NBER WORKING PAPER SERIES

PLAYLISTING FAVORITES:
MEASURING PLATFORM BIAS IN THE MUSIC INDUSTRY

Luis Aguiar
Joel Waldfogel
Sarah B. Waldfogel

Working Paper 29017
http://www.nber.org/papers/w29017

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2021

This research had no external support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w29017.ack

Playlisting Favorites: Measuring Platform Bias in the Music Industry
Luis Aguiar, Joel Waldfogel, and Sarah B. Waldfogel
NBER Working Paper No. 29017
July 2021
JEL No. K21,L12,L82

ABSTRACT

Platforms are growing increasingly powerful, raising questions about whether their power might be exercised with bias. While bias is inherently difficult to measure, we identify a context within the music industry that is amenable to bias testing. Our approach requires ex ante platform assessments of commercial promise - such as the rank order in which products are presented - along with information on eventual product success. A platform is biased against a product type if the type attains greater success, conditional on ex ante assessment. Theoretical considerations and voiced industry concerns suggest the possibility of platform biases in favor of major record labels, and industry participants also point to bias against women. Using data on Spotify curators' rank of songs on New Music Friday playlists in 2017, we find that Spotify's New Music Friday rankings favor independent-label music, along with some evidence of bias in favor of music by women. Despite challenges that independent-label artists and women face in the music industry, Spotify's New Music curation appears to favor them.

Luis Aguiar
Department of Business Administration
University of Zurich
Plattenstrasse 14
8032 Zurich
Switzerland
luis.aguiar@business.uzh.ch

Joel Waldfogel
Frederick R. Kappel Chair in Applied Economics
3-177 Carlson School of Management
University of Minnesota
321 19th Avenue South
Minneapolis, MN 55455
and NBER
jwaldfog@umn.edu

Sarah B. Waldfogel
Department of Economics
University of Wisconsin–Madison
swaldfogel@wisc.edu

1 Introduction

In the past few decades, many markets have become dominated by platforms that enjoy increasing returns, giving them power and the possibility of abusing it. Prominent examples include Google, whose search algorithm is alleged to favor properties in which it has a financial interest (Edelman, 2011) and Amazon, which may use its internal data to disadvantage its suppliers (Zhu and Liu, 2018). The music industry, too, is now dominated by a small number of platforms, chiefly Spotify and Apple Music, whose power has invited scrutiny about possible bias.[1]

At the same time, digitization has brought an explosion of new creative production; and large libraries of new and old songs are now directly available to consumers. For example, Spotify has over 40 million songs and adds about 40,000 per week. In this environment, platforms such as Spotify and Apple Music – through the songs they choose to promote – have substantial power to determine which songs and artists succeed. The main mechanism for promotion on these platforms is "playlists," which are both informative lists of songs as well as utilities for listening to these songs. Playlists can have many followers, and placement on influential lists has large impacts on artist and song success (Aguiar and Waldfogel, forthcoming).

Both theoretical considerations, as well as perspectives of industry participants, provide reasons for concern that such power might be exercised with bias (Hagiu and Jullien, 2011; Bourreau and Gaudin, 2018). First, some upstream suppliers – the "major" record labels – have ownership stakes in Spotify, which could give Spotify a reason to provide more advantageous promotion of their products. In parallel, representatives of independent record labels have voiced concern about bias against their products (Dredge, 2016, 2015). Industry participants have also raised concerns about platform roles in the music industry success of women. Although it is not clear a priori why platforms would have a financial incentive to engage in gender bias, there is reason for concern, particularly in the wake of the #metoo movement.[2]

[1] In 2019, both services together accounted for 55 percent of the global streaming subscription market. Spotify alone accounted for 36 percent of the market. See https://musically.com/2019/12/09/report-spotify-has-36-market-share-of-music-streaming-subs/

[2] See Smith et al. (2018), and Flanagan (2018). Similar concerns have been raised in the film industry. See, for example, Ellis-Petersen (2014).

Assessing these claims requires a method for measuring platform bias, which is inherently difficult. One could test for bias simply if one knew which factors legitimately affected the way in which platforms should rank products. Conditional on these factors, one could then ask whether different suppliers received differential treatment from a platform (for example, search Google rankings). In our context, one could ask whether different songs – by label type or artist gender – received different rankings on playlists. For example, if two songs had the same expected commercial promise, but the major-label song received a better ranking on the playlist, then one could infer bias in favor of major label music. One might be tempted to implement this by regressing songs' playlist ranks on characteristics reflecting commercial promise as well as label type, interpreting the label type coefficient as bias. But it is difficult to rule out the possibility that apparent bias reflects correlation with relevant but unobserved factors. So we regard this "conditioning on observables" (COO) approach as inherently unpromising, although we implement a version for comparison in section 5.3.

In the past few decades, researchers have developed alternative, *outcome-based* approaches for measuring bias. Rather than asking whether different groups are treated differently after controlling for available factors, these tests compare the ex post performance of different groups assessed to be of equal promise. This approach has two information requirements. First, one needs a measure of the decision maker's ex ante assessment of promise. Second, one needs measures of ex post performance. To use a traditional offline example, if a hiring firm gives applicants a score, then one can compare, say, the on-the-job performance of hired men and women. If women outperform men who had received the same score at hiring, then the firm's scoring procedure is biased against women. Researchers have applied this approach to measuring bias in policing, bail setting, hiring, and journal publication, among other contexts (Ayres and Waldfogel, 1994; Smart and Waldfogel, 1996; Hellerstein et al., 1999; Knowles et al., 2001; Arnold et al., 2018; Card et al., 2020).

The way that Spotify operates some of its playlists satisfies the information requirements for outcome-based bias testing, which surmounts the challenge inherent in the COO approach. First, we can observe curators' ex ante assessments of songs' commercial promise: Spotify maintains human-curated "New Music Friday" lists in each of the countries in which it operates.[3] Each week, Spotify curators rank songs on each country's list; and songs ranked

[3]We use the terms "New Music" and "New Music Friday" interchangeably in the text to refer to these lists.

higher on these lists ultimately stream more based, in part, on curators' ex ante assessments of songs' commercial promise (Aguiar and Waldfogel, forthcoming). Second, we can also observe ex post streaming: Spotify discloses daily streams for the top 200 songs by country. These two ingredients allow bias tests that compare eventual success, relative to ex ante curator assessments. Our contribution is the development of an implementable approach to measuring platform bias, illustrated with measurement of possible platform bias in Spotify's New Music playlists.

While our analysis focuses on a subset of the many playlists available at Spotify, we note that the New Music Friday lists are of particular relevance. They have arguably become "an important part of launching any new song" and the exposure they offer is often seen as "career-changing," also by improving the likelihood of appearing on other important Spotify playlists (Aswad, 2020). Because the New Music Friday lists serve as an important entry point for new products and artists, their potential biases may affect not only which existing products succeed, but also the type of products entering the market in the future.

Using data on 5,736 songs assigned to 14,747 top 20 entries on weekly New Music lists in 26 countries during 2017, we find the following. First, songs with better New Music ranks have more ex post streaming success on the platform, which provides evidence – to a first approximation – that curators rank songs to maximize eventual streams of songs on the playlists. Second, conditional on New Music ranks, songs on independent record labels stream less than songs on major record labels. We find a similar result by gender, although the gender differential in ex post streaming arises only for songs that curators rank outside the top 10. Hence, despite concerns about bias against independent-labels and women, these results indicate that independent music, and music by women, receive better ranks than their eventual on-platform streaming performance seems to warrant.[4] Independent music receives an average boost of about 2 ranks, while music by women receives an average boost of 1.4 ranks. Third, the degree of bias implied by the outcome-based tests differs systematically from the amount implied by a conditioning-on-observables regression of New Music ranks on song types and available predictors of commercial promise, reinforcing our concern about the adequacy of conditioning on observables. Fourth, our bias results do not appear to be driven by differential promotion by label type or artist gender. Finally, we estimate the relative

[4]We refer to "songs by women" rather than "songs by female artists" because we test whether an artist's perceived gender, not their biological sex, affects how the platform treats them.

weights that curators apply to independent-label music and music by women, relative to their respective complements: Curators behave as if they maximized weighted streams, where the weights are 40 percent higher for independent-label music, and 10 percent higher for music by women. We conclude that Spotify New Music playlist rankings use the platform's power to promote music from independent record labels and, to a lesser extent, women; and we suggest that – with access to data – our approach is applicable to other platform contexts.

The paper proceeds in five sections after the introduction. Section 2 provides background on playlists and possible reasons for concern about bias. We also place this work in context of the relevant existing literatures. Section 3 presents a simple model of the curator's ranking decisions in which curators maximize the weighted sums of streams for different types of music, and unequal weights reflect bias. Section 4 describes the data, and Section 5 presents results, including both direct evidence on the existence of bias and measures of the welfare weights curators attach to independent vs major music and music by women vs men. Section 6 concludes.

2 Background

2.1 Promotion, Playlists, and Digitization

Digitization has given rise to a large number of new music products as well as large gains in welfare from products that would otherwise not have been made available to consumers.[5] Increasingly, however, consumer discovery of new products depends on a few platform operators. Prior to digitization, success depended on decisions of which songs to promote on radio, made by program directors at thousands of independently owned radio stations and at thousands of independently-owned record stores. Promotion on radio, and availability in record stores, have long been understood to be valuable. As a result, record labels had a long history of compensating radio stations for airing their songs, sometimes illegally in the form of "payola" (Nayman, 2012).

With the rise of Spotify and Apple Music, both promotion and distribution now occur through streaming platforms, which simultaneously play the roles previously played by both

[5]The discussion of playlists in this section draws on Aguiar and Waldfogel (forthcoming).

radio stations and record stores. While platforms passively accept essentially all new music, platforms more consequentiallly decide which songs to feature on playlists. Playlists are lists of songs displayed to users, and they promote music in two ways. First, playlists are potentially informative lists of songs. Second, they are utilities for listening to music. A user can click on a playlist button and then be served the songs on the playlist, in either random or ranked order. Spotify users can choose to "follow" playlists, which causes the lists to appear on users' home screens. At Spotify, any user is free to create playlists, which any other user can follow. Despite free entry into playlist creation, the widely followed lists are controlled by Spotify. Of the top 1,000 playlists on Spotify according to the number of followers, 866 of these are controlled by Spotify, and their cumulative followers account for 90 percent of the followers of these 1,000 lists.[6]

Playlists take two broad forms, curated and algorithmic. The Global Top 50 simply algorithmically includes the top 50 streaming songs from the previous day. While the Global Top 50 is the same for all users, other algorithmic playlists such as "Discover Weekly" are customized to individual users. The New Music playlists that we study in this paper, by contrast, are curated by individuals who decide which songs to include on the lists; and they are controlled by Spotify employees, and each New Music list is the same for all of its users. These lists focus on newly released music that curators find notable and wish to bring to users' attention.

With a total number of 9 million followers across the 26 countries included in our analysis, the New Music Friday lists rank 7^{th} in the most followed Spotify lists on the platform. During 2017, curators added 50 new songs to each country's New Music list per week, and songs remained on these lists for seven days. As Aguiar and Waldfogel (forthcoming) document, playlist inclusion decisions have large impacts on streaming. The higher that curators rank songs on the New Music Friday playlists, the more likely the songs eventually appear on the top 200 streaming songs, and about half of this effect appears causal; the rest reflects curators' predictive ability. As we discuss further in Section 5, curators give more favorable ranks to songs that they correctly predict will become successful. While the effect is substantial for top-ranked songs, it is negligible for songs that curators rank outside the top 20.

[6]These calculations are based on the top 1,000 playlists listed at Chartmetric's playlist ranking (https://app.chartmetric.com/playlist-list/spotify) on November 7, 2020.

Policy makers and industry observers have demonstrated growing concerns about platform behavior, including concerns about biased exercise of their power. Perhaps the most prominent example of this arose in 2017 when the European Commission fined Google $2.7 billion for delivering biased search results which favored its own properties (Scott, 2017). Financial relationships between platforms and products among those they promote arises not only at Google but also at Spotify. The major record labels own substantial shares of Spotify: In 2008 they collectively owned 18 percent of Spotify; by 2018 this had fallen to about 6 percent.[7]

Music industry participants have concerns about the possibility of financial influence on promotion decisions, even apart from incentive issues arising from ownership stakes. According to Dredge (2015), independent record labels "worry that major labels will start putting more pressure on Spotify to include more of their songs, while others note that some independent creators of popular playlists are hoping to charge money to add tracks." A few years later, the concern about payola was reinforced when Spotify CEO Daniel Ek indicated, in a 2019 earnings call, that Spotify had found "a way to charge record companies to pay to promote their artists direct to their fans on our platform without p**sing off our users too much"; and in October of 2019 Spotify announced that labels could "pay to have their artists promoted to targeted fans within the Spotify ecosystem via a 'Brand New Music For You' pop-up visual ad" (Ingham, 2019). In 2020 Spotify announced a 'Discovery Mode' for artists in which their songs would be featured more in personalized listening sessions in exchange for a lower royalty (Stassen, 2020). While neither of these Spotify innovations directly affects placement on the New Music lists studied in this paper, they may fuel industry concerns about bias.

Other observers raise concerns about possible gender bias. For example, Smith et al. (2018) reports that of "600 popular songs on the Billboard Hot 100 year end charts from 2012 to 2017," it is "suspicious" that "women comprised just 22.4% of artists and 12.3% of songwriters." A widely read journalistic account (Pelly, 2018) reported, based on monitoring some popular Spotify playlists (including Today's Top Hits, New Music Friday, Rock This, RapCaviar, Hot Country, and ¡Viva Latino!), that many were "staggeringly male-dominated." Concerns about gender issues in recorded music are also reflected by events surrounding remarks of Recording Academy President Neil Portnow after the 2018 Grammy awards. Asked

[7]See Ingham (2018).

about gender imbalances in the awards, he indicated that women needed to "step up." This attracted criticism, and he resigned in 2019 (Flanagan, 2018).

2.2 Existing Evidence

This paper contributes to two existing literatures. First, there is a growing body of work on possible platform bias. Hagiu and Jullien (2011), Bourreau and Gaudin (2018), and Belleflamme and Peitz (2018) provide theoretical rationales for platform bias. On the empirical side, Edelman (2011) explores whether Google biases its search results in favor of its own properties, and Zhu and Liu (2018) study whether Amazon enters the markets for products established by its marketplace vendors. Hunold et al. (2020) show that travel platforms rank hotels less favorably if the hotel charges less off the platform. A related literature is concerned with gender bias in particular: Lambrecht and Tucker (2019) investigate the role of algorithms in possible gender bias of ad targeting. Datta et al. (2015) use field experiments to document that user behavior and characteristics such as gender affect the ads the user is shown. We extend this literature with direct test for platform bias.

Second, playlists are a form of product recommendation made more important by the large increase in the number of new products brought forth by digitization. Hence, this work is also related to both work on the long tail of new creative products (Anderson, 2006; Brynjolfsson et al., 2003; Aguiar and Waldfogel, 2018), as well as work on product recommendations, reviews, and product discovery (Sorensen, 2007; Reinstein and Snyder, 2005; Forman et al., 2008; Duan et al., 2008; Datta et al., 2018).

3 A Model of Playlist Rankings

We now introduce a model of curator playlist rank determination in order to guide our empirical work by providing both testable predictions and an explicit way to define and measure bias.

Consider a playlist curator who decides which songs to add to the lists, and in which order. We assume that songs differ in their *likeability*, which we define as the tendency for people, once exposed, to desire further engagement with the song. Playlists provide exposure, and positions higher on the list provide more exposure. Define l_j as the likeability of song j, and suppose that the amount of exposure a song receives depends on its playlist rank r according to Ae^{Br}, where $A > 0$ and $B < 0$. Then the expected streams eventually received by song j when it is placed at rank r is given by $s(j,r) = l_j Ae^{Br_j}$. Total expected streams for the ranked songs are given by $S = \Sigma s(j,r) = \Sigma l_j Ae^{Br_j}$, where the summation occurs over the playlisted songs. A curator maximizing listening to playlisted songs would rank chosen songs according to their likeability.

While we postulate playlist stream maximization as the curator's rationale for ranking songs already chosen for the list, a curator can choose songs for list inclusion with other rationales in mind. The curator might view the 50-song list as a bundle whose appeal depends on more than the songs' individual attributes. Thus, for example, a curator might include only one track from popular artist's newly released album, even if its second or third most appealing tracks were more "likeable" than other songs chosen for the playlist. We can distinguish between rationales for including songs on the list and the rationale for the ranking. Even if a curator does not include all of the most likeable songs on the list, the maximization of streams of the songs on the list entails ranking songs from most to least likeable. A curator seeking to maximize listening would rank listed songs according to their likeability. That is, if $l_1 > l_2$, then $l_1 Ae^B + l_2 Ae^{2B} > l_1 Ae^{2B} + l_2 Ae^B$. In words, expected streams are maximized by granting more exposure to more likeable songs.

We can augment the above framework to accommodate possible bias by allowing the curator to attach different weights to different types of music.[8] We define bias as a positioning that deviates from stream maximization.[9] For example, if the curator attaches different weights to major- and independent-label music, then the curator's maximand becomes $W = \sum_{j \in M} l_j Ae^{Br_j} + \sum_{j \in I} \psi l_j Ae^{Br_j}$, which can be written more succinctly as:

$$W = \sum_{j \in M \cup I} [(1 - \delta^I) + \delta^I \psi] l_j Ae^{Br_j}, \tag{1}$$

where δ^I is an indicator for independent-label songs, M is the set of major-label songs, I is

[8] Uncovering welfare weights underlying decision maker behavior has a long history. See Ahmad and Stern (1984) and Ross (1984).

[9] Our notion of platform bias recalls the definition of media bias in Gentzkow and Shapiro (2010), which measures bias as the deviation in product positioning from the location on the political spectrum that would maximize profits.

the set of independent-label songs, and ψ is the weight the curator attaches to streams of independent-label songs, relative to streams of major-label songs. Maximization of expected weighted streams, W, by the ranking of songs requires the curator to order songs by $[(1 - \delta^I) + \delta^I \psi] l_j$ rather than by simply l_j. Note that $[(1 - \delta^I) + \delta^I \psi]$ equals 1 for major-label songs and ψ for independent-label songs.

We have thus far described curators as maximizing weighted streams, but a platform facing differing streaming costs across types of songs might instead maximize the net benefit of streams, i.e. the benefit of generating streams less their costs. In that case ψ would have a different interpretation. Rather than simply reflecting the differential benefit that curators attach to, say, indie-label streams, the parameter ψ would instead reflect differential benefits less costs. Hence, a value of $\psi > 1$ would reflect some combination of curators attaching greater benefits to indie streams and the platform facing lower costs on those stream. While different, this interpretation also reflects a form of bias. Given Spotify's prohibitions on payment for playlist placement, we understand playlist ranking to reflect curators' editorial predictions of song likeability. Hence, we interpret differential playlist rankings based on costs as a form of bias as well.

This setup gives us a way to interpret the relationships among the New Music ranks chosen by the curator, eventual streams, and song type. First, monotonic decline in realized streams as songs are ranked worse is consistent with stream maximization by curators. Second, if realized streams vary by song type, conditional on curator rank – for example between major label and independent music – then we have evidence of bias. Moreover, we can identify the degree of bias (ψ).

Suppose we see an independent-label song at rank r streaming less than a major-label song that is ranked one place worse, at $r + 1$. We can rationalize this ranking as one that maximizes weighted streams with a weight ψ that the curator attaches to independent-label music. Figure 1 illustrates this with a stylized pattern of average streams by rank and label status. In the Figure, average streams for independent-label songs ranked 2 ($s(I, 2)$) fall short of average streams for major-label songs at rank 3 ($s(M, 3)$). We can rationalize this pattern of average streams by rank and song type as one that maximizes weighted streams if curators value independent streams more highly than major-label streams, or that $\psi s(I, 2) > s(M, 3)$. More generally, we obtain a lower-bound on ψ from: $\psi > \frac{s(M,r+1)}{s(I,r)}$.

We could satisfy this inequality with an arbitrarily large value of ψ. To pin down a finite estimate of ψ, we also need an upper bound. A second condition on the curator's weight, also illustrated in Figure 1, is that the weight on independent streams (ψ) cannot be so large that weighted streams of independent songs ranked 2 exceed streams of major-label songs ranked 1. That is, the upper bound on ψ is defined by $\psi s(I,2) < s(M,1)$ or, more generally, that $\psi < \frac{s(M,r)}{s(I,r+1)}$.

In short, we have upper and lower bounds on the weight the curator attaches to independent-label music based on the expected streaming success of major and independent music at adjacent ranks:

$$\frac{s(M,r+1)}{s(I,r)} < \psi < \frac{s(M,r)}{s(I,r+1)}. \tag{2}$$

We obtain similar conditions when considering the possibility of curators attaching different weights to songs by women and men. We can econometrically estimate ψ by obtaining the expected streams by rank and song type, then choosing a value of ψ to minimize the magnitude of the inequality violations, and we implement this in Section 5.5.

Although this method for measuring bias is built around features of our music-industry context, it is worth noting that the basic approach could be applied in other platform areas. For example, Google's organic search results are supposed to be ranked according to the quality of the link, reflecting users' tendencies to click through and to stay at the linked page. Conditional on their order in the search results, if links to Google's own pages had fewer, or less successful clicks, that could be interpreted as pro-Google bias.[10]

3.1 Testing for Bias

One way to test for bias is to estimate the parameter ψ in the foregoing model and check whether it deviates from unity. We will do this, but it is also useful to pursue a direct measurement of bias via our outcome-based approach.

As discussed in the introduction, one could try to measure bias, using the conditioning-on-observables (COO) approach, by regressing ranks on factors predictive of curator likeability

[10]See Varian (2009) for a discussion of search auctions.

assessments along with, say, a label type indicator. Although there are some available observables – such as artist past streaming success – that are predictive of song success, it is difficult to rule out the possibility that unobservable song characteristics predictive of success are correlated with song characteristics of interest such as label type or artist gender.

Our preferred outcome-based approach avoids these concerns. While neither likeability (l_j) nor the degree of exposure associated with each rank (Ae^{Br_j}) is observable, realized average streams – the empirical analog of expected streams – for each rank are observable. That is, expected streams $s(j, r)$ for each rank can be measured based on realized averages. This generates both a way to see whether curators are maximizing streams and a way to measure bias. If curators are maximizing expected streams, then we should see curators rank-ordering songs by likeability. We do not observe likeability, but we do observe its consequence, realized streams; and, as we document below, we see higher realized streams for songs ranked better.

Interpreting this as evidence that curators rank order by likeability is complicated by the possibility that playlist position may reflect not only curators' assessments of likeability but also a causal influence of playlist ranks on eventual streams. In the extreme, the entire relationship between ranks and eventual streaming could be the causal influence of the exposure associated with the rank, and lower streams for songs with worse ranks might not provide evidence about curator ordering based on likeability. If all songs were equally likeable, then realized streams would still decline as rank worsened simply because of exposure. Aguiar and Waldfogel (forthcoming) document that songs better ranked on the New Music playlists experience more streaming success but – importantly – that only about half of the raw relationship between rank and streaming success is causal. This provides evidence that curators allocate the more commercially promising (likeable) songs to better playlist positions and, in turn, indicates that curator maximize listening, to a first approximation. Beyond that, differential realized streams, conditional on curators' ex ante assessments – the New Music ranks – are evidence of bias.[11]

The outcome-based test avoids the need to find sufficient measures of likeability to circumvent the problem of potentially correlated unobservables. Implementing the OB test requires information on the curator's ex ante likeability assessments. While it is generally challenging

[11]In addition to systematic evidence of the rank-streams relationship, we can also point to curator claims that they rank songs according to predicted promise. According to Spotify's Head of Global Hits, "We highlight the songs that will probably get the most streams at the top..." See Aswad (2020).

to know how commercially promising the curator assesses teach song to be, our context provides us with such a measure through the rank that is assigned to each song included on the New Music lists.

All of Spotify's playlists have ranks, but the New Music lists have the unique feature that songs are given a rank when placed on the list; and this rank remains constant for the seven days that songs remain on the lists. Other lists operate differently: A song enters a list at some rank, and the rank can change repeatedly while the song remains on the list, often for months. Hence, the New Music lists are well suited for implementing the outcome-based approach. The associated limitation is that our results are specific to the curation decisions for these lists and may not generalize to Spotify's playlist curation as a whole.

One important caveat is in order. Our approach tests for curatorial bias in choosing rankings, *taking consumer tastes as given.* If listeners are biased against some type of music, then catering to this taste by giving worse ranks to the sort of music that listeners dislike – even if unfairly – will not register as bias in our test. Our approach does not test whether there is bias in the environment. Rather, our test indicates whether curators' ranking decisions favor or disfavor particular kinds of music, given listener preferences. We return to this theme in the conclusion.

4 Data

The data for this study consist of the 14,747 top-20 entries on Spotify's New Music Friday playlists, for 26 countries during the period from early April until the end of 2017.[12] The 14,747 entries include 5,736 distinct songs. For each entry, we observe its New Music rank in each country where it is included and the week when it was included. We also observe four groups of variables related to the song. First, we observe measures of its streaming success, in particular its daily Spotify streams, by country, during the 52 weeks following its appearance on a New Music playlist, if it appeared in a country's daily top 200. Second, as detailed below, we observe whether a song is on a major or an independent label, as well as artist gender. Third, we observe some additional song characteristics, including artists' past (2016) streams, as well as sonic characteristics (key signature, etc). Finally, we observe

[12]We include only the top 20 New Music entries per country and week because Aguiar and Waldfogel (forthcoming) document that New Music inclusion effects are negligible for songs ranked outside the top 20.

measures of song promotion at Spotify, in particular the other playlists on which a song appears around the time it appears on a New Music list, as well as the numbers of followers for each of those playlists. We are able to obtain these playlist variables for 4,067 songs accounting for 12,366 playlist entries.

We also calculate artists' (previous-year) 2016 streams on Spotify, by country. Finally, we observe numerous song characteristics, including beats per minute, key signature, whether major or minor, and seven other characteristics of the songs measured on 100 point scales (danceability, valence, energy, acousticness, instrumentalness, liveness, and speechiness).

Our data are drawn from three sources. The data on the New Music playlist entries, and the data on song characteristics, are drawn from Spotontrack (https://www.spotontrack.com/). The data on Spotify streams are from https://kworb.net/spotify/. Data on the songs' playlist histories are from Chartmetric (https://www.chartmetric.com/).

We do additional work to create two variables of special interest for the study, whether a song is released by a major record label and the gender(s) of a song's artist(s). We measure whether a song is on a major record label according to whether the song is on a label we have classified as major, according to information at Wikipedia label pages.[13]

We used a three-pronged approach to determine artist gender. First, we searched the artist name on Musicbrainz, which provides genders for many individual artists. Second, we checked the artist names against first name databases.[14] If we lacked a Musicbrainz match and the first name was attached to a gender with at least 90 percent probability, we assigned that gender. We found the remaining artist genders by hand.[15] We discard songs when the genders of either of the first two listed artists is unknown. Assigning artist gender is complicated by the fact that may artists are not individual but are instead groups of people. In our main analysis we code a song as being by a woman if its sole artist is a woman, or in the case of multiple artists, the first two artists or bands are either women or

[13]We classify a label as major if its name includes any of the following names: Asylum, Atlantic, Capitol, Epic, Interscope, Warner, Motown, Virgin, Parlophone, Republic, Big Machine, Sony, Polydor, Big Beat, Def Jam, MCA, Universal, Astralwerks, WM, Tobago, RCA, Ultra Records, LLC, disco:wax, Island Records, Spinnin' Records, Black Butter, Fueled By Ramen, WEA Latina, UMLE, Hollywood Records, Ariola, EMI, b1, FFRR, Elektra France, WEA, Reprise, Arista, and Columbia.

[14]We use the U.S. Social Security names database as well as the website https://gender-api.com, which allows to determine gender based on a first name and covers non-US names.

[15]To do this, we visited artists' Wikipedia pages, social media pages, and interviews and looked for artists' pronouns. While the best measure of gender requires asking artists how they identify, our perception of the artist's gender - what we can observe - is likely to be the same information that is available to Spotify.

mixed ("women or mixed").[16]

Table 1 provides sample statistics. Of the New Music entries in the data, 66 percent are on major labels, and 21 percent are by women. Just over a third appear in the daily top 200 streaming data. Even assigning zero streams to songs not making the top 200, the average streams per song is close to 1 million, and the average first-week streams measure is 70,000. The streams distribution is highly skewed: the median is zero, and the 75^{th} percentile measure 0.026 million streams. For the songs with the additional playlist data, the average number of playlist followers that a sample song has at the time of its appearance on the New Music list is 3.37 million. Of this total, 88 percent are for lists curated by Spotify, while 7 percent are on major-label-controlled playlists.

5 Empirical Results

This section presents empirical results, in five parts. First, we present evidence that curators rank songs to maximize streams. Second, we provide the results of the outcome-based tests for bias. Third, we compare the degree of rank bias implied by the OB test with the rank bias implied by the inherently challenged COO approach. Fourth, we explore robustness of the OB bias results to the time horizons for streams and to the inclusion of controls for other playlist inclusion. Finally, we present estimated weights that curators attach to independent-label music, and songs by women, relative to other music using the inequalities approach in Section 3.

5.1 Curator Goals

The outcome-based test asks whether two songs that curators give the same initial assessment ultimately experience similar degrees of success. Implementing the outcome-based test requires both measures of ex post song success as well as measures of curators' assessments of song likeability at the time of release. Streams provide a clear measures of song success,

[16] In the appendix we explore the sensitivity of the gender results to three additional approaches. Our most restrictive measure treats a song as a woman-only song only if its sole artist, or both entities (artists or groups), are all women. We refer to this as the "women-only" measure. Less restrictive than our main approach is a measure that deems the song to be by women if the first artist (or band) is a woman regardless of the additional participants ("first artist woman"). Our most liberal measure classifies the song as by women if either the first or second entity is a woman ("either artist woman"). See Figure B.1 and Table B.1.

and the songs' New Music ranks are curators' ex ante assessments. This runs the risk of circularity. Curators' New Music ranks are, by assumption, reflections of their likeability assessments. Relying on this approach requires evidence that curators assign New Music ranks with stream maximization in mind. Fortunately, the relationship between curators' New Music ranks and songs' subsequent success provides strong evidence that curators' order songs to maximize streams. As depicted in Figure 2, songs that curators rank 1^{st} have over an 80 percent probability of appearing among the top 200 streamers, while songs ranked 5^{th} have roughly a 50 percent probability; and those ranked 10^{th} have about a 25 percent probability. The log of eventual streams bears a similar monotonic relationship with curator ranks.

Of course, the positive relationship between curators' ranks and eventual streaming success reflects two distinct mechanisms. First, as Aguiar and Waldfogel (forthcoming) document, there is a causal impact of New Music ranks on eventual streams, which presumably arises from exposure. Even if songs did not differ in likeability (l_j), their eventual streaming volumes would decline for worse ranks due simply to exposure differentials. But this causal effect, as shown by the authors, explains only half of the magnitude of the difference between the streaming success of the top-ranked and lower-ranked songs. The second mechanism is simply prediction: Curators choose songs that they predict will be appealing to consumers, so part of the relationship between curatorial ranks and streaming success arises from the curators' higher ranking of more likeable songs. The fact that the causal impact of New Music ranks delivers only half of the relationship between ranks and eventual streams means that ranks also reflect curator prediction and, in particular, that curators order songs according to their commercial prospects, or likeability.

We can also test whether curators rank songs according to likeability using data on song characteristics predictive of commercial success. One such variable is the past (2016) streams of songs by the artist whose song is currently on the New Music list. As Figure 3 shows, past streams – both the share of songs whose artists had past streams in the top 200 and average past log streams – decline essentially monotonically in current New Music ranks. That is, curators give better ranks to songs whose artists have had more prior success, on average.

We take the evidence in Figures 2 and 3 to indicate that curators rank songs on the New Music lists as if they were – broadly – trying to maximize eventual streams. We say "broadly" because specific kinds of deviations from this pattern constitute bias.

5.2 The Outcome-Based Test for Bias

Our outcome-based bias test is literally a test for bias in curators' assignment of ranks: Conditional on the rank assigned by a curator, does a song of one type stream more than a song of another type? If so, then the song was ranked in a biased fashion. For example, does a song by a woman need, on average, better prospects - measured, for us, by eventual streaming success - to be granted the same New Music rank as a song by a man? We can test this simply by comparing measures of streaming success for men and women, or major and independent-label, songs that curators assign the same New Music rank.

Figures 4 and 5 compare streaming success of songs by label type and artist gender, respectively, using two measures of streaming success: the share of songs appearing in the daily top 200 in the country in which they appear on the New Music list, and observed log streams (all streams for the days when the songs appear in the top 200). As Figures 4 and 5 show, conditional on New Music rank, independent-label songs – and songs by women – have less streaming success. Moreover, the differential streaming averages prevail throughout the top 20 for label type, while they are more pronounced outside of the top 10 playlisted songs for gender. These figures provide our first glimpse of the paper's main results. Conditional on how likeable curators assess a song to be, independent-label music, and music by women, becomes less successful. This is consistent with curator biases in favor of independent-label music and, to a more limited extent, music by women.

We perform the basic outcome-based tests via regressions of streaming success on New Music rank variables and song type indicators. In particular, we estimate the following equation:

$$s_{jc} = \beta Rank_{ic} + \alpha^{indie}\delta_j^{indie} + \alpha^{women}\delta_j^{women} + \mu_c + \varepsilon_{jc}, \tag{3}$$

where the dependent variable s_{jc} is a measure of streaming success for song j in country c. We consider two measures of streaming success. The first one is a binary measure of whether song j appears among the daily top 200 streaming songs in country c at some point after entering the New Music playlist. As a second measure of streaming success, we use the logarithm of the cumulative streams obtained by song j in country c after entering the New Music list. δ_j^{indie} and δ_j^{women} are indicator variables equal to 1 if song j is on an independent label or by a woman, respectively. $Rank_{jc}$ is the rank obtained by song j on the New Music Friday list of country c, μ_c are country fixed effects, and ε_{jc} is an error term. We alternatively use rank fixed effects instead of including the New Music rank linearly in some of our specifications.

Table 2 provides the results of estimating equation (3). The first two columns include indicators for New Music ranks as well as country indicators. The latter two columns repeat the exercise using the New Music rank linearly, rather than flexibly with rank indicators. Both songs by women and songs on independent record labels receive fewer streams – by both measures – than their men and major-label counterparts. The effect is larger for label type than gender. These results, along with Figures 4 and 5, indicate bias in favor of independent labels and songs by women.[17]

The linear specifications in the last two columns allow us to measure the average magnitude of the bias, in terms of rank, via the ratio of the indicator coefficient (α) and the linear New Music rank coefficient (β). For example – based on the last column – independent label songs have 0.962 fewer log streams, conditional on their ranks. Being ranked one worse is associated with 0.488 fewer log streams. Hence, the differential treatment of independent label songs is equivalent to favoritism in rank assignment of 1.97 ranks (0.962/0.488). The analogous rank differential for gender is 1.40 (0.682/0.488).

Table 2 and Figures 4 and 5 aggregate countries into single coefficients. According to Spotify, the New Music playlists are "curated by our playlist editors - genre, lifestyle, and culture specialists with diverse backgrounds from all around the world." Moreover, Spotify reports having curation teams "in every country."[18] It is possible that curation for the respective countries' New Music lists exhibits different degrees of bias. To explore this we estimate country-specific coefficients on gender and label status, using our binary measure of streaming success as the dependent variable (appearing in the top 200). We find strong evidence against coefficient constancy for label status (p-value < 0.000001) and weaker evi-

[17] In unreported regressions that allow gender and label type effects to differ for songs ranked 11-20, gender effects are insignificant in the top 10 using the share of songs streaming in the top 200.

[18] See https://artists.spotify.com/help/article/new-music-Friday and https://www.forbes.com/sites/dannyross1/2020/03/02/spotifys-head-of-music-explains-playlisting/.

dence against constancy for gender (p-value = 0.014).[19] The left panel of Figure 6 shows that the pro-indie bias in the aggregate coefficient is driven most strongly by Norway, Finland, Great Britain, the United States, and Spain. The right panel of Figure 6 shows that the degree of gender bias is more similar across countries. Here, Chile, Germany, the United States, Canada, and Great Britain have the largest bias in favor of women. While we reject coefficient constancy for label type in both tests and in one of two tests by gender, the gender coefficients are more similar across country than the label type coefficients. In that sense, the treatment of gender is closer to a platform-wide strategy.

5.3 Comparing Outcome-Based and Conditioning on Observable Tests

With the shortcomings of the conditioning on observables approach in mind, we can run a regression of New Music rank on factors plausibly associated with the ranks that curators would assign, including measures of past (2016) streaming by the artist, as well as song characteristics. Then we can compare measures of rank bias from the COO approach with measures of rank bias implied by our preferred estimates from outcome-based tests (in Table 2). In particular, we estimate the following equation:

$$Rank_{jc} = X_j\omega + \gamma^{indie}\delta_j^{indie} + \gamma^{women}\delta_j^{women} + \mu_c + \varepsilon_{jc}, \tag{4}$$

where X_j is a vector of song characteristics, including country-specific past log streams (and an indicator for zero past streams), song characteristics, including whether the song is of US-origin, the number of beats per minute, seven characteristics of the songs measured on 100 point scales (described above), whether the song is in a major or a minor key, and the song's key signature. We report the results of estimating equation (4) in the first column of Table 3; and it yields a coefficient of -0.758 for songs by women (indicating pro-woman bias) and a coefficient of 1.074 for independent label music (indicating an anti-independent label bias). The remaining columns present estimates of rank bias implict in the outcome-based test calculated from Table 2. The lower panel presents tests of equality between the rank

[19] Using the log streams measure, we reject constancy for label type (p-value < 0.000001) but not for gender (p-value=0.108).

bias estimated in the first specification and the rank biases in the last two columns. That is, we test whether $\gamma = \frac{\alpha}{\beta}$ separately for label type and gender. We reject equality of the bias for label type using both specifications and in one of two gender tests. We take this fairly systematic rejection as empirical evidence, in addition to the a priori concerns, of the shortcomings with the conditioning on observables approach.

5.4 Outcome-Based Test: Robustness

While the outcome-based approach obviates the concern about unobserved likeability determinants correlated with label type or gender, the OB approach has a vulnerability of its own. The OB test reveals, for example, that independent label music ultimately streams less, conditional on New Music rank. We interpret this as bias, but it is also possible that, conditional on New Music rank, independent label music receives less promotion than major label music, which could explain lower levels of subsequent streams. We have two ways to address this concern.

First, we can look at initial streams – while the songs are on the New Music Friday list – rather than eventual streams during the year following appearance on a New Music list. If playlist placement is responsive to streaming success, then we would expect a feedback loop between streaming, conditional on current playlist placement, and subsequent playlist placement. Eventual streams would therefore reflect not only initial New Music rank and listener response to the songs but also intervening promotion decisions to place the songs on more or less followed playlists, according to ongoing consumer response. This logic suggests that an outcome based test using eventual streaming, conditional on initial conditions, runs the risk of conflating later promotion with consumer response. A test using the streaming outcomes from the first week of post-New Music list streaming should be uncontaminated by feedbacks between streaming success and resulting - endogenous - promotion. Table 4 reproduces the analysis in Table 2, using measures of initial streams rather than eventual streaming success. Results are nearly identical, suggesting that our estimates may be interpreted as the result of bias rather than differential promotion.

Second, we can use the playlist follower data to directly control for additional promotional effects on the Spotify platform.[20] At the time a song is added to a New Music list, the

[20]While our playlists follower measures describe promotion on the Spotify platform, two of the measures –

song can also be added to other Spotify playlists specializing in new music. We observe all of the playlists to which the song is added as of two days after its inclusion on the New Music list, and we tabulate the numbers of followers on these lists. We do this separately for Spotify playlists controlled by Spotify itself, by the major record labels, and by other entities collectively. Because these playlist data are only available for a subsample, the first two columns of Table 5 repeat the basic outcome-based test regressions from the first two columns of Table 2. The third and fourth columns add the initial playlist follower measures. While the initial followers measures are strongly significant, the bias coefficients remain large and significant.[21] The last four columns of Table 5 reproduce the analysis using initial rather than eventual streams, with very similar results.

The evidence presented above indicates that Spotify's New Music ranks are biased; in particular, they favor music from independent labels and, to a lesser extent, music by women.

5.5 Curator Welfare Weights

The evidence in the foregoing section – that eventual success declines monotonically with the songs' ranks – shows that curators rank songs, broadly, to maximize streams. But we also see systematic deviations from stream maximization, in that songs from independent labels and by women receive better ranks than their eventual streaming success warrants. We can rationalize this behavior inside the stream maximizing framework by allowing curators to have different weights by label type or gender; and we can uncover these weights.

Starting with the inequalities from Section 3 – $\frac{s(M,r+1)}{s(I,r)} < \psi < \frac{s(M,r)}{s(I,r+1)}$ – the main implementation question is how to estimate the expected streams conditional on song type and rank, e.g. $s(j \in M, r)$ or $s(j \in I, r)$. We implement this by averaging across all of the country weeks in our data, to calculate $s(j \in M, r)$ and $s(j \in I, r)$ for $r = 1, \ldots, 20$, as well as the analogous measures for by gender. We then find the ψ that minimizes the violations of these two inequalities:

$\psi s(I, r) > s(M, r + 1)$ and $\psi s(I, r + 1) < s(M, r)$

those for major labels and others – reflect promotional efforts by entities that operate outside of the platform. Hence, including these controls may also account for off-platform promotional activities. If promotion outside of the Spotify platform affects streams in ways not directly correlated with our in-platform promotion measures, then it is possible that our outcome-based test results will be biased.

[21] We obtain similar results if we condition on playlist followers as of the end of the week the songs are on the New Music lists.

We implement this for both label type and gender, and we do so using four measures of streaming success employed above (appearance in the top 200 and log streams, for eventual and initial streaming). Table 6 reports the resulting welfare weight estimates, along with standard errors produced by 100 bootstrap replications in which we resample on country weeks in producing the $s(., r)$ measures.

Depending on the measure used, the weight on independent streams, relative to major-label streams, is between 1.40 (se = 0.05) and 1.45 (0.05). That is, curators appear to rank songs as if they attached 40-45 percent more weight to independent-label than to major-label songs. The weights on music by women, relative to men, range from 1.08 (0.03) to 1.17 (0.04), indicating that curators rank songs as if they attached about 10-15 percent more weight to streams for songs by women.

6 Conclusion

Large sectors of the economy have come to be dominated by platforms whose decisions can affect the fortunes of their suppliers and consumers. We examine the exercise of this sort of power by Spotify, the largest of the platforms serving the recorded music industry. We propose and implement an outcome-based test for bias that avoids the challenges of traditional "conditioning on observables" approaches to bias measurement but requires information on platform assessments of product quality along with measures of eventual product success. In our context, playlist ranks are the ex ante measure; and eventual streaming success is the ex post success measure.

We find that Spotify's New Music curators' ranking decisions largely reflect stream maximization but deviate from this basic goal by ranking independent music and songs by women more highly than the work's eventual streaming success warrants. While we do find evidence of bias, the apparent bias that we document is at odds with the allegations of industry critics. Despite substantial major-label ownership stakes, Spotify does not appear to bias its New Music rankings toward the majors. And despite industry criticism about label type and gender bias, Spotify's New Music curators appear to favor these groups.

While our study covers only the New Music Friday lists, we highlight that these can also affect the likelihood of appearing on other important Spotify playlists (Aswad, 2020) and play an important role in the discovery and success of new products (Aguiar and Waldfogel, forthcoming). Their biases may affect not only which existing products succeed on the platform, but also the type of products that will be brought forth in the future. Even if the New Music lists favor independent-label music and music by women, however, we cannot rule out that other playlists, and other promotional activity at Spotify, favor different sorts of music.

Still, if the results are correct, they beg the question of why Spotify would exercise bias in this fashion. We offer two pieces of speculation. First, Spotify may be responsive to the criticism – from independent labels and about treatment of women in the industry – that we cite above. This might lead them to actively promote work from the groups voicing concerns. Spotify's concern about gender issues led them to partner with Smirnoff on the "Smirnoff Equalizer," an app that linked to users' Spotify accounts to analyze "the percentage of music you listen to created by men versus women" (Bein, 2018). Second, to the extent that promotion of independent label music de-concentrates market power from Spotify's major-label suppliers, promotion of independent-label music may create a future environment more favorable for streaming rate negotiations.

Although our results are specific to one context – the New Music Friday lists at Spotify – our approach may have broader applicability. Our approach for measuring an agent's bias requires the agent's ex ante assessment of a product's commercial prospects, along with ex post measures of eventual commercial success. These elements are widely available in platform contexts. Platforms often display product options in some ranked order, descending in ostensible usefulness to consumers. Examples include Google search results, airline and hotel options are travel sites, and Marketplace vendors at Amazon. Conditional on the order in which options are ranked, do consumers ultimately purchase different types of products (for example, those with financial connections to the platforms) at different rates? With access to relevant data, our approach would allow easily implemented tests for bias.

An important last word is in order. Our finding that Spotify's New Music ranks display some bias in favor of independent label artists and women does not mean that these groups face a generally welcoming environment in the recorded music industry. There are reasons for concern that women are underrepresented among successful artists. Streams for music by women account for roughly a quarter of total streams, a share that is low compared with the shares of women in listening, among musicians, and – of course – in the population as a whole.[22] Our findings indicate that Spotify's New Music playlist rankings do not appear to compound the challenges women and independent label artists face in the recorded music industry.

[22]The US Bureau of Labor Statistics reported in 2020 that 31.8 percent of musicians and singers were women. See https://www.bls.gov/cps/cpsaat11.htm. On June 12, 2021, everynoise.com reported that 44.1 percent of Spotify listeners were women, while 22.8 percent of streams were of music by women. See https://everynoise.com/gender_tldr.html.

References

Aguiar, L. and J. Waldfogel (2018): "Quality predictability and the welfare benefits from new products: Evidence from the digitization of recorded music," *Journal of Political Economy*, 126, 492–524.

——— (forthcoming): "Platforms, Power, and Promotion: Evidence from Spotify Playlists," *Journal of Industrial Economics*.

Ahmad, E. and N. Stern (1984): "The theory of reform and Indian indirect taxes," *Journal of Public economics*, 25, 259–298.

Anderson, C. (2006): *The long tail: Why the future of business is selling less of more*, Hachette Books.

Arnold, D., W. Dobbie, and C. S. Yang (2018): "Racial bias in bail decisions," *The Quarterly Journal of Economics*, 133, 1885–1932.

Aswad, J. (2020): "Inside the Human Science of Spotify's New Music Friday Playlist," *Variety*, https://variety.com/2020/music/news/spotify-new-music-friday-playlist-1234843453/, Dec 4.

Ayres, I. and J. Waldfogel (1994): "A market test for race discrimination in bail setting," *Stanford Law Review*, 987–1047.

Bein, K. (2018): "Test Your Gender Bias with Honey Dijon, Smirnoff & Spotify's Equalizer: Exclusive." *Billboard Magazine*, https://www.billboard.com/articles/news/dance/8224527/honey-dijon-smirnoff-spotify-equalizer-womens-month., March 2.

Belleflamme, P. and M. Peitz (2018): "Inside the engine room of digital platforms: Reviews, ratings, and recommendations," .

Bourreau, M. and G. Gaudin (2018): "Streaming platform and strategic recommendation bias," .

Brynjolfsson, E., Y. Hu, and M. D. Smith (2003): "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers," *Management Science*, 49, 1580–1596.

Card, D., S. DellaVigna, P. Funk, and N. Iriberri (2020): "Are referees and editors in economics gender neutral?" *The Quarterly Journal of Economics*, 135, 269–327.

Datta, A., M. C. Tschantz, and A. Datta (2015): "Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination," *Proceedings on privacy enhancing technologies*, 2015, 92–112.

Datta, H., G. Knox, and B. J. Bronnenberg (2018): "Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery," *Marketing Science*, 37, 5–21.

Dredge, S. (2015): "10 things we learned from a day of indie labels talking digital music," *The Guardian*, https://www.theguardian.com/technology/2015/apr/10/things-we-learned-indie-labels-digital, April 10.

——— (2016): "Music curation and playlists: the new music battleground (#midem)," *Musically*, https://musically.com/2016/06/05/music-curation-and-playlists-the-new-music-battleground-midem/, June 5.

Duan, W., B. Gu, and A. B. Whinston (2008): "Do online reviews matter? – An empirical investigation of panel data," *Decision support systems*, 45, 1007–1016.

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

Ellis-Petersen, H. (2014): "Gender bias in the film industry: 75% of blockbuster crews are male," *The Guardian*, https://www.theguardian.com/film/2014/jul/22/gender-bias-film-industry-75-percent-male, July 22.

Flanagan, A. (2018): "Grammy President Neil Portnow To Step Down In 2019." *NPR.com*, https://www.npr.org/sections/therecord/2018/06/01/615889769/grammy-president-neil-portnow-to-step-down-in-201, June 1.

Forman, C., A. Ghose, and B. Wiesenfeld (2008): "Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets," *Information systems research*, 19, 291–313.

Gentzkow, M. and J. M. Shapiro (2010): "What drives media slant? Evidence from US daily newspapers," *Econometrica*, 78, 35–71.

Hagiu, A. and B. Jullien (2011): "Why do intermediaries divert search?" *The RAND Journal of Economics*, 42, 337–362.

Hellerstein, J. K., D. Neumark, and K. R. Troske (1999): "Wages, productivity, and worker characteristics: Evidence from plant-level production functions and wage equations," *Journal of labor economics*, 17, 409–446.

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

Ingham, T. (2018): "One Reason Why Spotify's Deals with the Major Labels are Balanced on a Knife-Edge." *Music Business Worldwide*, https://www.musicbusinessworldwide.com/one-reason-why-spotifys-deals-with-the-major-labels-rest-on-a-knife-edge/., November 13.

——— (2019): "Record Companies Can Now Pay Spotify to Promote Artists on the Platform Via Pop-up Music for You Alerts," *Music Business Worldwide*, https://www.musicbusinessworldwide.com/record-labels-can-now-pay-spotify-to-promote-artists-on-the-platform-via-pop-up-music-for-you-alerts/, October 24.

Knowles, J., N. Persico, and P. Todd (2001): "Racial bias in motor vehicle searches: Theory and evidence," *Journal of Political Economy*, 109, 203–229.

Lambrecht, A. and C. Tucker (2019): "Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads," *Management Science*, 65, 2966–2981.

NAYMAN, L. (2012): “Rock ’n’ Roll Payola: Dick Clark and Alan Freed,” *In These Times*, https://inthesetimes.com/article/13100/rock_n_roll_payola_dick_clark_and_alan_freed, April 24.

PELLY, L. (2018): “Discover Weakly.” *The Baffler*, https://thebaffler.com/latest/discover-weakly-pelly., June 4.

REINSTEIN, D. A. AND C. M. SNYDER (2005): “The influence of expert reviews on consumer demand for experience goods: A case study of movie critics,” *The journal of industrial economics*, 53, 27–51.

ROSS, T. W. (1984): “Uncovering regulators’ social welfare weights,” *The RAND Journal of Economics*, 152–155.

SCOTT, M. (2017): “Google Fined Record $2.7 Billion in E.U. Antitrust Ruling.” *New York Times*, https://www.nytimes.com/2017/06/27/technology/eu-google-fine.html., June 27.

SMART, S. AND J. WALDFOGEL (1996): “A citation-based test for discrimination at economics and finance journals,” *National Bureau of Economic Research*.

SMITH, S. L., M. CHOUEITI, AND K. PIEPER (2018): “Inclusion in the Recording Studio? Gender and Race/Ethnicity of Artists, Songwriters & Producers across 600 Popular Songs from 2012-2017,” *Annenberg Inclusion Initiative*, http://assets.uscannenberg.org/docs/inclusion-in-the-recording-studio.pdf.

SORENSEN, A. T. (2007): “Bestseller lists and product variety,” *The journal of industrial economics*, 55, 715–738.

STASSEN, M. (2020): “Spotify is Letting Record Labels Influence Personalized Recommendations So Long as they Pay for it in Royalties,” *Music Business Worldwide*, https://www.musicbusinessworldwide.com/spotify-is-letting-record-labels-influence-personalized-recommendations-so-long-as-they-pay-for-it-in-royalties/, November 20.

VARIAN, H. R. (2009): “Online ad auctions,” *American Economic Review*, 99, 430–34.

ZHU, F. AND Q. LIU (2018): “Competing with complementors: An empirical look at Amazon. com,” *Strategic management journal*, 39, 2618–2642.

A Figures and Tables

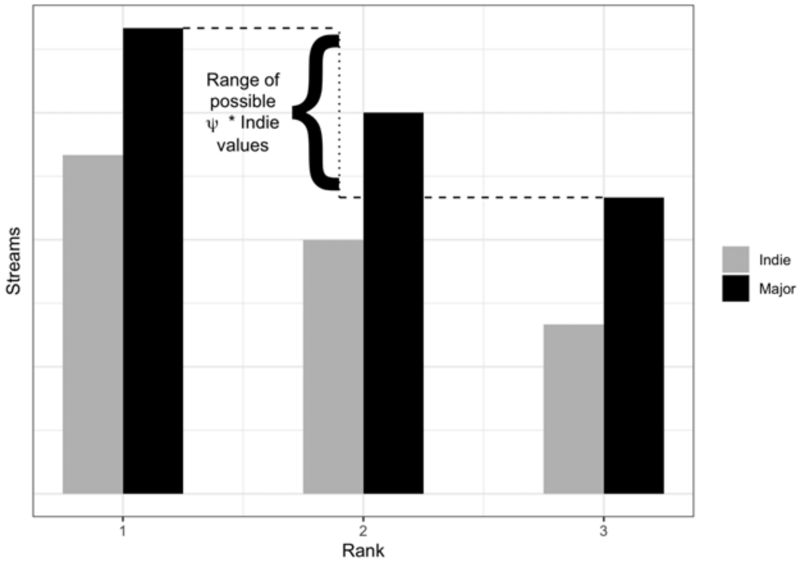

Figure 1: Theory Illustration.

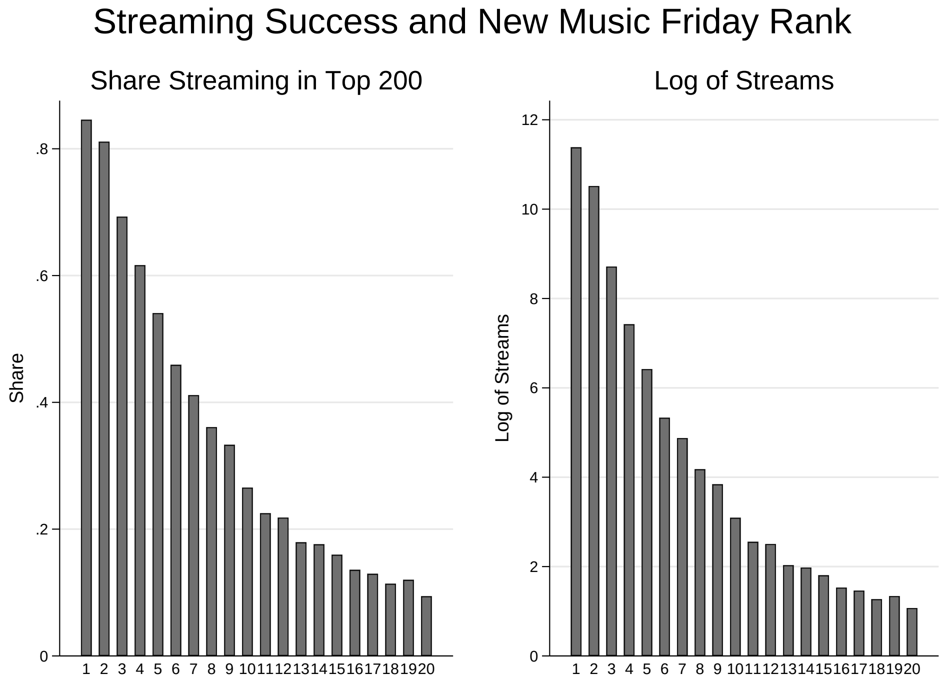

Figure 2: Share of Songs Streaming in Top 200 by New Music Friday Rank.

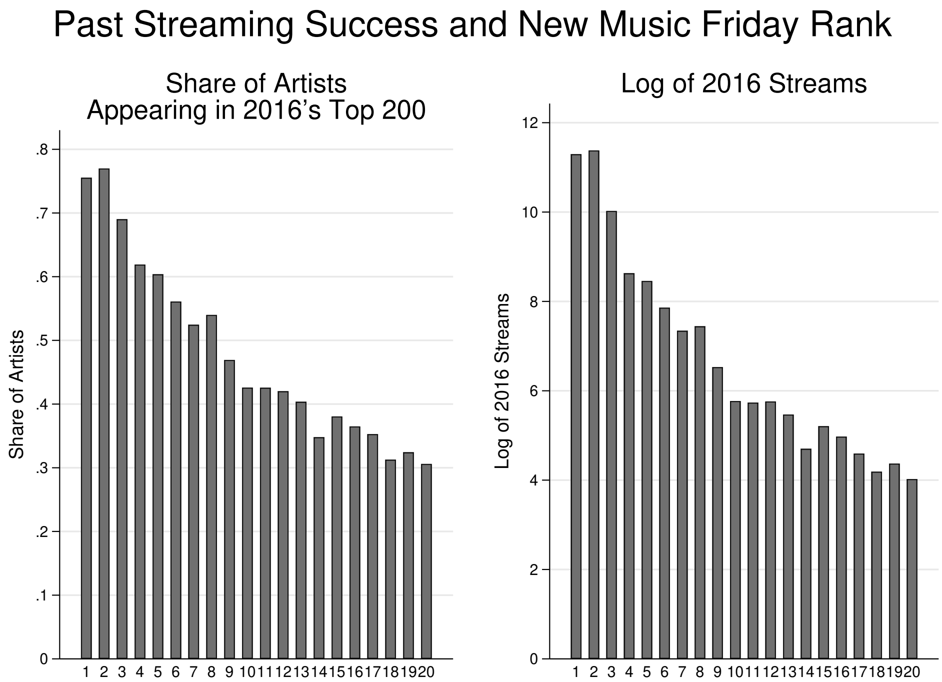

Figure 3: Artist Past Success by New Music Friday Rank.

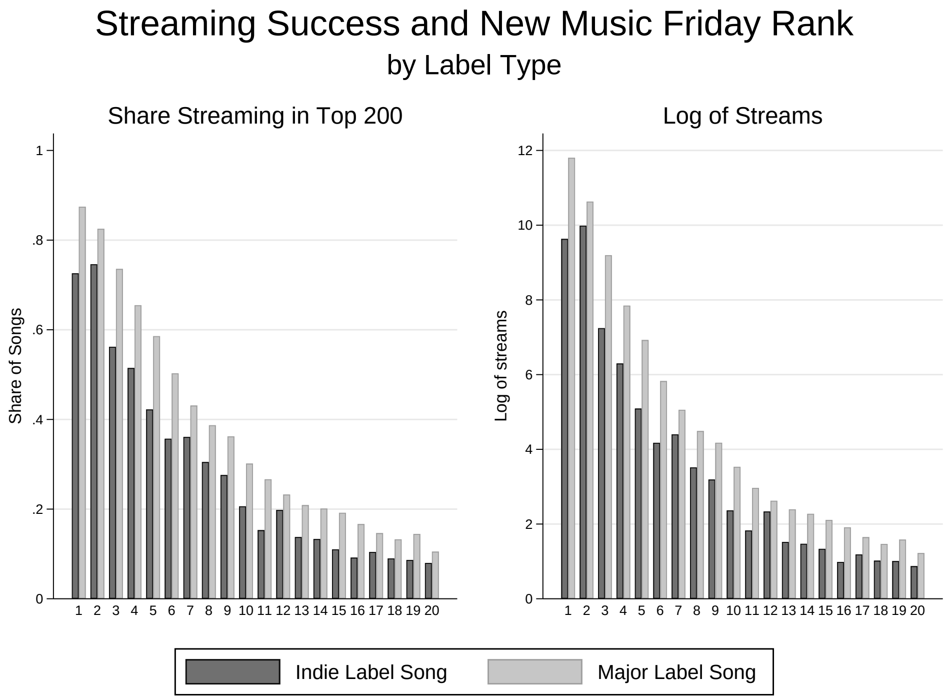

Figure 4: Song Success and New Music Friday Rank, by Label Type.

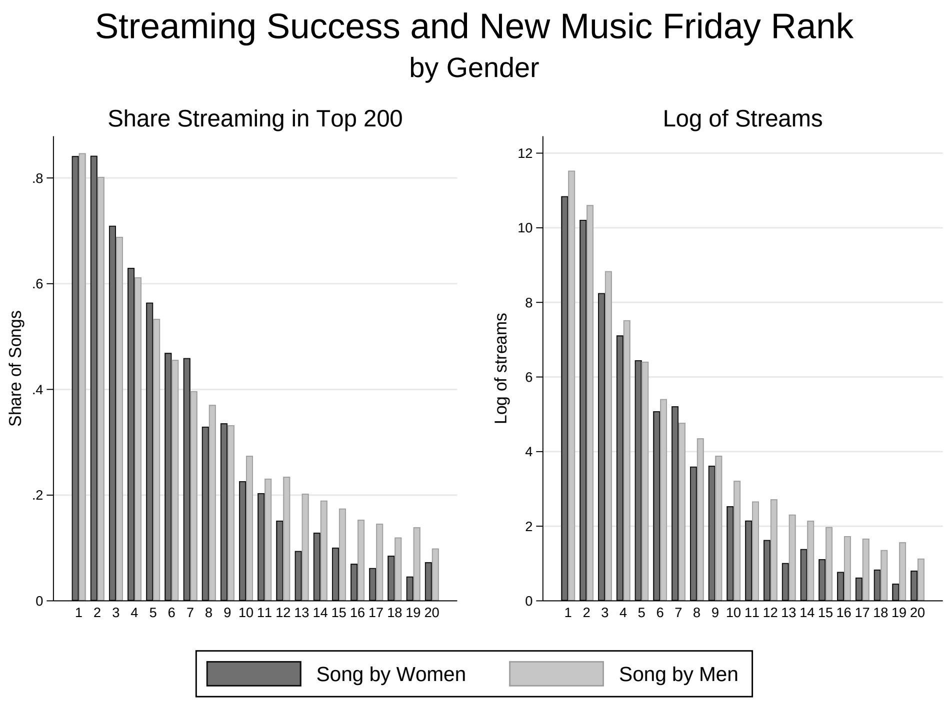

Figure 5: Song Success and New Music Friday Rank, by Gender.

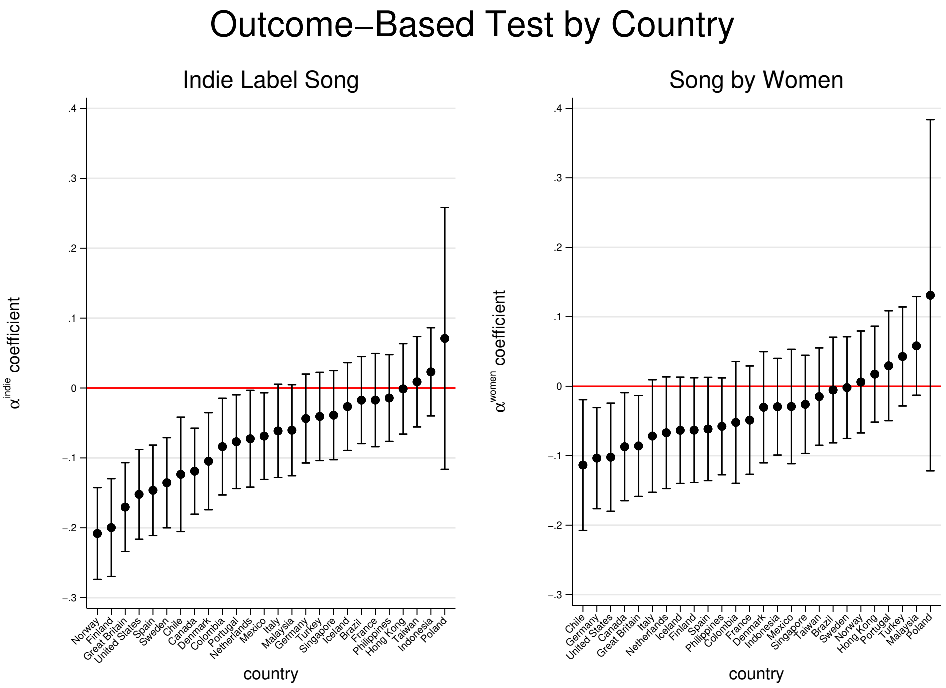

Figure 6: Outcome-Based Test, by Country.

Table 1: Descriptive Statistics.†

	Full Sample ($N = 14,747$)		Sub-Sample ($N = 12,366$)	
	Mean	Std. Dev.	Mean	Std. Dev.
New Music Friday (NMF) Rank	10.45	5.77	10.30	5.77
Song by Women	0.21	0.41	0.22	0.41
Major Label Song	0.66	0.47	0.68	0.47
Appears in Daily Top 200	0.35	0.48	0.35	0.48
Streams	0.96	5.64	0.93	5.74
First Week Streams	0.07	0.37	0.07	0.40
Spotify Playlists Followers, on NMF Add Day			2.96	8.74
Major-Label Playlists Followers, on NMF Add Day			0.23	0.93
Other Playlists Followers, on NMF Add Day			0.18	0.72

† Note: Streams and playlist followers are expressed in millions.

Table 2: Outcome-Based Tests for Bias: Does Indie and Music by Women Stream Differently Conditional on Initial Conditions? †

	(Top 200) Coef./s.e.	(Log Streams) Coef./s.e.	(Top 200) Coef./s.e.	(Log Streams) Coef./s.e.
Indie Label Song	-0.075***	-0.855***	-0.082***	-0.962***
	(0.007)	(0.083)	(0.007)	(0.085)
Song by Women	-0.035***	-0.659***	-0.036***	-0.682***
	(0.007)	(0.088)	(0.007)	(0.090)
New Music Friday Rank			-0.038***	-0.488***
			(0.001)	(0.007)
Rank Fixed Effects	✓	✓	✗	✗
Country Fixed Effects	✓	✓	✓	✓
R^2	0.359	0.363	0.334	0.325
No. of Obs.	14747	14747	14747	14747

† Dependent variables are measures of streaming success. Specifications (Top 200) measure streaming success with a binary measure of whether song j appears among the daily Top 200 streaming songs in country c during the year after entering the New Music Friday playlist. Specifications (Log Streams) use the logarithm of cumulated streams for song j in country c during the year after entering the New Music Friday playlist. Coefficients on independent label and songs by women show the degree of differential streaming success, conditional on initial New Music Friday rank assigned by curators. Robust standard errors are reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 3: Implied bias in conditioning-on-observables test compared with outcome-based test †

	New Music Friday Rank Coef./s.e.	Appears in Top 200 Coef./s.e.	Log of Streams Coef./s.e.
Indie Label Song	1.074***	-2.177***	-1.970***
	(0.102)	(0.190)	(0.182)
Song by Women	-0.758***	-0.967***	-1.398***
	(0.116)	(0.199)	(0.186)
R^2	0.116	0.334	0.325
No. of Obs.	14494	14747	14747
Bias Difference: Indie		3.251	3.044
P-value		0.000	0.000
Bias Difference: Women		0.209	0.639
P-value		0.361	0.003

† The dependent variable in the first specification is the New Music Friday rank. The regression also includes measures of past streams for the artists, country fixed effects and their interaction with past streams, as well as song characteristics. The second and third specifications report the degree of bias in New Music Friday Rank implied by the last two columns in Table 2. The lower panel reports results of tests of whether the two approaches deliver equal levels of rank bias. Robust standard errors are reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 4: Outcome-Based Tests for Bias: Initial Streaming Success †

	(Top 200) Coef./s.e.	(Log Streams) Coef./s.e.	(Top 200) Coef./s.e.	(Log Streams) Coef./s.e.
Indie Label Song	-0.071***	-0.782***	-0.078***	-0.871***
	(0.006)	(0.071)	(0.007)	(0.072)
Song by Women	-0.035***	-0.530***	-0.037***	-0.549***
	(0.007)	(0.076)	(0.007)	(0.078)
New Music Friday Rank			-0.037***	-0.426***
			(0.001)	(0.006)
Rank Fixed Effects	✓	✓	✗	✗
Country Fixed Effects	✓	✓	✓	✓
R^2	0.376	0.387	0.349	0.352
No. of Obs.	14747	14747	14747	14747

† Dependent variables are measures of streaming success. Specifications (Top 200) measure streaming success with a binary measure of whether song j appears among the daily Top 200 streaming songs in country c while the songs are on the New Music Friday playlist. Specifications (Log Streams) use the logarithm of cumulated streams for song j in country c while the songs are on the New Music Friday playlist. Coefficients on independent label and songs by women show the degree of differential streaming success, conditional on initial New Music Friday rank assigned by curators. Robust standard errors are reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 5: Outcome-based test, controlling for initial conditions. †

	Ultimate Streaming Success				Initial Streaming Success			
	(Top 200) Coef./s.e.	(Log Streams) Coef./s.e.	(Top 200) Coef./s.e.	(Log Streams) Coef./s.e.	(Top 200) Coef./s.e.	(Log Streams) Coef./s.e.	(Top 200) Coef./s.e.	(Log Streams) Coef./s.e.
Indie Label Song	-0.070***	-0.761***	-0.069***	-0.748***	-0.066***	-0.724***	-0.064***	-0.707***
	(0.007)	(0.093)	(0.008)	(0.094)	(0.007)	(0.079)	(0.007)	(0.080)
Song by Women	-0.032***	-0.627***	-0.035***	-0.644***	-0.033***	-0.510***	-0.034***	-0.523***
	(0.008)	(0.096)	(0.008)	(0.096)	(0.008)	(0.083)	(0.008)	(0.084)
Spotify playlist followers (mil)			0.000	0.001			0.000	0.006
			(0.001)	(0.010)			(0.001)	(0.009)
Major-label playlist followers (mil)			0.018***	0.169**			0.019***	0.156**
			(0.007)	(0.085)			(0.007)	(0.075)
Other playlist followers (mil)			-0.029***	-0.254**			-0.026***	-0.248***
			(0.008)	(0.105)			(0.008)	(0.092)
Rank Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Country Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.351	0.355	0.352	0.356	0.366	0.379	0.367	0.379
No. of Obs.	12366	12366	12366	12366	12366	12366	12366	12366

† Dependent variables are measures of streaming success. For the Ultimate Streaming Success panel, specifications (Top 200) measure streaming success with a binary measure of whether song j appears among the daily Top 200 streaming songs in country c during the year after entering the New Music Friday playlist, and specifications (Log Streams) use the logarithm of cumulated streams for song j in country c during the year after entering the New Music Friday playlist. For the Initial Streaming Success panel, specifications (Top 200) measure streaming success with a binary measure of whether song j appears among the daily Top 200 streaming songs in country c while the songs are on the New Music Friday playlist, and specifications (Log Streams) use the logarithm of cumulated streams for song j in country c while the songs are on the New Music Friday playlist. Coefficients on independent label songs and songs by women show the degree of differential streaming success, conditional on initial New Music Friday rank assigned by curators. Robust standard errors are reported in parenthesis.

* Significant at the 10% level.

** Significant at the 5% level.

*** Significant at the 1% level.

Table 6: Curator weights on independent (women) relative to major (men) streams.†

Song Type	Appearing in Top 200	Log of Streams	Appearing in Initial Top 200	Log of Initial Streams
Song by Women	1.08	1.17	1.10	1.17
	0.03	0.04	0.03	0.04
Indie-Label Song	1.40	1.41	1.44	1.45
	0.05	0.04	0.05	0.05

† Bootstrapped standard errors are reported below estimates ($N = 100$).

B Appendix

B.1 Alternative Song Gender Measures

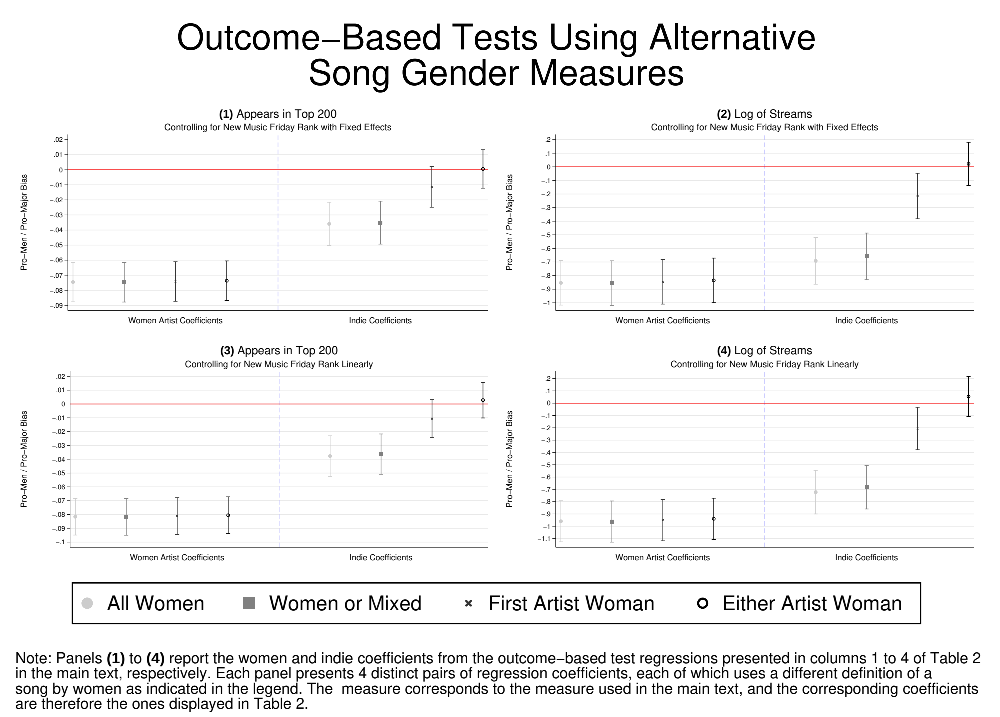

Figure B.1: Outcome-Based Tests Using Alternative Song Gender Definitions.

Table B.1: Curator weights on women relative to men streams: Robustness to alternative song gender definitions. †

Song Type	Appearing in Top 200	Log of Streams	Appearing in Initial Top 200	Log of Initial Streams
Women-Only	1.10	1.19	1.10	1.17
	0.03	0.05	0.03	0.04
Women or Mixed	1.08	1.17	1.10	1.17
	0.03	0.04	0.03	0.04
First Artist Woman	1.04	1.09	1.01	1.07
	0.02	0.03	0.03	0.04
Either Artist Woman	1.00	1.02	1.00	1.03
	0.03	0.04	0.03	0.03

† Bootstrapped standard errors are reported below estimates ($N = 100$).