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

Imke Reimers
Cornell University and ZEW

Joel Waldfogel
University of Minnesota, NBER, ZEW

October 1, 2023

DRAFT – COMMENTS WELCOME

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. 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. The EU's Digital Services Act's provision for researcher data access would allow easy implementation of our approach in contexts important to policy makers.

*We are grateful to Senay Sokullu, who served as a discussant of an earlier version 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, and Wake Forest 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] The paradigmatic example of the behavior that regulators seek to control is a platform's display ranking of products. While prohibiting self-preferencing sounds simple and appealing, its definition and measurement are not straightforward since it is not clear what rankings platforms' favored 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 platform bias in general and self preferencing in particular.

We attempt to assist this policy conversation with a simple equilibrium model that gives rise to a workable definition of platform 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 (for example its own products) might rank those products "too high," increasing its own revenue at the expense of overall seller and

[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 (2023) and Peitz (2022) 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`.

consumer surplus. Therefore, 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 seek to 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, measuring self preferencing with the platform coefficient. The OB approach infers bias from differential 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.

Data requirements for *detecting* bias are relatively light, whereas implementing our structural model imposes a heavier burden, requiring data on platform rankings, sales, product characteristics, and the causal purchase consequences of rankings. While this is a tall data order for researchers lacking cooperation of platforms, we note that the EU's Digital Services Act (DSA) has provisions allowing "vetted researchers" access to data. 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 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 the theory and 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 that correspond 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), implemented in 2022, "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."[4] 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."[5] 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."[6] Competition authorities in other countries are also concerned about self-preferencing among online platforms: India forbids platform sales of their own products.[7]

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 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.[8]

---

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

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

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

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

[8] 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).

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 the effects of self-preferencing on outcomes (Zou and Zhou, 2023). It is also relevant to work on platform decisions about whether to sell their own products (Anderson and Bedre-Defolie, 2021).

Second, it is closely related to work testing for platform bias. Some work 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 some detail in Section 3.

Third, a growing literature takes 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 this self-preferencing reflects 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. Our paper complements these studies by presenting a framework that allows for an explicit definition, and measurement, of bias.

2 Model

This section introduces a model of consumer choice, platform rankings, and the resulting surplus measures for consumers, sellers, and the platform. The basic model consists of 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 the Pareto frontier running between 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 platform-created ranked lists of products. These lists may arise as responses to search queries or may simply reflect ranked orders in which platforms promote products. We refer to the rank-ordering of products as R. For exposition, we use the logit functional form. Our illustrative applications use logit and nested logit.

Consumer i chooses among J products on the ranked list (and the outside option), based on product characteristics x_j, prices p_j, and each product's ranking in the search order, r_j.

The consumer's utility for product j when ranked at r_j is given by:

$$u_{ij} = x_j\beta + \alpha p_j + \gamma r_j + \gamma' r_j^0 + \xi_j + \epsilon_{ij},$$

where the outside good has utility 0, ξ_j represents unobserved product quality, r_j^0 is the initial rank of product j, and ϵ_{ij} is an extreme value error. Here, γ is the causal effect of rank on sales, while γ' reflects the systematic product quality variation across ranks that is not accounted for by x_j.[9] Define the mean *expected* utility of product j when ranked at r as $\delta_j(r) = x_j\beta + \alpha p_j + \gamma' r_j^0 + \gamma r$.[10] Even if product j were moved to a different rank r_j, the part of utility reflected by $\gamma' r_j^0$ would remain, while the part reflected by γr would change. We then define "rank-independent expected mean utility" as

$$\delta_j^0 \equiv x_j\beta + \alpha p_j + \gamma' r_j^0 = \delta_j(r) - \gamma r. \tag{1}$$

Product j's market share when ranked r^{th} is given by $s_j(r) = \frac{e^{\delta_j(r)}}{1+\sum e^{\delta_j(r)}}$. The way in which the products are ranked affects both consumer well-being and the propensity for consumers to purchase.

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. 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

[9] The causal effect of rank on utility γ may be interpreted as a reduced form characterization of search costs.

[10] The platform chooses a rank before realized quality, reflected in ξ_j, is known, so we set $\xi_j = 0$ to calculate expected $\delta_j(r)$.

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.

As a step toward defining the Pareto frontier, it is helpful to first consider two rankings that maximize consumer and seller surplus, respectively. 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 $\gamma < 0$, consumer surplus (CS) is maximized by ranking products in declining order according to rank-independent mean utility δ_j^0. This 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. Second, we assume zero marginal costs and therefore that the price equals per-unit variable profits.[11] Then, expected 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 solution is to rank products according to $p_j e^{\delta_j^0}$, which we could term "rank-independent expected seller surplus." The challenge with this approach is that the

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

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. Hence, it is not obvious *a priori* that ranking according to $p_j e^{\delta_j^0}$ maximizes total revenue. In Appendix Section A.1, we show that ranking according to $p_j e^{\delta_j^0}$ produces highly accurate estimates of maximum seller surplus; and we proceed with this approach.[12]

Maximizing the individual components provides two endpoints of the welfare frontier. Ranking by $e^{\delta_j^0}$, or by its monotonic logarithmic transformation δ_j^0 directly, maximizes CS, while ranking by $p_j e^{\delta_j^0}$ (or by $\ln(p_j) + \delta_j^0$) maximizes seller surplus. We obtain a welfare frontier between consumer and seller surplus by ranking products according to an index that is a weighted sum of these two terms,

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

for varying combinations of κ_1 and κ_2. This 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 values 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

$$I_j = \kappa_1 \ln(p_j) + \kappa_2 \delta_j^0 + \psi \mathbb{1}_j + \varepsilon_j, \tag{4}$$

[12] 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. We find higher seller surplus only in 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 deliver both higher revenue and consumer surplus.

which we term the platform's supply function. The presence of additional factors (besides $\ln(p_j)$ and δ_j^0) affecting the platform's chosen rankings gives rise to interior departures from the consumer-seller frontier. We propose to measure the welfare effect of biases by estimating the supply function in Equation (4) from the ordinal relationship between ranks and the explanatory variables. If we estimate $\psi \neq 0$, then we have evidence of bias. Furthermore, we can debias the rankings by solving the model with $\psi = 0$.[13]

2.3 Discussion of our bias definition

For an intuitive sense of our 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.[14] 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. In that sense, bias gives rise to an inefficiency.

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 social 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). However, 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 discipline against self-preferencing. A retailer facing competition, on the

[13] 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 distributions of model actual rankings are very similar to those for observed rankings.

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

other hand, might be limited in its use of self preferencing. And, indeed, critics of platform self-preferencing make a distinction between large platforms and other retailers. For example, the EU's Digital Services Act prohibitions on self-preferencing are directed specifically at 17 "very large service operators" and two "very large online search engines."[15]

We view our framework as a tool for measuring the extent of bias in a platform's ranking behavior. 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.

3 Comparing bias-testing approaches

In our theoretical framework, bias enters through the platform supply function in Equation (4) 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 ranks 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 in Monte Carlo simulations.

3.1 Conditioning on observables tests

The supply function in Equation (4) 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 and not the cardinal information in I_j, researchers using COO tests treat the index as both cardinal and proportional to I_j. They estimate the ranking/supply function directly

[15] See `https://ec.europa.eu/commission/presscorner/detail/en/IP_23_2413`.

by regressing ranks on factors relevant to the ranking 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:

$$r_j = X_j\lambda + \alpha p_j + \psi\mathbb{1}_j + \nu_j, \tag{5}$$

where X_j is a set of "controls" presumed to reflect all of the relevant determinants of δ_j^0. Then, the coefficient on the platform-preferred product indicator reflects just platform bias. This, for example, is the approach of Jürgensmeier and Skiera (2023) and Farronato et al. (2023), who measure bias at Amazon by regressing search rankings on a platform indicator and controls.

Our theoretical framework points out both the appeal and potential shortcoming of this approach. Given cardinality and linearity assumptions, if the control variables X_j adequately proxy for rank-independent mean utility and contributions to seller surplus p_j, then the coefficient ψ in the regression in Equation (5) measures bias. In particular, if $\psi < 0$, then ranks of platform products are 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 coefficient on the platform indicator would reflect a combination of possible bias and unobserved determinants of demand and therefore rankings.[16]

[16] 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.

3.2 Outcome-based tests

Outcome-based tests for bias, as implemented in Aguiar et al. (2021), may also be viewed through the lens of our supply function I_j. Two products with the same value of I_j would merit the same rank. 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. Rearranging terms, this indicates that we can test for bias by estimating

$$\delta_j^0 = \mu_r - (\kappa_1/\kappa_2) \ln(p_j) + (\psi/\kappa_2)\, \mathbb{1}_j + \nu_j, \tag{6}$$

where μ_r is a rank fixed effect, and the coefficient on $\mathbb{1}_j$ determines bias. Again, δ_j^0 is not directly observed, but in this approach, we observe the outcome variable with error, which does not bias our estimate of $\psi' = \psi/\kappa_2$. Here, if two products with the same δ_j^0 have the same expected sales q_j conditional on search ranks and prices p_j, then a regression of 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.

The OB approach frees us from the need for functional form assumptions on I_j and the need to control for product appeal, but it brings another data requirement. We need to observe the outcome affected by the rank (the quantity sold for each product), along with some of the variables that are also required by the COO approach: the price, the rank assigned by the platform, and whether the product is potentially platform-preferred.

The OB approach can also produce a measure of bias in terms of rank positions. After removing the causal effects of ranks, the rank-independent quantities can still vary with ranks, as indicated by the γ' term in Equation (1). Consider a variant on Equation (6) replacing the rank fixed effects with a linear term in rank. Rank bias would then be given by the ratio of the platform indicator coefficient – providing a measure of the bias in terms of the outcome – to the rank coefficient, which translates the outcome differential to a rank differential.

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 (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.

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 the index based on 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 100 iterations for 200 "markets," with 50 ranked products in each market.

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.[17] To this end, we use our simulated data to perform 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 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 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, relative to true 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 detecting the true bias, ranging from yellow (100 percent correct) to purple (zero percent).

[17] Varying both ψ and the correlation ρ between Z and the variable underlying $\mathbb{1}$, we consider a two-way grid of $\psi = -100, -75, \ldots, 100$ and $\rho = -0.75, -0.5, \ldots, 0.75$.

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), in particular when the correlation is substantial and when the true bias is small.[18] 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 ($\psi = 75$) and no correlation between Z and $\mathbb{1}$, as well as with zero bias ($\psi = 0$) and a correlation ρ of 0.75.[19]

The bottom panels of Figure 1 report the OB tests from linear and logarithmic specifications. In contrast 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 using quantity (and 96 percent of cases using log quantity). 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. Second, the context must feature ranked product listings and ranks that affect the products' sales.

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

[19] 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.

To implement the COO approach for detecting rank bias, we also need detailed product characteristics, including prices. The OB approach does not require 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.

It would be very desirable, for example, to see ranked choice sets presented to individuals, including product characteristics and prices, as well as the consumers' choices. It would be further advantageous to have a way to measure causal rank effects, 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.

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, about one sixth 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.

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 ranks rather than directly observed; and there is some question about the relative accuracy of sales estimates for Amazon vs non-Amazon products.[20]

The data for the 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.[21] This is a total of 76 daily promotions of 3,738 promotional listings. The promotional listings include 2,892 distinct titles because some titles are promoted on more than one day. Second, we have a measure of daily Amazon sales (inferred from sales rankings) and prices for each of these ebook titles, from Bookstat.[22] 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 each title, we also observe whether Amazon is the publisher and its sales during 2021.

For the day of the promotion and the following day, sales average 138.9 for non-Amazon, and 63.7 for Amazon, ebooks; 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 ranking of 26.5 for other books.

4.2 Expedia hotel listings in 2013

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.[23] The data include 399,342 hotel searches at Expedia during

[20] 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.

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

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

[23] 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

2013, among them 121,545 in which Expedia randomized the rank ordering of the hotels. 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.[24] 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-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 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 negligible 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 data analyzed in Aguiar et al. (2021), the top 20 New Music listings for 26 countries during 2017.

document causal impacts of rank positions.

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

The Spotify context has a number of features that merit description. First, Spotify does not produce music, so there is no threat of explicit self-preferencing. That said, the major record labels have substantial ownership stakes in Spotify, and observers have raised concerns about possible bias in favor of major-label music. Following Aguiar et al. (2021), we explore this possible dimension of bias. Second, the context lacks song-specific product prices, as users have subscription access. Third, while ranks are not randomized, we see the same songs at different curator promotional ranks in different countries, as in the Amazon context. Finally, because these are digital products, marginal costs to the upstream sellers are zero.

The Spotify data have strengths and weaknesses. In addition to having the ranked lists of promoted songs, we also observe detailed song characteristics that allow us to implement COO and OB bias tests.[25] While we do have usage data, these measures are available only for the top 200 songs by day and country, requiring us to impute the missing usage observations with the lowest-observed usage.

We have a total of 18,489 listings. Of these, 6,637 (35.9 percent) achieve sufficient streaming success to appear among a country's top 200 daily songs. 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.

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

[25] 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.

approaches.

We summarize our results in Table 1. The first panel 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[26] better (lower) than their non-platform counterparts, conditional on observable characteristics.

Column (2) reports an outcome-based 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 also delivers a measure of rank bias from the relationship between rank-independent sales and the rank, along with the platform effect. The fourth column reports a regression of rank-independent log sales ($\ln(q_j) - \gamma \ln(r_j)$) on log rank, log price, and the Amazon indicator.[27] The platform coefficient, -0.714 (0.042), divided by the log rank slope coefficient (-0.190), shows that 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$^{\text{nd}}$ 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 for Amazon.[28]

[26] $23.6 = 100 \times \left(e^{-0.269} - 1\right)$

[27] We present an estimate of γ in the third column, from a regression of log sales on log prices and log ranks, including a title fixed effect (-0.266 (0.106)).

[28] 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.

The second panel of Table 1 reports tests using Expedia data. The first column, using a conditioning on observables 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 outcome-based 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 using the randomized sample, giving a rank coefficient of -0.00031 with a causal interpretation. 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 conditional on observables 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.[29] The second column, using an outcome-based 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. 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. That is, the outcome-based approach indicates that the platform is biased by 4.2 rank positions against major-label music, whereas the COO test indicates a bias of 1.2 in its favor. 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

[29] 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.

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.[30] That is, we estimate

$$\ln(s_j) - \ln(s_0) = x_j\beta + \alpha p_j + \gamma r_j + \xi_j \tag{7}$$

The vector x_j includes characteristics of title j: whether it is an Amazon product, 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 next-day fixed effect. The first column of Table 2 reports results from this demand model. 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), indicating that consumers attach lower utility to Amazon products, even conditional on rank.

[30] Because our goal is the illustration of the framework rather than the estimates themselves, we take a simple approach to the problem. One might alternatively estimate a variety of more flexible models, including both richer parameterizations or search models, as in Ursu (2018).

The supply model requires us to obtain a rank-independent mean utility estimate, $\delta_j^0 = \delta_j - \gamma r_j$. We predict δ_j from Equation (7), and we obtain 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.[31] 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 day-promotion 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 effect of higher prices on rankings reflects the platform's concern about revenue; and the result that the price coefficient is significantly below the utility coefficient indicates that the platform is concerned about CS and not simply revenue. 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 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

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

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.[32] 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. 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. Hence, 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 level is the decision of whether to book a hotel among those listed.[33] For this, we estimate a logit model relating the binary choice to book 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 appealing.[34] 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

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

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

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

prices and the rank-independent utilities δ_j^0.[35] 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 price coefficients are smaller than the utility 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, suggesting small bias.

Panel B of Figure 2 illustrates the degree of chain bias at Expedia using kernel density plots for the chain hotels' ranks for 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 12.19, are close to the observed actual rankings, which average 12.30. Second, the debiased search ranks – setting the bias parameter ψ to zero – average 11.41 for chain hotels, indicating a bias of 0.78 rank positions against chains. The other bias tests also find small amounts of bias. The COO approach produced bias of 0.57 while the OB test delivered 1.77.

5.3 The platform's objective and welfare estimates

We model the welfare frontier as the CS and revenue combinations resulting from rankings of $\kappa_1 \ln(p_j) + \kappa_2 \delta_j^0$ (setting $\psi = 0$) for various values of κ_1 and κ_2, with $\kappa_1 \geq 0$ and $\kappa_2 > 0$. The frontier extends from maximal revenue when $\kappa_1 = \kappa_2$, to the revenue associate 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

[35] 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.

Expedia contexts, respectively, relative to their CS and revenue maxima. The Figures also show the model depictions of the actual and debiased choices.

These figures answer two questions of interest. First, they depict the disposition of the platform toward consumers versus sellers. We summarize this information, derived from the κ_1 and κ_2 coefficients in the previous section, using the ratios of debiased to maximal revenue and maximal CS, respectively. The figures also show the welfare cost of bias. We measure this based on the differences in CS and revenue between the model's debiased and actual rankings.

Table 4 reports both welfare results for the Amazon and Expedia examples. The top panel shows the platforms disposition 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 choice sacrifices proportionally more revenue than consumer surplus, indicating a concern for consumers and not simply revenue. Secondly, the bottom panel reports our measures of the welfare losses from the bias in actual rankings. Actual rankings here forgo 3.30 percent of the CS in debiased rankings, and they forgo 5.30 percent of the revenue resulting from debiased rankings. Both results are also illustrated in Panel A of Figure 3.

The second column of Table 4 presents results for Expedia. Their 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 and not just seller revenue. 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 0.057 percent of CS and 0.23 percent of revenue. 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.

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 the challenges faced by 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 allows inference on both the platform's balancing of consumer and seller interests, as well as the welfare cost of platform bias.

The data requirements for implementing our approach are in principle simple but in practice difficult without data internal to platforms. Yet, the Digital Services Act includes a provision to allow researchers access to data. Data obtained with such access would allow relatively straightforward implementation of our approach in meaningful contemporary contexts. In the meantime, this paper provides a framework for analyzing and detecting platform bias, although 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. It is possible that suppliers would endogenously change their prices if they faced different amounts of bias. Still, we hope that our framework provides a useful input into the analysis of platform bias as regulators forbid the practice.

References

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.*

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.

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

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*.

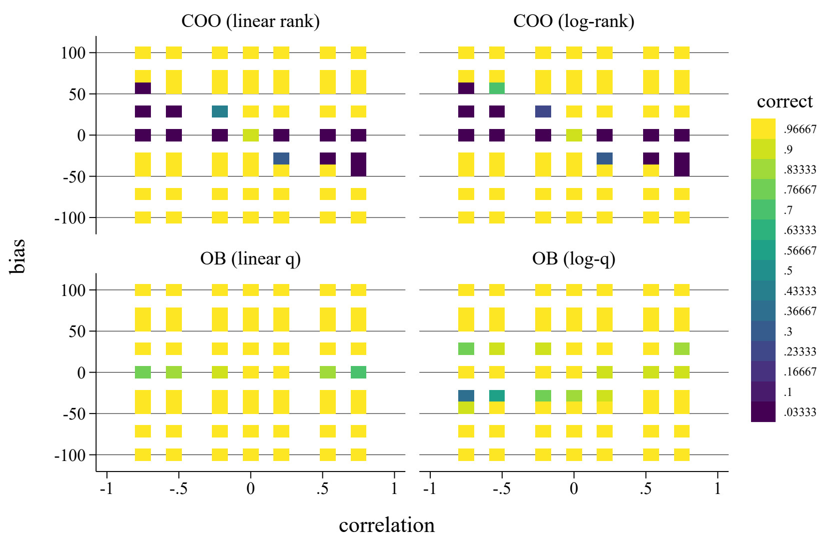
Train, K. E. (2009): *Discrete choice methods with simulation*, Cambridge university press.

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.

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

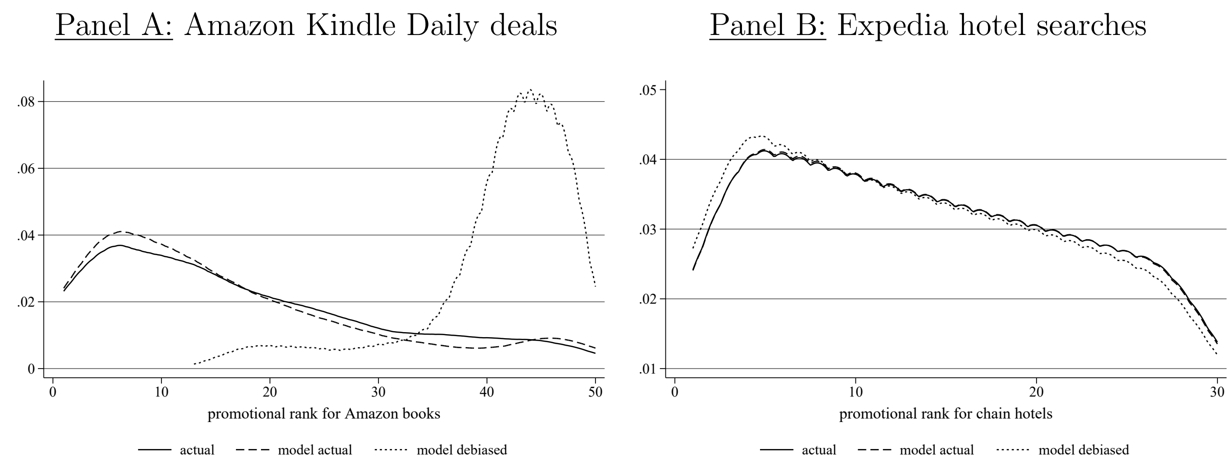
7 Figures and Tables

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 is detected by conditioning on observables (COO) and outcome-based (OB) tests. The vertical axis shows the true degree of bias, while 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.

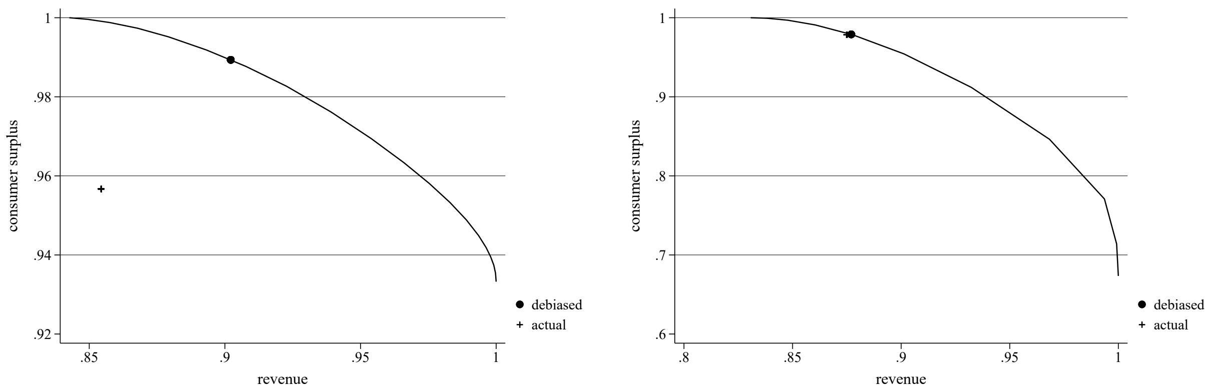
Figure 2: Actual and debiased ranks



Note: The figures show kernel density plots of rankings for the platform, or 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 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)	(2)	(3)	(4)
	ln rank	ln quantity	ln quantity	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)	(2)	(3)	(4)
	rank	$\mathbb{1}$(buy)	$\mathbb{1}$(buy)	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)	(2)	(3)	(4)
	rank	log streams	log streams	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 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 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 by OLS.

Table 2: Amazon demand and supply estimates

	demand	supply	
	(1)	(2)	(3)
	logit	linear	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 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. 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 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. Robust standard errors are reported in parentheses.

Table 4: Welfare effects of bias

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

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 consumer surplus and revenue attained from the debiased rankings relative to the maximum levels of CS and revenue on the frontier. The bottom panel compares outcomes associated with the model's depiction of actual rankings with those for the debiased rankings. 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 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 from $R^{\text{max Rev}}$ necessarily reduces the numerator of this formula. Hence, revenue can only *rise* with if such a switch reduces the denominator proportionally more than it reduces the numerator.

This, in turn, occurs when

$$\frac{(p_m e^{\delta_m^0} - p_k e^{\delta_k^0})}{\sum_{j\in J} p_j e^{\delta_j^0} e^{\gamma r_j}} > \frac{(e^{\delta_m^0} - e^{\delta_k^0})}{1+\sum_{j\in J} e^{\delta_j^0} e^{\gamma r_j}},$$

or when

$$\frac{(p_m e^{\delta_m^0} - p_k e^{\delta_k^0})}{(e^{\delta_m^0} - e^{\delta_k^0})} > \bar{p}(1-s_0),$$

where

$$\bar{p} = \frac{\sum_{j\in J} p_j e^{\delta_j^0} e^{\gamma r_j}}{\sum_{j\in J} e^{\delta_j^0} e^{\gamma r_j}},$$

and $(1-s_0)$, the inside share, is given by

$$(1-s_0) = \frac{\sum_{j\in J} e^{\delta_j^0} e^{\gamma r_j}}{1+\sum_{j\in J} e^{\delta_j^0} e^{\gamma r_j}}.$$

Intuitively, violations (increases in revenue with rank order permutations relative to $R^{\text{max Rev}}$) arise when the permuted products have very similar revenue (i.e. similar $p_j e^{\delta_j^0}$) but different prices. Raising the rank of a product with only slightly lower expected revenue but with a higher price necessarily places a less appealing (lower δ_j^0) product at a better rank. Because the less appealing product now receives a lower rank-induced weight ($e^{\gamma r}$), the probability of purchase declines more than the numerator.

Our proposed welfare frontier arises from ranking products according to the weighted sum, $\kappa_1 \ln(p_j) + \kappa_2 \delta_j^0$. To this end, we first use the 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 delivers a direct violation such that both CS and revenue lie beyond the pre-permutation proposed frontier.

At the revenue-maximizing endpoint of the frontier, permutations 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 an average of 99.99892% of the maximal revenue that we calculate.

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 general 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.