they correctly identify the absence of bias 93 percent of the time. The log specification, however, only correctly identifies the absence of bias 34 percent of the time, reflecting the importance of functional form.

using logs). The instances in which the OB test is "incorrect" reflect failures to detect small amounts of bias.[24] While the OB test never finds significant bias inconsistent with the true level, the COO test finds significant bias of the wrong sign in 28.6 percent of instances when the true bias is 1/3 of a standard deviation of q, and 14.3 percent of the instances when the true bias is 2/3 of a standard deviation. Our simulation reinforces our concerns about unobservables undermining COO tests and leads us to conclude that if one can observe outcomes and ranks, the outcome-based test would be preferred.

4 Data and descriptive evidence

Before describing the data we use to illustrate our approaches, it is helpful to outline what we need in a context, and in data on that context, for the analyses we envision. First, we require a context with a relevant form of possible bias, for example a platform that sells its own, or otherwise potentially favored, products alongside those of other suppliers. Second, the context must feature ranked product listings and display ranks that affect the products' sales. Third, to implement the COO approach for detecting rank bias, we also need detailed product characteristics, including prices. The OB approach does not require detailed product characteristics but instead requires direct measures of the product sales quantities that the ranks affect, along with prices. Finally, implementing the full welfare analysis requires all of the above, as well as a way to estimate a causal rank effect, for example with randomly assigned product rankings or products that appear repeatedly at different rank positions.

The datasets and contexts we are able to examine – from Amazon, Expedia, and Spotify – have some but not all of the features we would ideally have. Still, they allow illustration of our approaches. Our analyses of them emphasize the features we need to implement our approach.

[24] For example, a bias of 25 is roughly a third of a standard deviation of q.

4.1 Amazon Kindle Daily Deals

Each day, Amazon selects about 50 ebooks for their "Kindle Daily Deal" page. These ebooks are displayed at the site in ranked order, and the title list is also emailed to interested customers. This context has several strengths. First, a significant share of the titles are published by Amazon Publishing, making self-preferencing a possible concern. Second, while rankings are not randomized, the same titles are promoted at different ranks on different days, giving a plausible strategy for measuring rank effects. Finally, marginal costs of ebooks are the same (zero) for all products, so that prices directly reflect per-unit variable profits.

The data for this application are drawn from two sources. First, for each date between April 4 and July 12, 2022, we collected data on the titles promoted on the Kindle Daily Deal page, as well as the rank order of the promoted titles, directly from Amazon.[25] This is a total of 76 daily promotions and 3,738 promotional listings for 2,892 distinct titles. Our second source, Bookstat, provides a measure of daily Amazon sales (inferred from sales rankings) and prices for each of these ebook titles.[26] For each title, we also observe whether Amazon is the publisher and its sales during 2021.

Our main analyses make use of data for the day of the promotion and the following day, as the promotions affect sales for two days. For the day of the promotion and the following day, ebook sales average 138.9 for non-Amazon, and 63.7 for Amazon; and the price averages \$4.17 for non-Amazon books and \$2.26 for Amazon. Despite the apparent differences in popularity, Amazon ebooks are ranked highly. The average promotional ranking for Amazon books is 17.5, compared to an average of 26.5 for other books.

While the context is of great interest, the data here have some shortcomings. First, rather than being choice-level data as with Expedia (see below), the Amazon data are at the product level. Second, we have high-frequency quantity data, but they are inferred from sales

[25] See `https://www.amazon.com/b?node=6165851011` and `https://www.amazon.com/s?rh=n%3A6165851011&fs=true` for the current list.

[26] See `https://bookstat.com/`.

ranks rather than directly observed; and there is some question about the relative accuracy of sales estimates for Amazon vs non-Amazon products.[27]

4.2 Expedia hotel listings

Expedia is a site where consumers can search for, and book, travel. As part of a data mining competition in 2013, Expedia made available a dataset of hotel searches, including all of the product options presented to consumers and information on which (if any) hotel was chosen by the Expedia user.[28] The data include 399,342 hotel searches at Expedia during 2013. Of these, 277,797 were organic searches, and Expedia randomized the rank ordering of the hotels in the remaining 121,545 searches. Available hotel characteristics include the price, the star type, the average consumer rating of the hotel (on a five point scale), a property location score, and whether the hotel is part of a chain. The dataset includes 8,624,781 listings, for an average of 21.6 listings per search. Among the organic search results, 91.6 percent of searches result in a booking (reflecting over-sampling of successful searches), while only 12.5 percent of random-order searches produce a booking.[29] Finally, 64.5 percent of the listings are for chain hotels.

In many respects, these resemble ideal data. We see the products presented to individuals, as well as the product ranks, prices, and characteristics; and we also see which products consumers chose. Moreover, because of randomization of product rank orderings, it is easy to estimate causal rank effects. Despite these significant advantages, the context has four shortcomings. First, Expedia does not own hotels, so there is no direct possibility of self-

[27] It is not clear whether the data can accurately distinguish between consumption from Kindle Unlimited borrowing and a la carte sales. Virtually all of the Amazon Publishing books are available through Kindle Unlimited, while only some of the other books are, raising a question about interpreting differences in reported sales as bias.

[28] The data, at `www.kaggle.com/c/expedia-personalized-sort/data`, were made available through the International Conference on Data Mining (ICDM 2013) and Kaggle.com. Ursu (2018) uses the data to document causal impacts of rank positions.

[29] See `www.kaggle.com/c/expedia-personalized-sort/data`.

preferencing in this context. Instead of studying self-preferencing, we explore the possibility of bias with respect to whether a hotel is part of a chain. Second, the data cover 2013, which is by now of essentially historical interest. Third, the over-sampling of successful organic searches may bias the estimates. Finally, hotel services are not digital products, and they have marginal costs that we do not observe. We instead assume that marginal costs of hotel stays are similar for both chain and non-chain hotels.

4.3 Spotify New Music Friday

Each week, Spotify creates country-specific lists of 50 new songs for their New Music Friday playlists. As Aguiar and Waldfogel (2021) document, Spotify appears to rank the chosen songs in descending order of expected promise. Here, we use the top 20 New Music listings for 26 countries during 2017 (Aguiar et al., 2021) to study potential bias according to whether music is released by major labels. While Spotify does not produce music, the major record labels have substantial ownership stakes in Spotify, and observers have raised concerns about possible bias in this dimension.

We observe usage data for the top 200 songs by day and country, and we have a total of 18,489 listings. In our data, 62.1 percent of the listings are for major-label songs, and 43.0 percent of the major-label songs appear among their countries' top 200 songs, compared with 24.2 percent for independent songs. In addition to having the ranked lists of promoted songs, the Spotify data also include detailed song characteristics that allow us to implement COO and OB bias tests.[30]

A few additional features of the Spotify context merit discussion. First, while ranks are not randomized, we see the same songs at different curator promotional ranks in different countries, as in the Amazon context. Second, because these are digital products, marginal

[30] For each song we observe the artist's previous-year streams, some musical characteristics of the songs such as danceability, beats per minute, speechiness, etc., and whether the song is produced by a major record label.

costs to the upstream sellers are zero. Finally, because users buy subscription access rather than purchasing songs, there is no product-specific revenue; and welfare analysis would require a different framework.

4.4 Comparing bias tests using platform data

While the Monte Carlo exercise in Section 3.3 provides one basis for comparison, we also implement the COO and OB tests using the data from Amazon, Spotify, and Expedia. The goal of this exercise is to gauge the similarity of bias test results from the contrasting approaches.

We summarize our results in Table 1. Panel A reports bias tests using the Amazon data. The first column reports a COO regression of a product's log rank on an Amazon dummy and the product's log price. The regression also includes (unreported) controls for star ratings, the number of reviews, and the title's sales during 2021, the year before the daily promotions we study. The coefficient on the Amazon dummy, -0.269 (se=0.025), indicates that Amazon books receive ranks that are 23.6 percent[31] better (lower) than their non-platform counterparts, conditional on observable characteristics.

Column (2) reports an outcome-based (OB) test using the log of realized quantity sold as the dependent variable and rank fixed effects, price, and the Amazon indicator as explanatory variables. The -0.696 (0.042) coefficient on the Amazon dummy indicates that, conditional on the rank the platform assigns to them, Amazon books sell 50 percent less ($= e^{-0.696} - 1$). Hence, this OB test also indicates bias in favor of Amazon books.

So far we have one measure – from the COO approach – of the bias in terms of rank positions. The OB approach can also deliver a measure of rank bias if we parameterize the relationship between rank-independent sales and the rank for each product type. We calculate rank-independent sales by subtracting the causal rank effect from log sales: $\ln(q_j^0) =$

[31] $23.6 = 100 \times \left(e^{-0.269} - 1\right)$

$\ln(q_j) - \gamma \ln(r_j)$, where we use the column (3) estimate of γ. The fourth column reports a regression of rank-independent log sales on log rank, log price, and the Amazon indicator. The platform coefficient, -0.714 (0.042), divided by the log rank slope coefficient (-0.190) delivers an estimate of rank bias: Absent platform bias, the promotional ranks of platform products would be $e^{-0.714/-0.190} = 42.9$ times higher (worse). In effect, this means that platform products ranked 2^{nd} or worse should instead have been ranked last among the 50 products. While the OB and COO tests give the same direction of bias, the magnitudes are very different in our Amazon estimates.[32]

The second panel of Table 1 reports tests using Expedia data. The first column, using a COO approach, shows that chain hotels (the group whose potential platform bias we explore) receive rankings that are 0.67 units higher (worse) than they should be. The second column, using an OB approach and a linear probability model, echoes the finding of anti-chain bias, showing that chain hotels are 0.8 percentage points (0.02) more likely to be booked, conditional on search rank. Columns (3) and (4) quantify the OB results in terms of rank positions. The third column reports a regression of the probability of booking a hotel on rank using the randomized sample, giving a causal rank coefficient of -0.00031 that allows calculation of a rank-independent booking probability. Finally, Column (4) – using the rank-independent booking probability as the dependent variable – indicates that the platform puts chain hotels 2.1 (=0.00710/0.00342) rank positions worse than what they deserve. While still small, this is roughly three times the rank bias implied by the COO approach.

The third panel of Table 1 reports results for Spotify. The first column, using a COO test, shows bias in favor of major-label songs. Songs from major record labels are ranked 1.2 positions better than their observable characteristics warrant.[33] The second column, using an

[32] When we specify the models in levels rather than logs, the COO approach gives a rank bias of 5.2 positions in favor of Amazon products, while the OB approach delivers bias of 42.9 rank positions.

[33] The control variables include the artist's song streams in the previous year, the song's beats per minute, as well as other characteristics known as: valence, energy, accousticness, instrumentalness, danceability, liveness, and speechiness.

OB test, finds the opposite direction of bias. Conditional on rank, major-label songs stream 44.0 percent more than other songs, indicating bias against major-label songs.[34] Columns (3) and (4) deliver the OB estimate of rank bias. Using Column (4), dividing the major-label coefficient (0.44) by the rank coefficient (-0.106) gives a rank bias of 4.2 positions. The Spotify context is noteworthy in that COO and OB tests detect biases of opposite signs.

The comparisons have a few implications. First, the COO and OB tests sometimes give different answers for whether there is bias. Second, even when the tests give the same direction of bias, the magnitudes differ substantially. These results reinforce the a priori concerns – and the Monte Carlo results above – about the reliability of the COO approach.

5 Model estimates and simulation

This section presents estimates of the demand and supply models for Amazon and Expedia (the two contexts with product prices), as well as structural estimates of rank bias and its welfare cost.

5.1 Actual and debiased rankings at Amazon

We estimate demand for ebooks on the Kindle Daily Deals pages using a plain logit approach. That is, we estimate

$$\ln(s_j) - \ln(s_0) = x_j\beta + \alpha p_j + \gamma r_j + \xi_j \tag{8}$$

The vector x_j includes characteristics of title j: whether it is an Amazon product and title sales in the previous year. Although we have suppressed time subscripts, the estimates include both the day of the promotion and the following day, as well as a fixed effect for the following day. The first column of Table 2 reports results from this demand model.

[34]Because we observe streams only for songs appearing among the top 200 for a country and day, we must impute the usage for songs not making the top 200. We assign them the lowest observed usage.

The coefficient on the log sales rank is -0.405 (0.027), capturing both the tendency to place more appealing books at better ranks and a causal impact of ranks on sales. The Amazon coefficient is negative ($-0.669, se = 0.045$), indicating that consumers attach lower utility to Amazon products, even conditional on rank.

The supply model requires a rank-independent mean utility measure, $\delta_j^0 = \delta_j - \gamma r_j$. We predict δ_j from Equation (8), where we use a causal estimate of the rank effect γ from a separate regression including a title fixed effect, identifying a γ of -0.335 (0.103) from within-title variation in its Kindle Daily Deals rank.[35] We then use the calculated $\hat{\delta}_j^0$ to estimate the supply model using a rank-order logit on

$$r_j = \kappa_1 \ln(p_j) + \kappa_2 \hat{\delta}_j^0 + \psi \mathbb{1}_j + \epsilon_j.$$

For ease of interpretation, Column (2) of Table 2 first reports results from a linear regression with promotion-day fixed effects. Both price and utility terms have negative coefficients, indicating that both higher utility and higher prices give rise to better (lower) ranks. Platform products, too, receive better ranks, indicating bias. Column (3) of Table 2 reports the rank-order logit results (Hausman and Ruud, 1987), using promotion-day combinations as groups. The results are normalized so that positive coefficients correspond to better (lower) ranks. This model, which relaxes the cardinality assumption, also indicates that both higher prices and rank-independent mean utilities are associated with better rankings. The positive price coefficient reflects the platform's concern about revenue; and the result that the utility coefficient exceeds the price coefficient indicates that the platform is also concerned about CS. Finally, the ψ coefficient provides the direct test for bias, and it is consistent with platform self-preferencing.

The estimated supply function allows calculation of both debiased and model-actual

[35] Of the 2,835 titles in the dataset, 564 appear on the list more than once.

promotional ranks. Ranks reflecting the estimated bias ψ produce Amazon product ranks averaging 17.5, whereas debiased ranks (based on a recalculated supply function that sets $\psi = 0$) generate an average Amazon product ranking of 41.1. This indicates that the actual rankings are biased by 23.6 rank positions in favor of Amazon. Panel A of Figure 2, showing kernel density plots of the actual and debiased rank distributions of Amazon products, illustrates the bunching of Amazon products near the tail with the debiased ranks.[36] This result is consistent with the finding in the OB test indicating a rank bias of 42.9 positions. By contrast, the linear version of the COO test gave a bias of just 5.2 rank positions.

5.2 Expedia

Table 3 reports model estimates for Expedia. The first two columns report results from our demand model, which we estimate as a nested logit model in 2 parts.[37] First, consumers choose among the displayed options for the cases in which a hotel is chosen. Second, they decide whether to book a hotel based on the inclusive value from the first part. Column (1) reports results for the lower nest using a conditional logit model and the data with organic display ranks. The rank coefficient (-0.129, se=0.0004) captures both rank effects and underlying quality differences across search rankings. The coefficients appear reasonable: Consumers attach higher utility to hotels with more stars and better user ratings, and they dislike higher prices. Finally, consumers attach additional utility worth roughly \$16.64 (=0.127/0.00763) to chain hotels.

The upper nest is the decision of whether to book a hotel among those listed. For this, we estimate a logit model relating the binary booking choice to the inclusive value from the lower nest. Column (2) of Table 3 is estimated on both the randomized and organic search results and shows that consumers are more likely to book when the choice set is more

[36]Figure 2 also shows that the model matches observed rankings well: Observed Amazon product ranks also average 17.5.

[37]Our implementation of the nested logit follows Chapter 4 of Train (2009).

appealing.[38] The coefficient on the inclusive value – which gives the nested logit substitution parameter σ – is 0.535 (0.007).

The next two columns report the supply function estimates relating search rankings to prices and the rank-independent utilities δ_j^0.[39] Column (3) reports a linear regression of the rank on the supply parameters, along with hotel search fixed effects, and Column (4) reports the rank-ordered logit, using searches as groups. Both specifications show that the platform gives better ranks to hotels that are more appealing to consumers and to hotels with higher prices. The utility coefficients exceed the price coefficients, indicating that rankings are chosen with a concern for consumers and not simply revenue. Finally, the platform assigns worse ranks to chain hotels, reflecting an apparent bias against chains, although the bias coefficients are small in comparison with the price and utility terms.[40]

Panel B of Figure 2 illustrates the degree of chain bias at Expedia using kernel density plots for the chain hotels' actual rankings, as well as the model's versions of actual and debiased rankings. The figure shows two things. First, the model's actual chain rankings, which average 12.2, are close to those observed in the data, which average 12.3. Second, the debiased search ranks – setting the bias parameter ψ to zero – average 11.4 for chain hotels, indicating a bias of 0.8 rank positions against chains. Recall that the other bias tests also find small amounts of bias. The COO approach produced bias of 0.6 while the OB test delivered 1.8.

[38]This is likely driven in part by the oversampling on organic searches resulting in purchase.

[39]To obtain δ_j^0, we first estimate the conditional logit model on the randomized sample, which gives us a causal ranking parameter γ of -0.0799 (0.0013). We then calculate a rank-independent utility by subtracting γr_j from the utility function associated with each product's conditional choice probability.

[40]The chain bias we identify here is analogous to the forgone profit opportunity from incorrectly-chosen ranks in Compiani et al. (2021). In our context, deviations from surplus maximization that serve the platform's interest illustrate our notion of platform self-preferencing.

5.3 The platform's objective and welfare estimates

We model the welfare frontier as the locus of CS and revenue combinations arising from rankings of the unbiased supply function in Equation (4): $I'_j = \kappa_1 \ln(p_j) + \kappa_2 \delta^0_j$, for various values of κ_1 and κ_2. The frontier extends from maximal revenue, when $\kappa_1 = \kappa_2$, to the revenue associated with maximal CS, when $\kappa_1 = 0$, so the relevant welfare frontier segment is downward-sloping. The left and right panels of Figure 3 depict the average welfare frontiers for the Amazon and Expedia contexts, respectively, relative to their CS and revenue maxima. The figures also show the model depictions of the actual and debiased choices. The welfare cost of bias is the difference in CS and revenue between the model's debiased and actual rankings. The debiased location on the frontier shows the platform's attitude toward consumers versus sellers.

Table 4 reports both welfare results for the Amazon and Expedia examples. The top panel reports our measures of the welfare losses from the bias in actual rankings. Actual rankings here forgo 3.3 percent of the CS in debiased rankings, and they forgo 5.3 percent of the corresponding revenue. The bottom panel shows the platforms' dispositions toward consumers and sellers. After debiasing (setting $\psi = 0$), Amazon's rankings deliver a surplus combination that achieves 98.9 percent of maximal CS and 90.2 percent of maximal revenue. The platform's debiased choice sacrifices proportionally more revenue than consumer surplus, indicating relatively high platform concern for consumers. Both results are also illustrated in Panel A of Figure 3. These estimates, showing that Amazon ranks as if it valued CS more than seller revenue, are similar to previous findings (Gutierrez, 2021; Reimers and Waldfogel, 2017) that Amazon's pricing attaches high value to consumers.

The second column of Table 4 presents results for Expedia. Because we estimate only negligible chain bias, debiasing Expedia's rankings has a very small effect on welfare. Expedia's bias in the treatment of chain hotels during 2013 sacrificed only 0.06 percent of CS and 0.23 percent of revenue. Expedia's debiased rankings deliver a point on the welfare frontier that achieves 97.6 percent of maximal CS and 87.8 percent of maximal revenue, indicating – as in the Amazon example – platform concern for consumers. See also Panel B of Figure 3.

It should be emphasized, again, that these results are more illustrative of the approach than they are informative about actual platform bias.

6 Conclusion

Growing concern about platforms' potentially biased exercise of power creates a pressing need for tools and frameworks for evaluating platform bias. This paper provides a few steps in this direction. First, we develop a simple theory of demand and supply in platform contexts. Consumers choose among platform-ranked products, and platforms choose ranks that balance the interests of consumers, sellers, and the potentially biased platform itself.

This framework provides a definition of bias – platform rankings that create deviations from the welfare frontier – and a way to think about various tests for bias. We implement two such tests, a conditioning on observables (COO) and an outcome-based (OB) bias test, in three contexts. The regression results, along with a Monte Carlo study, reinforce our concerns about the COO approach relative to the OB approach. We then implement the equilibrium framework directly, using illustrative data from Amazon and Expedia. The structural model provides estimates of rank bias, and it allows us to measure the welfare cost of platform bias and to estimate the platform's balancing of consumer and seller interests.

The data requirements for implementing our approach are in principle simple but in practice difficult without data internal to platforms. Data that regulators could obtain for enforcement would allow relatively straightforward implementation of our approach in meaningful, contemporary contexts. While the Digital Services Act includes a provision to allow researchers access to data for studying "systemic risk" to the European Union, it is unclear whether platform self-preferencing falls into that category.[41] In the meantime, this paper provides a framework for analyzing and detecting platform bias. Some issues may merit further attention. For example, we have taken suppliers' prices to be given and fixed even as we counterfactually eliminate platform bias. While prices at Amazon are apparently invariant to exogenous changes in search ranks, it is nevertheless possible that prices are endogenous to platform bias more generally. Still, we hope that our framework provides a useful input into the analysis of platform bias as regulators forbid the practice.

[41] See `https://algorithmic-transparency.ec.europa.eu/news/faqs-dsa-data-access-researchers-2023-12-13_en`.

References

Abaluck, J. and A. Adams-Prassl (2021): "What do consumers consider before they choose? Identification from asymmetric demand responses," *The Quarterly Journal of Economics*, 136, 1611–1663.

Aguiar, L. and J. Waldfogel (2021): "Platforms, power, and promotion: Evidence from Spotify playlists," *The Journal of Industrial Economics*, 69, 653–691.

Aguiar, L., J. Waldfogel, and S. Waldfogel (2021): "Playlisting favorites: Measuring platform bias in the music industry," *International Journal of Industrial Organization*, 78, 102765.

Anderson, S. P. and Ö. Bedre-Defolie (2021): "Hybrid platform model," .

Armstrong, M. and J. Zhou (2011): "Paying for prominence," *The Economic Journal*, 121, F368–F395.

Bourreau, M. and G. Gaudin (2022): "Streaming platform and strategic recommendation bias," *Journal of Economics & Management Strategy*, 31, 25–47.

Chen, N. and H.-T. Tsai (2019): "Steering via algorithmic recommendations," *Available at SSRN 3500407*.

Compiani, G., G. Lewis, S. Peng, and W. Wang (2021): "Online search and product rankings: A double index approach," *Available at SSRN 3898134*.

Cure, M., M. Hunold, R. Kesler, U. Laitenberger, and T. Larrieu (2022): "Vertical integration of platforms and product prominence," *Quantitative Marketing and Economics*, 1–43.

De Corniere, A. and G. Taylor (2019): "A model of biased intermediation," *The RAND Journal of Economics*, 50, 854–882.

Dubé, J.-P. (2022): "Amazon Private Brands: Self-Preferencing vs Traditional Retailing," *Available at SSRN 4205988*.

Edelman, B. (2011): "Bias in search results: Diagnosis and response," *Indian JL & Tech.*, 7, 16.

Edelson, L., I. Graef, and F. Lancieri (2023): "Access to Data and Algorithms: For an Effective DMA and DSA Implementation," *CERRE, March*.

Farronato, C., A. Fradkin, and A. MacKay (2023): "Self-preferencing at Amazon: evidence from search rankings," Tech. rep., National Bureau of Economic Research.

Goeree, M. S. (2008): "Limited information and advertising in the US personal computer industry," *Econometrica*, 76, 1017–1074.

Gutierrez, G. (2021): "The welfare consequences of regulating Amazon," Tech. rep., mimeo, New York University.

Hagiu, A. and B. Jullien (2014): "Search diversion and platform competition," *International Journal of Industrial Organization*, 33, 48–60.

Hagiu, A., T.-H. Teh, and J. Wright (2022): "Should platforms be allowed to sell on their own marketplaces?" *The RAND Journal of Economics*, 53, 297–327.

Hagiu, A. and J. Wright (2015): "Marketplace or reseller?" *Management Science*, 61, 184–203.

Hausman, J. A. and P. A. Ruud (1987): "Specifying and testing econometric models for rank-ordered data," *Journal of econometrics*, 34, 83–104.

Hunold, M., R. Kesler, and U. Laitenberger (2020): "Rankings of online travel agents, channel pricing, and consumer protection," *Marketing Science*, 39, 92–116.

Husovec, M. (2023): "How to Facilitate Data Access under the Digital Services Act," *SSRN: https://ssrn.com/abstract=4452940*.

Jürgensmeier, L. and B. Skiera (2023): "Measuring Fair Competition on Digital Platforms," *arXiv preprint arXiv:2303.14947*.

Klepper, S., D. Nagin, and L.-J. Tierney (1983): "Discrimination in the criminal justice system: A critical appraisal of the literature," *Research on sentencing: The search for reform*, 2.

Lam, H. T. (2021): "Platform search design and market power," *Job Market Paper, Northwestern University*.

Lee, K. H. and L. Musolff (2021): "Entry into two-sided markets shaped by platform-guided search," *Job Market Paper, Princeton University*.

Leerssen, P. et al. (2021): "Platform research access in Article 31 of the Digital Services Act: Sword without a shield?" .

Parker, G., G. Petropoulos, and M. W. Van Alstyne (2020): "Digital platforms and antitrust," *Available at SSRN 3608397*.

Peitz, M. (2022): "The Prohibition of Self-Preferencing in the DMA," *CERRE Issue Paper*.

——— (2023): "How to apply the self-preferencing prohibition in the DMA," *Journal of European Competition Law & Practice*.

Raval, D. (2022): "Steering in One Click: Platform Self-Preferencing in the Amazon Buy Box," *Unpublished manuscript*.

Reimers, I. and J. Waldfogel (2017): "Throwing the books at them: Amazon's puzzling long run pricing strategy," *Southern economic journal*, 83, 869–885.

Train, K. E. (2009): *Discrete choice methods with simulation*, Cambridge university press.

Ursu, R., S. Seiler, and E. Honka (2023): "The Sequential Search Model: A Framework for Empirical Research," .

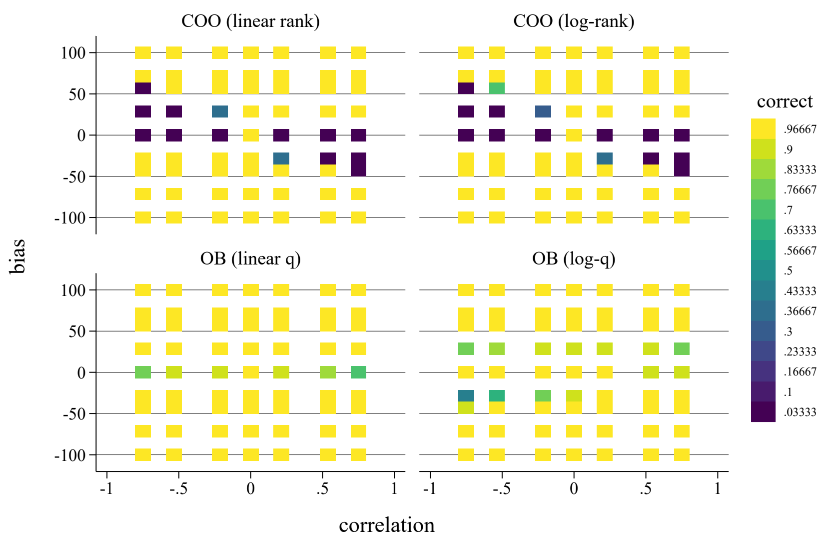
Ursu, R. M. (2018): "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions," *Marketing Science*, 37, 530–552.

Waldfogel, J. (2024): "Amazon self-preferencing in the shadow of the Digital Markets Act," .

Zou, T. and B. Zhou (2023): "Self-preferencing and Search Neutrality in Online Retail Platforms," .

7 Figures and Tables

Figure 1: Monte Carlo comparison of COO and OB tests

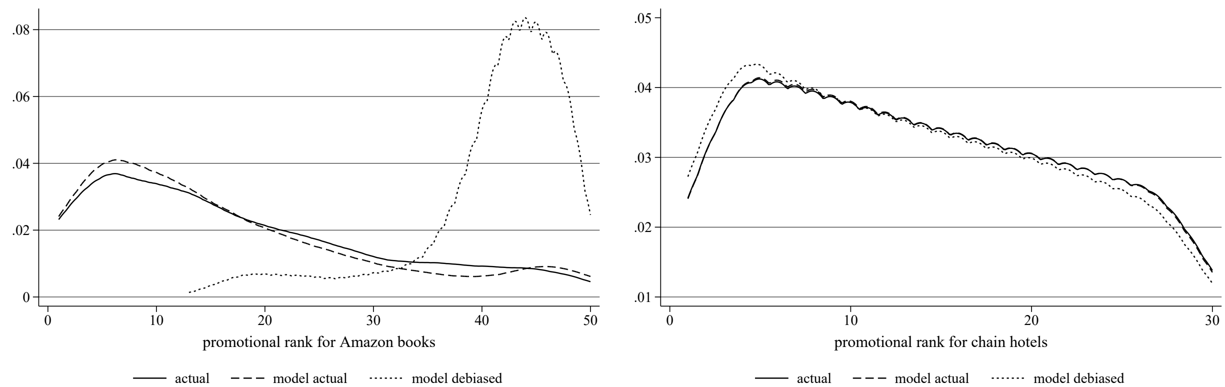


Notes: The figure reports results of Monte Carlo exercises comparing how often the true bias (or its absence) is detected by conditioning on observables (COO) and outcome-based (OB) tests. The vertical axis shows the true degree of bias, where a bias of 100 corresponds to about 4/3 of a standard deviation of the realized quantities. The horizontal axis shows the correlation between the determinant of the platform indicator D and unobserved determinants of expected sales Z. The upper panels describe COO tests based on regressions of platform ranks (and log ranks) on the observable X and the platform dummy D. The bottom panels describe OB tests based on regressions of realized sales (and log sales) on rank fixed effects and the platform dummy D. Yellow indicates a high probability of finding the true bias direction, while darker colors indicate lower probabilities. Each cell in each figure reflects percentages based on 500 draws.

Figure 2: Actual and debiased ranks

Panel A: Amazon Kindle Daily deals

Panel B: Expedia hotel searches

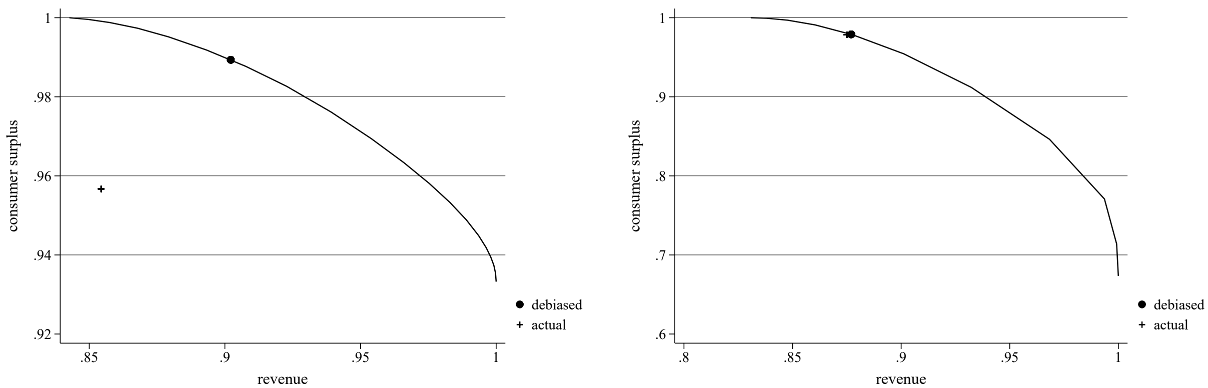


Note: The figures show kernel density plots of rankings for the potentially platform-favored products. The left figure describes Amazon Publishing books in the Kindle Daily Deal pages, and the right figure describes chain hotel rankings in Expedia searches. The solid lines show observed ranking distributions, the dashed lines show distributions of our model's actual rankings, and the dotted lines show our model depictions of debiased rankings.

Figure 3: Actual and debiased welfare outcomes

Panel A: Amazon Kindle Daily deals

Panel B: Expedia hotel searches



Note: The figures show the welfare frontiers arising from rankings that maximize weighted sums of revenue and consumer surplus for Amazon (left) and Expedia (right), as percentages of maximum revenue (x axes) and maximum consumer surplus (y axes). The points along the frontiers are outcomes associated with the model depictions of debiased rankings, showing the platforms' relative dispositions toward consumers vs sellers. The deviation between "actual" points (+), interior to the frontiers, and the debiased points shows the welfare cost of the bias in actual rankings.

Table 1: Bias test comparisons

Panel A: Amazon Kindle Daily deals

	COO	Outcome-based		
	(1) ln rank	(2) ln quantity	(3) ln quantity	(4) rank-ind ln(q)
Amazon	-0.269***	-0.696***		-0.714***
	(0.025)	(0.042)		(0.042)
ln rank			-0.266**	-0.190***
			(0.106)	(0.018)
ln price	0.049***	-0.306***	-0.523***	-0.304***
	(0.012)	(0.020)	(0.023)	(0.020)
Observations	6796	6826	6617	6826

Panel B: Expedia hotels

	COO	Outcome-based		
	(1) rank	(2) $\mathbb{1}$(buy)	(3) $\mathbb{1}$(buy)	(4) rank-ind pr
chain	0.57107***	0.00821***		0.00719***
	(0.00728)	(0.00021)		(0.00021)
rank			-0.00040***	-0.00407***
			(0.00001)	(0.00001)
price	0.01236***	-0.00017***		-0.00018***
	(0.00005)	(0.00000)		(0.00000)
Observations	5971587	5971587	2568446	5971587

Panel C: Spotify New Music Monday

	COO	Outcome-based		
	(1) rank	(2) log streams	(3) log streams	(4) rank-ind ls
major label	-1.212***	0.385***		0.440***
	(0.087)	(0.036)		(0.037)
rank			-0.126***	-0.106***
			(0.003)	(0.003)
Observations	18233	18489	13467	18489

Note: Column (1) of each panel implements a COO test with a regression of rank on (unreported) observables and a platform-preferred product indicator: "Amazon," "chain," and "major label." Column (2) implements the OB approach with regressions of sales outcomes on rank fixed effects and the platform indicator. Column (3) measures the causal rank effect using product fixed effects or (for Expedia) order randomization. Column (4) reports regressions of rank-independent outcomes on ranks, allowing quantification of the rank bias based on the OB test. All models are estimated using OLS.

Table 2: Amazon demand and supply estimates

	demand	supply	
	(1) logit	(2) linear	(3) r.o. logit
price	-0.0391***		
	(0.00435)		
ln rank	-0.405***		
	(0.0272)		
Amazon product	-0.669***	-64.20***	12.74***
	(0.0452)	(0.663)	(0.169)
ln daily pre-promo sales	0.0406*		
	(0.0226)		
ln price		-19.11***	3.815***
		(0.294)	(0.0592)
rank-indep mean util		-93.76***	19.26***
		(0.995)	(0.261)
Observations	6826	6826	6826

Note: Column (1) presents logit estimates of the demand for ebooks offered in the Kindle Daily Deal. "Amazon product" refers to a title published by Amazon. An observation is a promotion-day title, and we include two days of data for each promotion. Column (2) reports a linear regression of the platform-chosen rank on its log price and its rank-independent mean utility, along with the Amazon indicator. The regression includes promotion-day fixed effects. Column (3) reports the analogous specification using a rank-order logit model, with promotion days as groups and normalized so that positive coefficients deliver better ranks. Robust standard errors are reported in parentheses.

Table 3: Expedia demand and supply estimates

	(1) c logit	(2) logit	(3) linear	(4) r.o. logit
price	-0.00763***			
	(0.0000502)			
rank	-0.129***			
	(0.000374)			
# reviews	0.171***			
	(0.00304)			
stars	0.275***			
	(0.00355)			
chain	0.127***		2.296***	-0.623***
	(0.00524)		(0.00486)	(0.00123)
location score	0.0954***			
	(0.00270)			
inclusive value		0.535***		
		(0.00681)		
ln(price)			-8.552***	2.391***
			(0.00546)	(0.00181)
rank-indep mean util (δ_j^0)			-10.73***	3.638***
			(0.00319)	(0.00169)
Observations	6,196,924	397,720	6,640,113	6,640,113

Note: Columns (1) and (2) together provide nested logit estimates of demand. Column (1) reports conditional logit estimates on the choice of hotels, among the organic hotel searches. Column (2) reports a logit on the decision to book a hotel, and it is estimated across both randomized and organic searches. The coefficient on the inclusive value is the substitution parameter σ. Column (3) reports a linear regression of the Expedia search rank on its log price and its rank-independent mean utility, along with the chain hotel indicator, using organic searches. The regression includes hotel search fixed effects. Column (4) reports the analogous specification using a rank-order logit, with searches as groups and normalized so that positive coefficients deliver better ranks. Robust standard errors are reported in parentheses.

Table 4: Welfare effects of bias

	Amazon	Expedia
welfare change from bias		
% ΔCS	-3.30	-0.057
% ΔREV	-5.30	-0.232
debiased point on frontier		
CS relative to CS^{max}	0.989	0.976
REV rel to REV^{max}	0.902	0.878

Note: This table shows results from the structural model, for Amazon's Kindle Daily Deal pages (left column) and Expedia's hotel search rankings (right column). The top panel shows CS and revenue forgone with the model's depiction of actual rankings relative to the debiased rankings. The bottom panel shows CS and revenue attained from the debiased rankings relative to the maximum levels of CS and revenue on the frontier. Because of data shortcomings discussed in the text, the results are best viewed as illustrative of the method.

A Appendix

A.1 Evaluating our proposed welfare frontier

This section discusses the accuracy of our proposed welfare frontier. Ranking products by rank-independent mean utility delivers CS maximization. The other endpoint of the welfare frontier is more complicated. Our proposed solution for maximizing revenue is to order products according to $p_j e^{\delta_j^0}$, a ranking we term $R^{\text{max Rev}}$.

Total revenue per consumer in the market is given by

$$\frac{\sum_{j \in J} p_j e^{\delta_j^0} e^{\gamma r_j}}{1 + \sum_{j \in J} e^{\delta_j^0} e^{\gamma r_j}}.$$

A switch in the ranking order away from $R^{\text{max Rev}}$ necessarily reduces the numerator of this formula. Hence, revenue can only *rise* if such a switch reduces the denominator proportionally more than it reduces the numerator.

This, in turn, occurs when the ratio of the changes to the numerator and denominator of the total revenue function exceed the per-consumer total revenue with our proposed ranking. Suppose product a has a higher rank-independent revenue than product b: $p_a e^{\delta_a^0} > p_b e^{\delta_b^0}$, so that our proposed ranking would entail $r_a < r_b$. Rearranging terms, revenue rises with a switch of the rankings of the two products a and b if:

$$\frac{p_a e^{\delta_a^0} - p_b e^{\delta_b^0}}{e^{\delta_a^0} - e^{\delta_b^0}} < \frac{\sum_{j \in J} p_j e^{\delta_j^0} e^{\gamma r_j}}{1 + \sum_{j \in J} e^{\delta_j^0} e^{\gamma r_j}}.$$

To see how violations are unlikely, it is helpful to consider a variant of this inequality replacing the per-person total revenue (the current RHS) with average revenue among products sold. This would be achieved by removing the 1 from the denominator. That inequality would hold any time the marginal revenue of the switch (the current LHS) fell short of the average price. The actual denominator of the per-person total revenue expression is much higher, substantially reducing the probability that the inequality holds (and hence that our proposed ranking could be improved).

We explore this with our data from Amazon. Permutations – switching adjacently ranked products – deliver higher revenue than our proposed solution in 6 of 3,800 cases. When reversals occurred, the $R^{\text{max Rev}}$ ranking delivered at least 99.99981% of the revenue of the alternative. Hence, our approach delivers virtually all of the maximal revenue that we calculate; and permutations of non-adjacent products produced no revenue increases.

Beyond exploring revenue violations, we can also look for permutations that deliver outcomes beyond the entire frontier. Our proposed welfare frontier arises from ranking products according to the weighted sum, $\kappa_1 \ln(p_j) + \kappa_2 \delta_j^0$. We use Amazon data to create a proposed frontier using an 11-point grid of weighted sums of (κ_1, κ_2) from (0, 1) to (0.5, 0.5), as depicted in Figure 3. To evaluate our proposed solution, we calculate how often permutations of adjacently-ranked products increase revenue and/or CS. For each rank position, promotion day, and location along the 11 grid points on the proposed frontier (41,800 observations in total), we calculate whether CS and/or revenue based on the permuted ranks exceeds the proposed frontier value. None of the permutations deliver a direct violation such that both CS and revenue lie beyond the pre-permutation proposed frontier.

A.2 Unbiased commissions with positive marginal costs

When selling third-party product j, the platform receives a commission c_j that is proportional to the price. The platform maximizes its revenue by ranking products according to $c_j p_j e^{\delta_j^0}$. When marginal costs are zero, so that the price reflects per-unit variable profits, a constant commission across products ($c_j = c \forall j$) would lead to the Pareto frontier and would reflect the absence of platform bias.

If marginal costs are positive, per-unit seller surplus is $v = p - mc$ and not simply the price. Hence, maximization of seller surplus would be achieved by ranking products according to $v_j e^{\delta_j^0}$ rather than $p_j e^{\delta_j^0}$. The platform would naturally achieve the welfare frontier if a constant commission τ were levied against v rather than p. Given that commissions are charged against prices, however, it is of interest to derive unbiased commissions for the case with positive marginal costs.

We can calculate a commission c_j levied against the price p_j that is equivalent to a proportional commission τ on v. Slight rearrangement of $c_j p_j = \tau(p_j - mc_j)$ gives

$$c_j = \tau \frac{p_j - mc_j}{p_j}.$$

That is, product j's PS-maximizing commission is proportional to the share of the price that is a markup over marginal cost. A platform facing these commissions and maximizing its own revenue would also maximize overall seller surplus.