Journal of Economic Behavior and Organization 194 (2022) 550–567

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Bestseller lists and product discovery in the subscription-based market: Evidence from music streming

Jaeung Sim[a], Jea Gon Park[b], Daegon Cho[c,*], Michael D. Smith[d], Jaemin Jung[e]

[a] College of Business, Korea Advanced Institute of Science and Technology (KAIST), Seoul, South Korea
[b] Hyperconnect Inc., Seoul, South Korea
[c] College of Business, Korea Advanced Institute of Science and Technology (KAIST), Seoul, South Korea
[d] Heinz College, Carnegie Mellon University, Pittsburgh, PA, United States
[e] Moon Soul Graduate School of Future Strategy, Korea Advanced Institute of Science and Technology (KAIST), Seoul, South Korea

ARTICLE INFO

Article history:
Received 20 March 2020
Revised 12 December 2021
Accepted 20 December 2021
Available online 13 January 2022

JEL codes:
D83
L82
M31
O39

Keywords:
Streaming music
Bestseller lists
Saliency effect
Observational learning

ABSTRACT

Our study analyzes the impact of hourly-updated bestseller lists on music discovery in a digital streaming platform to provide evidence of whether and why bestseller lists affect consumer decisions in the subscription-based market. We circumvent the problem of demand-popularity simultaneity by leveraging high-frequency data and a regression discontinuity design. We find that being added to the top 100 charts increases song discovery by 11–13%. Furthermore, a series of analyses suggest that the saliency effect, instead of observational learning, is more likely to drive this behavioral change among streaming users. Specifically, we find that a song's chart entrance increases repeat consumption, normative rank positions within the top 100 lists do not demonstrate significant discontinuity, and an artist or a song's prior popularity does not moderate this effect.

1. Introduction

Ample literature has shown that consumers learn the quality and fit of products by observing other consumers' choices—i.e., observational learning—via bestseller lists, reviews, recommendations, and information spillovers (Cai et al., 2009; Godinho de Matos et al., 2016; Hendricks and Sorensen, 2009; Kumar et al., 2014; Sorensen, 2017; Waldfogel, 2016). Digital technology and Internet use have boosted, and in some cases enabled, the use of these information sources. On the demand side, online channels allow customers to easily obtain product information updated in real-time. On the supply side, providing information on consumers' past purchases is inexpensive and easily automated. Such real-time information has become commonplace in today's online websites and is an important determinant of demand.

While numerous technologies have led consumers to rely on such information sources in making decisions, some advances may empower consumers to make independent decisions. For instance, online subscription-based platforms, including Netflix, Spotify, and Apple Music, provide bundles of online content at zero marginal cost to their users. Under this

* Corresponding author.
E-mail addresses: jaeung@kaist.ac.kr (J. Sim), jpnova24@gmail.com (J.G. Park), daegon.cho@kaist.ac.kr (D. Cho), mds@cmu.edu (M.D. Smith), nettong@kaist.ac.kr (J. Jung).

https://doi.org/10.1016/j.jebo.2021.12.030
0167-2681/

all-you-can-eat pricing plan, subscribers do not encounter the financial burden of purchase decisions, as they did under à la carte channel sales (Aguiar and Waldfogel, 2018a; Chen et al., 2016; Godinho de Matos et al., 2016). Datta et al. (2017) found that adopting music streaming services allowed consumers to discover lesser-known artists and songs. Because of these reduced financial costs, consumers may have lower incentives to learn from others' choices in subscription-based markets. Despite the economic significance for content suppliers and platforms, little research has clearly examined this possibility so far.

Our study aims to capture whether and why peer-based information sources affect product discovery in digital streaming services, focusing on bestseller lists and music streaming. Answering this question involves extending the literature to address two main empirical challenges: (1) the simultaneity of consumption and positions on the lists in most aggregate data (Aguiar and Waldfogel, 2018b), and (2) separating repeat consumption from the number of new listeners (Aguiar and Waldfogel, 2018b; Chen et al., 2016; Datta et al., 2017). We tackle these problems using a novel and unique dataset with hourly ranking charts from a large digital music platform in South Korea, where digital music streaming services were the first and most prevalent in the world. Since the platform updates rank positions hourly and streaming counts in real-time, we can sharply distinguish consumption changes due to shifts in rank positions. Furthermore, this dataset allows us to separate the number of new listeners from the total number of times a track was played, enabling us to count the repeat consumption.

Our study uses a sample that includes the top 250 songs every hour played between March 5 and April 2, 2017 and information on the exact timing and number of plays on the service platform. It is worth noting that the top 100 most played songs are visible to users on the front page, but the remaining 101–250 selling lists are not publicly available. Leveraging this sharp discontinuity on the charts, we use a regression discontinuity design (RDD)—which compares songs just above the cutoff with those just below the cutoff—to identify the causal relationship between being listed on the ranking charts and song discovery.

We document a positive impact of placing on the top 100 charts as a discovery-based consumption. The regression results show that being added to the top 100 charts leads to 11–13% more new listeners, and as a result, total streaming counts increases by 2–3%. We provide a series of results supporting the causal interpretation that the ranking charts affect music discovery. We obtain consistent results across different margins of the RDD, functional forms, and sample selections. We also conduct a falsification test by utilizing placebo cutoffs instead of the top 100th position and find no significant relationships, indicating that our results are not spurious.

Contrary to the prior studies, we find that the increase in streaming consumption is more consistent with a saliency effect than with observational learning. First, we find that being listed on the ranking charts also increases repeat consumption—that is, bestseller lists affect streamers even when they are already aware of a song's quality. Second, normative positions within the top 100 ranking charts—such as top 10, top 20, top 25, and top 50—do not create a significant discontinuity, suggesting that streaming users are unlikely to choose content based on perceived quality gaps which normative numbers might imply. Lastly, we examine the contingency of this effect by factors associated with awareness of artists and songs and find no significant heterogeneity across these factors. These findings imply that bestseller lists still significantly influence consumer choices in subscription-based streaming services, but they do so through an alternative mechanism, possibly due to negligible marginal costs of consumption and choice overload in the digital music market.

The remainder of this paper is structured as follows. Section 2 provides background and related literature on two issues related to our research question: music streaming and the impact of bestseller lists. In Section 3, we describe our dataset, variables, and identification strategy. Section 4 presents our empirical analysis results, including the RDD estimates and a set of robustness checks. In Section 5, we discuss our results and conclude.

2. Background

2.1. Music streaming

Streaming services provide consumers with real-time access to a bundle of digital content without transferring ownership of the content. This model contrasts with an ownership model where consumers pay to download and own individual pieces of content. In the context of digital music, streaming services can be categorized as either interactive or non-interactive (Aguiar and Waldfogel, 2018a). Interactive (non-interactive) services do (not) allow consumers to choose which songs to stream. Most major digital music distributors, including Apple Music, Spotify, and YouTube Music are interactive services.

Subscription services are frequently sold through either (or both) a monthly subscription (flat-rate) model or an ad-supported free model (Thomes, 2013). The subscription model provides unlimited access to a bundle of music without commercials, while the ad-supported model provides (limited or unlimited) access to the music with advertisements. In 2015, the market size of the subscription model was estimated at 2 billion USD, whereas the ad-supported free model was estimated at 634 million USD (IFPI, 2016), with subscription-based services growing by 58.9% annually, and ad-supported services growing by 11.3% annually.

2.2. Related literature

This study is closely related to the literature on bestseller lists and product discovery. Bestseller lists are considered to be an impactful market mechanism that affects product discovery, along with reviews, recommendations, and information spillovers (Sorensen, 2017). Bestseller lists represent consequences of previous consumer choices and can affect subsequent choices of new consumers in several ways (Carare, 2012; Cai et al., 2009). First, a consumer may choose a product following other consumers because she believes that these consumers have superior information to her on product quality (i.e., observational learning). Second, a consumer may imitate others' choices because she prefers to act according to the dominant trend in the market (i.e., the conformity effect). Third, bestseller lists might increase demand because when consumers are not aware of the entire choice set, they may be more likely to choose more salient options (i.e., the saliency effect).

Numerous studies have examined the effect of bestseller lists on product sales in a variety of contexts. Salganik et al. (2006) developed a large artificial music market and conducted experiments to examine the impact of information on previous participants' choices. The authors found that popularity information shapes participants' choices even after listening to songs by themselves, and the salience of such information amplifies this effect. Sorensen (2007) found that a book's listing on the New York Times bestseller lists increases its sales by 4.3% on average, and this effect is more substantial for new authors. Cai et al. (2009) conducted a randomized field experiment in a restaurant setting and showed that providing top-selling lists increases dishes presented on the lists, whereas lists of randomly selected dishes have no significant impact on consumer choices.

Recent studies have investigated platform-offered bestseller lists in digital markets. Tucker and Zhang (2011) examined how popularity-based rankings affect the number of website visits and found that these rankings disproportionately benefit niche products with narrow appeal. Carare (2012) examined the impact of today's rank on subsequent demand in a mobile app marketplace. The author found that today's bestseller rank increases tomorrow's mobile app demand, and the effect is substantially greater for paid apps than free apps, supporting observational learning instead of the saliency effect. Godinho de Matos et al. (2016) analyzed the impact of randomly changed lists of most popular movies on video-on-demand sales. They found that artificial swaps of rank positions have a significant impact on demand in the short term, but consumers rapidly adjust their choices toward the true quality of movies based on outside information. Our study is most closely related to Aguiar and Waldfogel (2018b), who examined the economic impact of the inclusion and exclusion of songs from Spotify playlists on streaming revenue. They found that being added to Today's Top Hits on Spotify increases streaming counts by nearly 20 million and is worth between $116,397 and $162,956.

To the best of our knowledge, our research provides the first empirical evidence that observational learning has a limited moderating impact on the relationship between bestseller lists and product discovery. Unlike most of the related studies demonstrating significant heterogeneity across factors associated with awareness of products, we find no evidence that these factors lead to a significant contingency of the bestseller-list effect. Furthermore, we find that being added to the ranking charts increases repeat consumption as well, whereas normative numbers within the top 100 cutoff do not make a significant discontinuity. These results are more consistent with the saliency effect than observational learning, which most of the prior works supported.

We suggest that zero marginal costs and large product catalogs in subscription-based services are responsible for these findings. To be specific, music streamers on subscription-based platforms do not pay additional costs for listening to new music. Thus, they might be insensitive to quality signals in choosing songs from large catalogs. While such results are consistent with Carare (2012), who found that free apps are less sensitive to product rankings than paid apps, observational learning could be further weakened because music listening accompanies smaller search costs for product quality and more frequent decisions to use products than mobile app downloads.

Our study also makes several empirical contributions to the literature. First, we use fine-grained hourly data to provide relatively more precise estimates of the impact of bestseller lists on streaming demand. Since Spotify only provides daily updated information of top-selling charts and streaming counts, Aguiar and Waldfogel (2018b) noted that they could not distinguish treated and untreated days sharply. In our analysis, we overcome this limitation by using hourly updated data, which can better capture the exact moment of rank changes presented to consumers, allowing us to identify treated and untreated songs and immediate responses to the chart entrance. Second, our data allow us to separate the number of new listeners from repeat consumption, which comprises previous listeners' behaviors. By doing so, we can distinguish the informational effect of ranking charts from other confounders.

We summarize the related literature and our contributions in Table 1.

2.3. Research framework

Fig. 1 illustrates the conceptual framework for ranking charts and music consumption. Music consumption of a focal song and its competitors in the previous hour determines its current ranking. Being listed on the charts may make the song more salient and help consumers learn about its quality. If the path through the product saliency is dominant the charts will significantly increase repeat consumption as well as the number of new listeners. Conversely, if the path through observational learning is the primary channel, the ranking charts will mainly affect consumers unaware of the focal song. Moreover, the significance of this path will be moderated by outside factors affecting product awareness, such as the focal song's artist, label, and promotion activities (Dewan and Ramaprasad, 2012; Hendricks and Sorensen, 2009). It is worth noting that songs listed on the ranking charts may concurrently appear on other popular playlists, which could further boost demand for those songs (Aguiar and Waldfogel, 2018b).

As described in the following section, our dataset allows us to observe current rank positions, the number of new listeners, total streaming counts, and song/artist characteristics. We alleviate the issue of unobserved promotion activities using

Table 1
Summary of related literature.

Study	Research context (marginal price)	Independent variable	Dependent variables (results)	Explored mechanisms (supported)
Salganik et al. (2006)	An artificial market for music downloads (Y)	Download counts Salience of the counts	Inequality of music success (+ for download counts) (+ for saliency of the counts)	Social influence (Y) Salience of social information (Y)
Sorensen (2007)	Physical book sales for hardcover fiction titles (Y)	Being listed on the New York Times bestseller list	Sales of bestsellers (+4%) Sales of non-bestsellers similar to bestsellers (+)	Informational effect (Y) Promotional effect (N)
Cai et al. (2009)	Restaurant dining (Y)	Being listed on the top five most popular dishes	Sales of dishes (+13–20%)	Observational learning (Y) Saliency effect (N)
Tucker and Zhang (2011)	A website that lists wedding service vendors (N)	Websites popularity rankings	Website clicks (+)	Observational learning (Y)
Carare (2012)	Apple's App Store (Y/N)	Being listed on the top lists and sales rankings	Downloads of paid apps (+) Downloads of free apps (n.s.)	Observational learning (Y) Saliency effect (N)
Godinho de Matos et al. (2016)	A video-on-demand market in a telecommunications company (Y)	Peer-rating rankings	Video-on-demand sales for popular movies (+ for a short term)	Herding effect (Y) Outside information (Y)
Ursu (2018)	Hotel search results at Expedia (Y/N)	Search rankings	Search intensity (+) Purchase (+$1.92/position) Purchase conditional on search (n.s.)	Lowering search costs (Y) Affecting consumer expectations/utility (N)
Aguiar and Waldfogel (2018b)	Daily music streaming on Spotify (N)	Being listed on the Global Top 50 playlist	Streaming counts (+4%)	Not suggested
This study	Hourly music streaming in a Korean platform (N)	Being listed on the hourly top 100 ranking charts	New streamers (+11–13%) Streaming counts (+2–3%)	Observational learning (N) Saliency effect (Y)

Note. (Y): Yes; (N): No; (Y/N): Different across dependent variables; (n.s.): not significant.

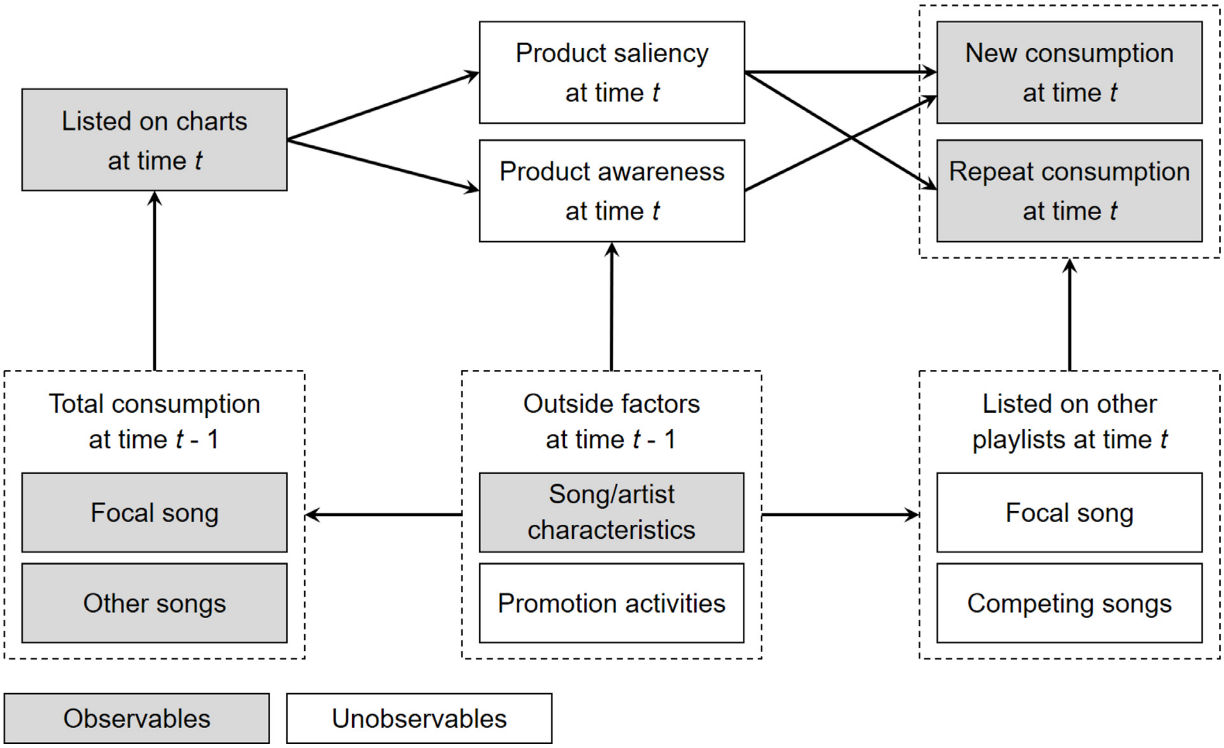

Fig. 1. Conceptual framework for ranking charts and music consumption.

Table 2
Summary statistics.

Variable	Description	Obs.	Mean (Std. Dev.)	Min.	Median	Max.
Rank	Rank of a song displayed at time t determined by consumption at time $t - 1$	123,818	101.01 (61.38)	1	97	250
Streaming counts	Total number of streaming counts at time t	123,818	2012.55 (2624.54)	112	1150.5	56,043
New streamers	The number of unique streamers who listen to a song at time t for the first time	123,818	114.31 (465.37)	0	47	23,203
Major label[a]	1 if a song is released by a major label, 0 otherwise	123,818	0.41 (0.49)	0	–	1
Foreign	1 if an artist is not a Korean artist, 0 otherwise	123,818	0.13 (0.34)	0	–	1
Years since debut	The number of years passed since an artist's debut	121,459[b]	5.81 (4.96)	0	5	27
Days since release	The number of days passed since a song's release date	123,818	237.18 (353.22)	0	112	5265
Album title	1 if a song is an album title track, 0 otherwise	123,818	0.23 (0.42)	0	–	1

Notes. The raw data consist of 161,750 observations. Among them, 123,818 observations were consecutively observed since time t - 2.
[a] Major labels indicate Universal Music, Warner Music Group, Sony Entertainment (international labels), CJ ENM, JYP Entertainment, SM Entertainment, YG Entertainment, Loen Entertainment, and FNC Entertainment (domestic labels). The rest of the labels are categorized as indie labels.
[b] 2359 observations from 13 unique artists are dropped due to a lack of debut information.

two approaches: 1) employing a first-difference model with lagged differences and fixed effects, which might capture linear increases in demand induced by promotions, and 2) examining interaction effects between the ranking charts and cross-sectional variations of product awareness captured by music labels, artists, and song characteristics. Although we do not track other playlists, as explained in Section 3, the ranking charts of interest were updated far more frequently (hourly) than other playlists (daily). Therefore, it is doubtful that the platform added the vast majority of songs to other playlists exactly when these songs newly entered the ranking charts. We leverage this institutional background by focusing on demand changes right after a song was added to/removed from the top 100 charts.

3. Identification strategy

3.1. Data source and measurement

We obtained data from one of the largest digital music platforms in South Korea, with approximately two million monthly active users and over 15 million tracks available for streaming. We developed a web crawler to retrieve ranking charts and streaming counts of listed songs and collected the data at hourly intervals from March 5 to April 2 in 2017, except for March 16, when the platform was temporarily unavailable due to checking its server. During the research period, the platform displayed hourly-updated ranking charts of the top 100 tracks on its web and mobile services.[1] Also, the platform disclosed the top 101–250 selling lists via accessible web addresses, but not through their built in web interface at that time. Thus, consumers were not exposed to such lists on the graphical user interface.[2] This provides us with a unique opportunity to compare chart rankers with unranked tracks.

In this platform, each track's cumulative counts of total streaming and unique listeners were updated in real-time. To calculate hourly streaming counts and new listeners, we subtract the cumulative values in the previous hour from those in the current hour. Note that the subtracted number of unique listeners does not include any streamer who had previously listened to this track because his/her repeated consumption increased total streaming only.

Table 2 describes our variables and their summary statistics. Correlations among these variables are shown in Table A1. We define *rank* as a position displayed at time t, which is constructed based on consumption at time t – 1. Prior studies, such as Aguiar and Waldfogel (2018a, 2018b), were limited by Spotify listings to capturing *streaming counts*, expressed as the sum of initial and repeat consumption. In our data, we can observe the number of *new streamers* to rule out the influence of ranking charts in previous hours. From our data, we observe that repeat consumption represents 94% of total consumption on average, supporting the necessity of separating the number of new listeners from repeat consumption.

Fig. 2 provides a graphical analysis of demeaned streaming counts and new streamers for each rank just above and below the cutoff of the ranking charts. Note that the horizontal axis indicates rank positions displayed at time t, which were constructed based on consumption at time t - 1, and the vertical axis denotes consumption occurred at time t. We observe a discontinuous jump in streaming demand at the cutoff. Also, we see that the number of new streamers does not

[1] Track rank was constructed by the weighted sum of streaming counts and downloads.
[2] Shortly after our research period, the platform included top 101-200 lists in its visible ranking charts.

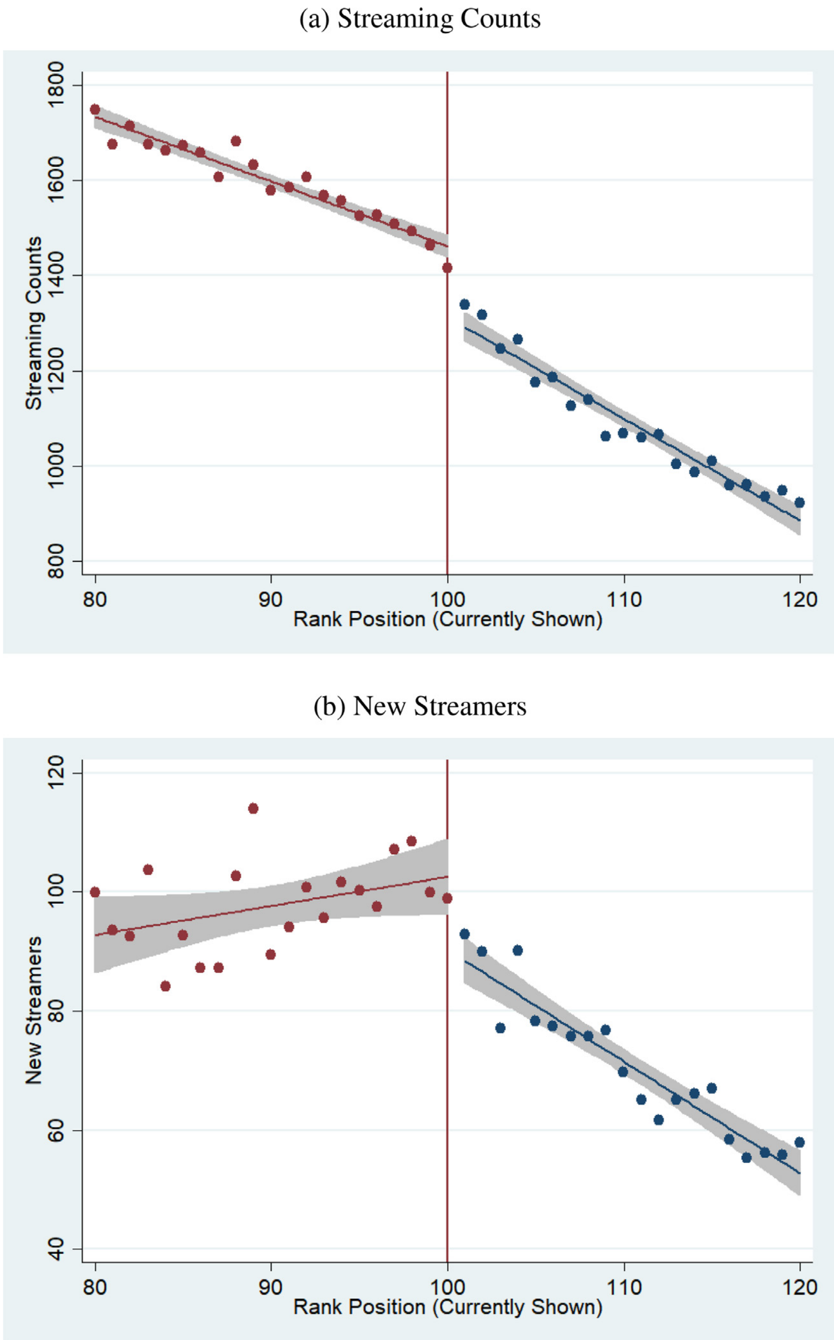

Fig. 2. Average streaming counts and new streamers around the top 100th position.

significantly increase as a song's rank position increases (the left side in Fig. 2b). This implies that the rank differences in total streaming across chart rankers close to the top 100 boundary are mainly attributable to repeat consumption rather than new listeners.

It is worth noting that Fig. 2 does not indicate whether a song maintained or lost its listeners compared to the previous hour. To focus on how differently the numbers changed across rank positions, we visualize the differences in streaming counts and new streamers in Fig. 3. The results demonstrate relatively sharp contrasts around the top 100 cutoff than Fig. 2, suggesting that the relative demand better captures the effect of being listed on the top 100 charts. We also observe that the differences in streaming counts and new streamers decrease as rank positions increases (numerically smaller) around

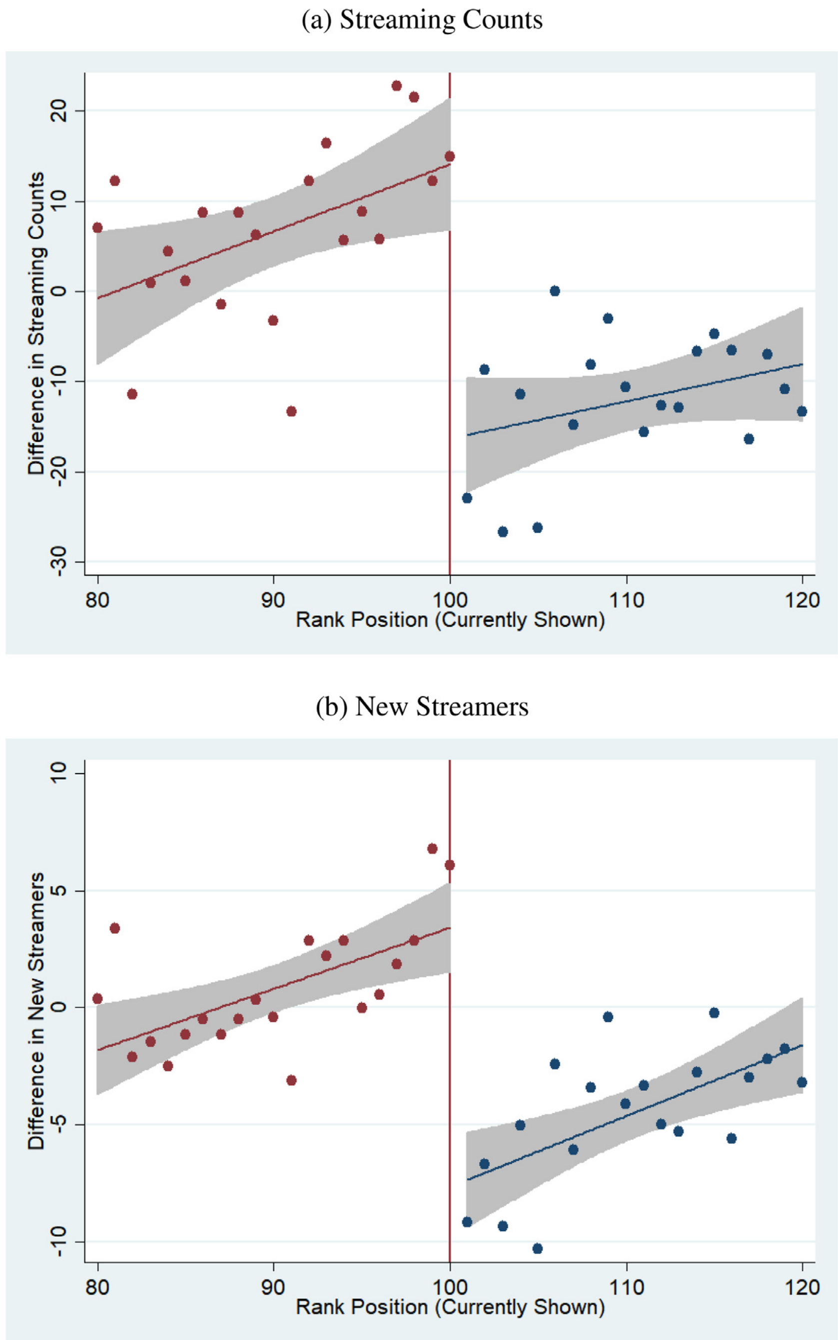

Fig. 3. Average differences in streaming counts and new streamers around the top 100th position.

the top 100 cutoff. Considering that the differences are relatively flat when rank positions are far from this boundary (see Fig. A2), such patterns seem attributable to entering and exiting from the ranking charts.[3]

[3] Rank positions just above (below) the boundary are more likely to be newly added to (removed from) the top 100 lists in the current hour. Therefore, the differences in streaming counts and new listeners are more probable to be contributed by chart entrance (exit) than rank positions farther from the cutoff. Our estimates of the main specification (Table 3) also support this argument. We find no statistically positive coefficients for [Rank - 100] (t) and [Rank - 100] x Top 100 Listed (t) from the first-difference models, suggesting that these patterns are mostly attributable to being newly added to (removed from) the top 100 lists and other control variables.

3.2. Econometric specification

We use a regression discontinuity design (RDD) to estimate the causal effect of being added to the top 100 ranking charts on the discovery of songs. The RDD is a quasi-experimental approach that employs discontinuous changes in the probability of receiving treatment (Hahn et al., 2001). We exploit the fact that being added to the lists is determined by the sharp cutoff. Specifically, if a song's streaming count is sufficiently close to that of the top 100th song, we can postulate that the difference in rank between the two songs mostly originate from random perturbation, which is not associated with future streaming demand. In other words, a group of songs just below the top 100 cutoff are statistically indistinguishable from a group just above.

We perform the RDD by using a local linear regression (Imbens and Lemieux, 2008; Lee and Lemieux, 2010). Using the 100th song's hourly streaming counts as a benchmark, we include songs of which streaming counts fall into a certain range. For the songs within this range, we estimate the following models:

$$\operatorname{arcsinh}(y_{it}) = \beta_1 \cdot \text{Top 100 Listed}_{it} + \gamma_{11} \cdot [\text{Rank}_{it} - 100] + \gamma_{12} \cdot [\text{Rank}_{it} - 100] \cdot \text{Top 100 Listed}_{it} + \alpha(i, t : \theta) + \varepsilon_{it}, \tag{1}$$

where subscripts i and t refer to songs and time, respectively; y_{it} is a dependent variable either the number of new streamers or total streaming counts at time t; $\operatorname{arcsinh}(y_{it}) \equiv \ln(y + \sqrt{y^2 + 1})$; Top 100 Listed$_{it}$ is a dichotomous variable indicating whether a displayed rank position at time t, Rank$_{it}$, is higher (numerically smaller) than or equal to the top 100th position; $\alpha(i, t : \theta)$ indicates a set of controls comprising the number of streaming counts and new streamers at time t – 1, hour-of-the-day fixed effects, day fixed effects, and song fixed effects; ε_{it} is an error term clustered at the song level to account for autocorrelation in the data (Bertrand et al., 2004; Moulton 1990).[4]

The focus of this model is the magnitude and statistical significance of β_1. If the coefficient is positive and statistically significant, we conclude that being added to the top 100 chart increases discovery of songs. γ_{11} (γ_{12}) presents the local linear relationship between the number of new streamers and rank positions just below (above) the top 100th position.

Our dataset includes several zero values of new streamers. In this case, taking the logarithm by adding one to the variable prior to its transformation might distort the actual elasticity of our estimates. To alleviate this possibility, we employ the inverse hyperbolic sine (or arcsinh) transformation that allows us to retain zero-valued observations (Bellemare and Wichman 2020; Burbidge et al., 1988; MacKinnon and Magee 1990). We quantify the local treatment effect by using Bellemare and Wichman (2020)'s small-sample bias corrected approximation of arcsinh-linear elasticity with dummy independent variables as follows:

$$\frac{\hat{\tilde{P}}}{100} \approx \exp\left(\hat{\beta} - 0.5\widehat{\text{Var}}\left(\hat{\beta}\right)\right) - 1,$$

where $\hat{\tilde{P}}$ is an approximation of the proportional effect for large dependent variables; $\widehat{\text{Var}}(\hat{\beta})$ is the estimated variance of $\hat{\beta}$, i.e., the coefficient of a dummy independent variable.

We utilize several choices of range in order to assess sensitivity of our findings (Cao et al., 2019; Luca 2011). Specifically, we use the following margins from the top 100th song's streaming count: [−10%, +10%], [−5%, +5%], [−2%, +2%], and [−1%, +1%]. For instance, if the top 100th song was played 100,000 times at time t – 1, songs that were played between 95,000 and 105,000 times belong to the range, [−5%, +5%]. We also restrict our observations to a specific range of rank positions to focus on the local treatment effect around the top 100 cutoff.

One may raise a concern that entering the top 100 charts might be associated with other popular playlists. Also, consumers may respond more positively to songs to which they are repeatedly exposed (Montoya et al., 2017). We can mitigate these concerns by employing a first-difference model capturing the effects only when songs just entered or were removed from the charts. During the research period, most playlists were offered by the platform and updated daily except the hourly top 100 charts. In this regard, songs were very unlikely to be newly added to the top 100 charts and daily updated playlists at the same time. Thus, we estimate the following specification:

$$\begin{aligned}\Delta\operatorname{arcsinh}(y_{it}) = {} & \beta_2 \cdot \Delta\text{Top 100 Listed}_{it} + \gamma_{21} \cdot \Delta[\text{Rank}_{it} - 100] + \gamma_{22} \cdot \Delta[\text{Rank}_{it} - 100] \cdot \Delta\text{Top 100 Listed}_{it} \\ & + \alpha(i, t : \theta) + \varepsilon_{it},\end{aligned} \tag{2}$$

where $\Delta(x_{it})$ is an operator denoting $x_{it} - x_{it-1}$; and other variables are identically defined as Eq. (1).

ΔTop 100 Listed$_{it}$ is zero when Top 100 Listed$_{it}$ = Top 100 Listed$_{it-1}$ regardless of whether a song was listed on the top 100 charts. In the first-difference model, song fixed effects control for unobserved heterogeneity in consumption decay paths across songs, instead of the scale of streaming consumption, alleviating concerns about the endogeneity of treatment with respect to the shape of a song's demand path (Hendricks and Sorensen, 2009; Kumar et al., 2014).

[4] We also estimate alternative models by excluding a set of fixed effects and control variables and changing the error structure. We find that the estimates are qualitatively and quantitatively similar to those of the current model. The estimated results are available upon request.

Table 3
Regression discontinuity estimates: level and first-difference form.

Rank bandwidth:	[96, 105]			
Prior count bandwidth:	[−5%, +5%]			
Specification:	Level Form		First Difference	
Dependent variable:	New Streamers	Streaming Counts	New Streamers	Streaming Counts
Top 100 listed (t)	0.208***	0.0499**	0.125***	0.0253**
	(0.0394)	(0.0162)	(0.0173)	(0.00875)
[Rank - 100] (t)	−0.00978	−0.00136	0.000765	0.000117
	(0.0126)	(0.00496)	(0.00130)	(0.000549)
[Rank - 100] x Top 100 Listed (t)	0.0202	−0.00116	−0.00483*	−0.000506
	(0.0136)	(0.00481)	(0.00231)	(0.000774)
arcsinh(streaming counts) (t - 1)	0.255**	0.708***	0.526***	0.128*
	(0.0839)	(0.0459)	(0.105)	(0.0571)
arcsinh(new streamers) (t - 1)	0.355***	−0.0235	−0.463***	−0.0617**
	(0.0668)	(0.0203)	(0.0538)	(0.0196)
Hour-of-day fixed effects	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes
Song fixed effects	Yes	Yes	Yes	Yes
No. of songs	126	126	123	123
Observations	1401	1401	1390	1390
Within R-squared	0.806	0.929	0.493	0.677

Notes. Standard errors in parentheses are robust and clustered by song. $^{+}p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

4. Results

4.1. Main results

We report the estimates of Eqs. (1) and 2 based on the rank range [96, 105] and the streaming count range [−5%, +5%] in Table 3. Across all model forms and dependent variables, we find positive relationships between being listed on the top 100 charts and streaming demand; these relationships are statistically significant at the 0.1% (1%) level for the number of new streamers (total streaming counts).

A notable observation is that the coefficients of streaming counts are much smaller than those of new streamers. Specifically, the estimated effect in the first column (+23.0%) is 3.5 times larger than the estimate in the second column (+5.1%). Likewise, the third column (+13.3%) shows the effect 4.2 times larger than that in the last column (+2.6%). This could be explained by repeat consumption which is relatively less likely to be affected by chart exposure. These findings support the necessity to separate new listeners from previous listeners to measure the causal effect of the ranking charts on music discovery.

In addition, we observe that the first-difference form (i.e., Eq. (2)) yields substantially smaller estimates than the level form (i.e., Eq. (1)). For instance, compared to the estimated effect in the first column (+23.0%), the third column demonstrates 42.2% or 9.7% points smaller result (+13.3%). This might suggest that the first-difference model partially rules out confounding factors—such as being listed on other playlists and repeated exposure to a specific song—by capturing only the moment of either entering or dropping out of the top 100 lists. Therefore, we focus on the first-difference model in our following analyses.

Based on the results in the third and fourth columns, the average gain in new streamers (streaming counts) due to the top 100 lists is about 12.4 people (34.8 streams) per hour.[5] In other words, 64.5% of the increase in total streams is attributable to repeat consumption. We also estimate our main specifications for repeat consumption in Table A2 and find that the coefficients are positive and significant but smaller than the estimates for total streaming counts. This finding suggests that music listeners who had already consumed and been aware of a song were also affected by the song's chart entrance.

4.2. Robustness checks

We assess the sensitivity of our results using various combinations of rank range and streaming count bandwidths: [91, 110] and [95, 105] for rank positions, and [−10%, +10%], [−5%, +5%], [−2%, +2%] and [−1%, +1%] for prior streaming counts. The RDD estimates for these combinations are presented in Table 4. We find that the estimates of new streamers are statistically significant and quantitatively similar across all sampling criteria; the estimated effects range from 10.9% and 13.3%. The estimates for total streaming counts are statistically and economically more significant for the relatively narrow

[5] We use the average hourly performance of 101st ranked songs—92.9 people and 1339.1 streams for new streamers and streaming counts, respectively—as the baseline to calculate the top 100 list's effects.

Table 4
Alternative bandwidths and estimated effects: first-difference form.

Dependent variable:	New streamers		Streaming counts	
Prior count bandwidth	Coefficient (Std. Err.)	Observations (No. of Songs)	Coefficient (Std. Err.)	Observations (No. of Songs)
Rank bandwidth: [91, 110]				
[−10%, +10%]	0.104***	3476	0.0145*	3476
	(0.0147)	179	(0.00581)	179
[−5%, +5%]	0.121***	2082	0.0162*	2082
	(0.0165)	154	(0.00821)	154
[−2%, +2%]	0.124***	1223	0.0180+	1223
	(0.0209)	130	(0.00978)	130
[−1%, +1%]	0.125***	921	0.0174	921
	(0.0263)	119	(0.0126)	119
Rank bandwidth: [96, 105]				
[−10%, +10%]	0.118***	2070	0.0236***	2070
	(0.0153)	146	(0.00611)	146
[−5%, +5%]	0.125***	1390	0.0253**	1390
	(0.0173)	123	(0.00875)	123
[−2%, +2%]	0.116***	949	0.0258*	949
	(0.0217)	109	(0.0106)	109
[−1%, +1%]	0.106***	786	0.0212+	786
	(0.0269)	107	(0.0126)	107

Notes. Standard errors in parentheses are robust and clustered by song. $^{+}p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$. Covariates and fixed effects used in Table 3 are included but not reported for brevity.

Table 5
Placebo cutoffs and estimated effects: first-difference form.

Prior count bandwidth: Dependent variable:	[−5%, +5%] New streamers		Streaming counts	
Cutoff [rank bandwidth]	Coefficient (Std. Err.)	Observations (No. of Songs)	Coefficient (Std. Err.)	Observations (No. of Songs)
120	0.0116	1206	−0.00616	1206
[115, 125]	(0.0212)	162	(0.00835)	162
110	0.00884	1261	−0.00196	1261
[105, 115]	(0.0252)	159	(0.00711)	159
90	−0.00959	1651	−0.00196	1651
[85, 95]	(0.0132)	101	(0.00711)	101
80	0.0169	1721	0.00754+	1721
[75, 85]	(0.0112)	99	(0.00434)	99

Notes. Standard errors in parentheses are robust and clustered by song. $^{+}p < 0.10$, $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Covariates and fixed effects used in Table 3 are included but not reported for brevity.

bandwidth of rank positions. Given the narrower bandwidth, the estimated effects are about 2.1 - 2.6%. The results indicate that our findings remain unchanged when we adopt the alternative margins of the RDD.

We further examine whether the inverse hyperbolic sine function and selection of control variables spuriously drive our findings. Table A3 reports the estimated effects of an alternative specification replacing the transformed variables with non-transformed or counting variables. We find that the estimated effects are still positive and significant. Notably, they are quantitatively similar to the converted numbers obtained from the estimates in Table 3. We also check if a song's re-entering or re-exiting from the ranking charts affects our findings. To address this possibility, we re-estimate our specifications by excluding observations where songs entered or exited more than once and observe similar results (see Table A4).

To rule out the possibility that correlations between consumption and ranks spuriously drive our results, we conduct a falsification test that utilizes placebo cutoffs which should not present sharp discontinuity (Allcott, 2011; Anderson and Magruder, 2012). By doing so, we can assess whether the estimated discontinuity at the pre-determined threshold is spurious. Given that the digital music platform did not provide discontinuous interfaces concerning the top 120th, 110th, 90th, and 80th positions, we expect these alternative thresholds not to pick up any effect in the falsification regressions.

Table 5 presents the results of our falsification test. We find that the placebo thresholds do not show statistically significant relationships with the number of new streamers and streaming counts. This provides evidence in favor of a non-coincidental effect picked by the sharp cutoff at the top 100th position in the main analysis.

Table 6
Entering or exiting from top 100 charts and new streamers: first-difference form.

Rank bandwidth:	[96, 105]			
Prior count bandwidth:	[−5%, +5%]			
Chart dynamics:	Newly added to the charts at t		Newly removed from the charts at t	
Dependent variable:	New streamers	Streaming counts	New streamers	Streaming counts
(t - 1) - (t) pairs where Top 100 Listed (t - 1) $\neq$ Top 100 Listed (t)				
Top 100 listed (t)	0.0646*	−0.00641	0.0944***	0.0100
	(0.0266)	(0.0139)	(0.0241)	(0.0114)
No. of songs	478	478	442	442
Observations	92	92	95	95
(t - 1) - (t) pairs where Top 100 Listed (t - 2) = Top 100 Listed (t - 1) $\neq$ Top 100 Listed (t)				
Top 100 listed (t)	0.0875	−0.0400	0.144**	0.0257
	(0.0765)	(0.0333)	(0.0461)	(0.0195)
No. of songs	265	265	249	249
Observations	77	77	68	68

Notes. Standard errors in parentheses are robust and clustered by song. $^{+}p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$. Covariates and fixed effects used in Table 3 are included but not reported for brevity.

4.3. Being added to and removed from bestseller lists

We also examine how gaining or losing a position on ranking charts affects streaming demand differently to provide insights on how in-out dynamics around the cutoff shape the total effect of being on the top 100 lists. To do so, we re-estimate the RDD estimates of our main specification by restricting our observations to the following cases. First, we consider a case where a song is newly added to the top 100 charts. In this case, we include (t - 1, t) observation pairs where Top 100 $\text{Listed}_{it-1} = 0$ and Top 100 $\text{Listed}_{it} = 1$. In the second case, we include observations where a song is newly removed from the lists; namely, we consider (t - 1, t) pairs where Top 100 $\text{Listed}_{it-1} = 1$ and Top 100 $\text{Listed}_{it} = 0$. For both cases, we expect positive coefficients of Top 100 Listed_{it} as our specification is unchanged.

Table 6 presents the estimates of Eq. (2) for samples constructed as follows. The first panel concerns all (t - 1, t) observation pairs satisfying the aforementioned conditions. In the second panel, we additionally restrict our sample to songs of which Top 100 Listed_{it} did not change consecutively; that is, one of ΔTop 100 Listed_{it} and ΔTop 100 Listed_{it-1} is zero. This restriction allows us to rule out the possibility that being listed on the charts at t - 2 affects music consumption at t - 1. From both panels, we find that the coefficients of Top 100 Listed_{it} are larger when being removed from the top-ranking charts than being added to the lists. We conjecture that repeated exposure for several hours further increases a song's saliency, and thus dropping out leads to more significant differences in music consumption.[6]

4.4. Observational learning vs. saliency effect

We have found significant effects of being listed on the top 100 selling charts even though consumers do not have to pay monetary costs for consuming new songs. There could be several drivers of this result. First, music streamers might add the entire top 100 songs to their private playlists for convenience rather than selecting particular songs among the listed ones. Second, consumers may perceive being listed on top 100 lists as a quality signal, that is, observational learning (Cai et al., 2009; Carare, 2012; Sorensen, 2007). Third, consumers might be more willing to choose the listed songs because being listed on ranking charts makes songs more salient than non-listed songs, i.e., the saliency effect (Cai et al., 2009).

Regarding the first possibility, streaming services provide a great environment for inattentive music consumption by removing marginal monetary costs and convenient interfaces. It might affect music consumption substantially, but in our data which, were recorded at hourly intervals, such behaviors are unlikely to be captured because a consumer who begins to play the top 100 songs from the first ranked song will listen to the 100th ranked song after five hours (when we assume that the average duration of songs is three minutes). Therefore, the inattentive music discovery does not seem to drive our results.

If the second possibility—i.e., observational learning—were dominant, we would not observe a significant behavioral change among users who had already listened to a song. However, our results support the opposite; being listed on the top

[6] Previous studies suggested that repeated exposure to stimuli increases content saliency (Eden et al., 2014), and that relative saliency induces more positive and emotional responses from individuals (Mrkva and Van Boven, 2020). This conjecture is consistent with our finding that the coefficient in the second panel (being listed on the charts at least twice—i.e., t - 2 and t - 1) was more significant than the estimate in the first panel (being listed on the charts at t - 1). Furthermore, as the top 100 charts are updated hourly, chart entrance and exit are unlikely to coincide with the timing of being added to or removed from daily updated playlists.

Table 7
Normative cutoffs and estimated effects: first-difference form.

Prior count bandwidth:	[−5%, +5%]			
Dependent variable:	New streamers		Streaming counts	
Cutoff [rank bandwidth]	Coefficient (Std. Err.)	Observations (No. of Songs)	Coefficient (Std. Err.)	Observations (No. of Songs)
50	0.00214	1845	−0.00240	1845
[45, 55]	(0.0111)	109	(0.00517)	109
25	−0.0125	1566	−0.00909	1566
[20, 30]	(0.0129)	73	(0.00765)	73
20	−0.0121	1572	−0.0231***	1572
[15, 25]	(0.0109)	67	(0.00599)	67
10	0.0267	1376	0.0243*	1376
[5, 15]	(0.0169)	41	(0.0101)	41

Notes. Standard errors in parentheses are robust and clustered by song. $+p < 0.10$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Covariates and fixed effects used in Table 3 are included but not reported for brevity.

100 lists increased repeated consumption of a song as well, suggesting that bestseller lists affect streaming users' choices even when the users are aware of product quality.

Moreover, we employ the following approaches to disentangle observational learning and the saliency effect. First, we assess whether consumers responded to the top 100 threshold because they perceived this number as a normative threshold. The extant literature suggested that people may infer a higher quality level from rank positions beyond a certain level (e.g., Isaac and Schindler, 2014). If so, music discovery would demonstrate similar discontinuities around other normative numbers, such as 10, 20, 25, and 50. To examine this possibility, we re-estimate our model by replacing the top 100 cutoff with these alternative thresholds. Table 7 reports the estimated results. We find that being ranked in higher or equal positions of the suggested numbers does not significantly affect the number of new streamers. These results suggest that streaming users consider only the availability or saliency of options instead of the perceived quality threshold.

Second, we estimate the interaction effects between Top 100 Listed and songs' observable characteristics associated with information discovery: Major Label, Foreign, Years Since Debut, Days Since Release, and Album Title (see Table 2 for variable description). Albums released by major labels are relatively more promoted than those from indie labels, so they tend to benefit less from observational learning (Dewan and Ramaprasad, 2012). Similarly, artists benefit less from information discovery in their home market than in other markets (Hendricks and Sorensen, 2009). Hence, we will observe that the impact of bestseller lists will decrease (increase) for major labels (foreign artists) if observational learning mainly drives this phenomenon. We also consider how much time had passed since a song's release or its artists' debut. As relatively old artists and songs are likely to be exposed to consumers more widely, we will obtain negative and significant interaction terms for these variables if observational learning is dominant. Lastly, we consider if bestseller lists affect album title songs differently. Title tracks are more likely to be advertised and exposed to consumers than non-title tracks. Furthermore, consumers may perceive non-title songs to be narrow-appeal and thus infer higher quality for them compared to album titles when both are listed on the top-selling lists (Tucker and Zhang, 2011). Hence, we will observe a negative interaction term between Top 100 Listed and Album Title if our findings are mainly attributable to observational learning.

The estimated interaction terms are presented in Table 8. Interestingly, we find no statistically significant interaction term. Considering that the literature on observational learning showed that minor producers, artists outside their home market, and seemingly narrow-appeal products benefitted more from enhanced information (Hendricks and Sorensen, 2009; Kumar et al., 2014; Tucker and Zhang, 2011), the results suggest that observational learning does not seem to significantly drive our results. For these reasons, we conjecture that the saliency effect seems to lead to the increase in the discovery of songs.

5. Conclusions

Our study aims to quantify the impact of bestseller lists on product discovery in a subscription-based platform. Using fine-grained data collected from one of the largest digital music platforms in South Korea, we find that entrance to the top 100 charts immediately increases the discovery of songs by 11–13%. Moreover, the observed relationships persist when we alter the margin of the RDD or the functional form of streaming demands, and importantly, we do not observe similar patterns from placebo cutoffs.

Importantly, we reveal that the saliency effect contributes to the effects of the ranking charts more significantly than observational learning. Being listed on the top 100 charts increases repeat consumption as well. Other normative positions within the charts do not create a significant discontinuity. Furthermore, factors associated with awareness of artists and songs do not lead to a significant contingency of this effect.

We suggest that zero marginal costs and large product catalogs in streaming services may drive these results. Specifically, music streamers might be insensitive to quality signals due to the lack of marginal monetary costs in subscription-based services. It is consistent with the previous findings that free apps are less sensitive to product rankings than paid apps

Table 8
Estimates of heterogeneous treatment effects: first-difference form.

Rank bandwidth:	[96, 105]					
Prior count bandwidth:	[−5%, +5%]					
Dependent variable:	New streamers					
Top 100 listed	0.138***	0.123***	0.132***	0.113**	0.126***	0.148*
	(0.0240)	(0.0181)	(0.0273)	(0.0396)	(0.0203)	(0.0574)
x Major label	−0.0314					−0.0316
	(0.0325)					(0.0361)
x Foreign artist		0.0182				0.0251
		(0.0305)				(0.0475)
x Years since debut			−0.00127			−0.00103
			(0.00344)			(0.00348)
x arcsinh(days since release)				0.00220		−0.000655
				(0.00807)		(0.00913)
x Album title					−0.00370	−0.00963
					(0.0333)	(0.0357)
[Rank - 100]	0.000783	0.000768	0.000759	0.000770	0.000769	0.000788
	(0.00131)	(0.00131)	(0.00130)	(0.00130)	(0.00130)	(0.00131)
[Rank - 100] x top 100 listed	−0.00485*	−0.00483*	−0.00486*	−0.00484*	−0.00482*	−0.00485*
	(0.00230)	(0.00232)	(0.00231)	(0.00231)	(0.00231)	(0.00230)
arcsinh(streaming counts) (t - 1)	0.527***	0.524***	0.526***	0.525***	0.526***	0.524***
	(0.105)	(0.106)	(0.105)	(0.106)	(0.105)	(0.106)
arcsinh(new streamers) (t - 1)	−0.463***	−0.463***	−0.463***	−0.463***	−0.463***	−0.463***
	(0.0541)	(0.0538)	(0.0538)	(0.0540)	(0.0539)	(0.0543)
Hour-of-day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Song fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of unique songs	123	123	122	123	123	122
Observations	1390	1390	1389	1390	1390	1389
Within R-squared	0.493	0.493	0.493	0.493	0.493	0.493

Notes. Standard errors in parentheses are robust and clustered by song. $^{+}p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$. Due to unobserved values of 'Years Since Debut', one observation is dropped in the third and sixth columns.

in the mobile application market (Carare, 2012). Second, music listeners might suffer from choice overload such that large choice sets can reduce their satisfaction with and the quality of their decisions (Besedeš et al., 2015). Digital music platforms provide millions of tracks, whereas users know a tiny part of their options. Furthermore, streaming users need to create their own playlist that consists of multiple songs. For these reasons, streaming users become more reliant on salient options such as songs listed on the ranking charts (Helmers et al., 2019).

This limited contingency suggests that popular artists can similarly benefit from being added to the top-selling lists in streaming platforms, unlike traditional markets. Therefore, they can continue to boost their music revenues from promotion activities even after their music is widely known. In other words, lesser-known artists and indie labels can barely narrow the popularity gap from major artists by entering the ranking charts. These results suggest that bestseller lists may exacerbate the concentration of product consumption more severely under all-you-can-eat pricing than à la carte.

It is also worth noting that the top 10 cutoff showed a 2.5% increase in total streaming counts, while it did not present a significant rise in new streamers (see Table 7). Given that the platform did not provide an exclusive playlist or interface for the top 10 songs, we may conclude that consumers perceive a song ranked on the top 10th or higher positions differently from other songs, particularly when they already listened to the song before. In this regard, the platform may consider promoting top 10 rankers to previous listeners by sending push notifications.

This study has several limitations. First, our identification strategy does not entirely rule out the simultaneity problem. Although our fine-grained data and first-difference model allow us to focus on the exact moment of entering and exiting from the charts, there still remains a possibility that a song may be added to other playlists—which were updated daily in the focal platform—in the same hour. Further, we do not take account of external factors, such as advertising and critic reviews, which may also affect the demand for songs. Future research may benefit from tracking other popular playlists, promotion activities, and media coverage.

Second, we acknowledge that future works should further validate the suggested mechanism. As noted above, it takes approximately five hours for inattentive consumers to reach the top 100th ranked song from the highest track. However, such consumers might reach the lowest ranker earlier if they use a shuffle mode for music listening. Although it is unlikely that the vast majority of consumers add the entire tracks on the charts and shuffle their playlists, the proportion of such inattentive consumers will be highly informative for streaming platforms.

Acknowledgment

The authors are grateful for the constructive comments and suggestions made by the co-editor, Scott Adams, and two anonymous referees. The authors also thank Minki Kim, Jiyong Park, Euro Bae, Jaehyeong Park, Robin Hwang, and participants at the Workshop on Information Systems and Economics (WISE) and the Symposium on Statistical Challenges in Electronic Commerce Research (SCECR) for many constructive comments. All remaining errors are ours.

Appendix

To explore the relative importance of bestseller lists in streaming platforms, we rely on playlist data from Spotify, the largest music streaming platform in the world, rather than our focal platform, which did not provide playlist following features during the research period, the number of playlist followers is not applicable in our context. Hence, we instead collect the lists of most popular playlists and their corresponding follower counts from Soundcharts.

Fig. A1 depicts the top 10 most popular playlists and their number of followers on October 22, 2020. Two bestseller lists—Today's Top Hits and Global Top 50—are the first and second most subscribed playlists. Notably, the follower distribution is highly concentrated on relatively popular playlists. We observe that the sum of their followers (42.6 million) occupies 35.8% of the top 10 playlists' follower counts (119.1 million). It is also worth mentioning that none of the top 10 lists are offered by independent providers, indicating that consumers heavily rely on platform-provided information to discover new music. These results suggest that bestseller lists are the most popular form of music discovery in music streaming platforms.

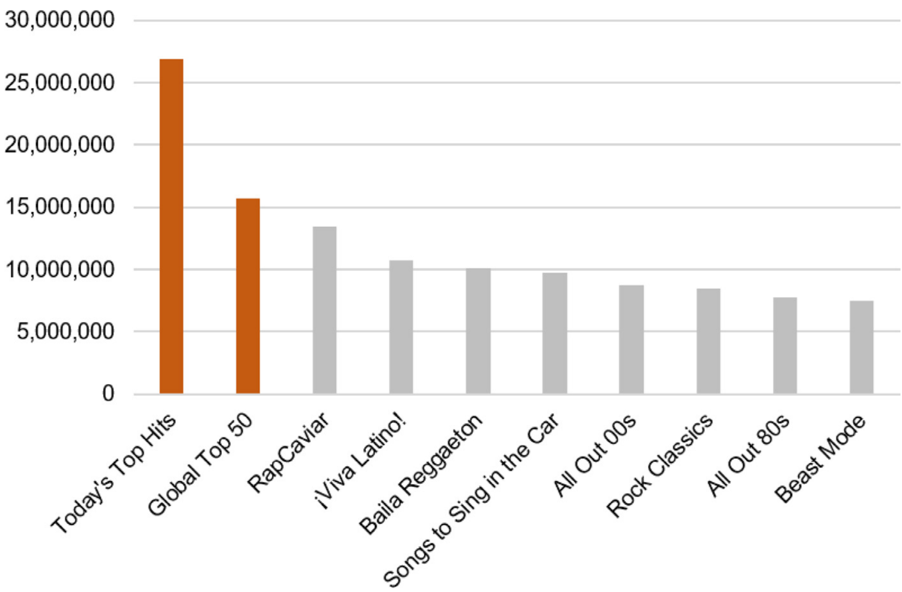

Fig. A1. Number of followers among Spotify's Top 10 most popular playlists
Note. This graph is based on the top 10 playlists on Spotify on October 22, 2020, collected by soundcharts.

Table A1
Correlation matrix.

Variables		1.	2.	3.	4.	5.	6.	7.
1.	Rank	–						
2.	Streaming counts	−0.5674	–					
3.	New streamers	−0.1898	0.6155	–				
4.	Major label	−0.1322	0.0539	−0.0301	–			
5.	Foreign	0.2087	−0.1198	−0.0154	−0.2166	–		
6.	Years since debut	0.1001	−0.1087	−0.0364	−0.0686	0.3301	–	
7.	Days since release	0.1532	−0.1548	−0.0788	−0.0913	0.1305	0.1915	–
8.	Album title	−0.0272	−0.0171	−0.0114	−0.0979	0.0899	−0.0407	−0.0148

Note. All correlations are statistically significant at the 0.1 percent level.

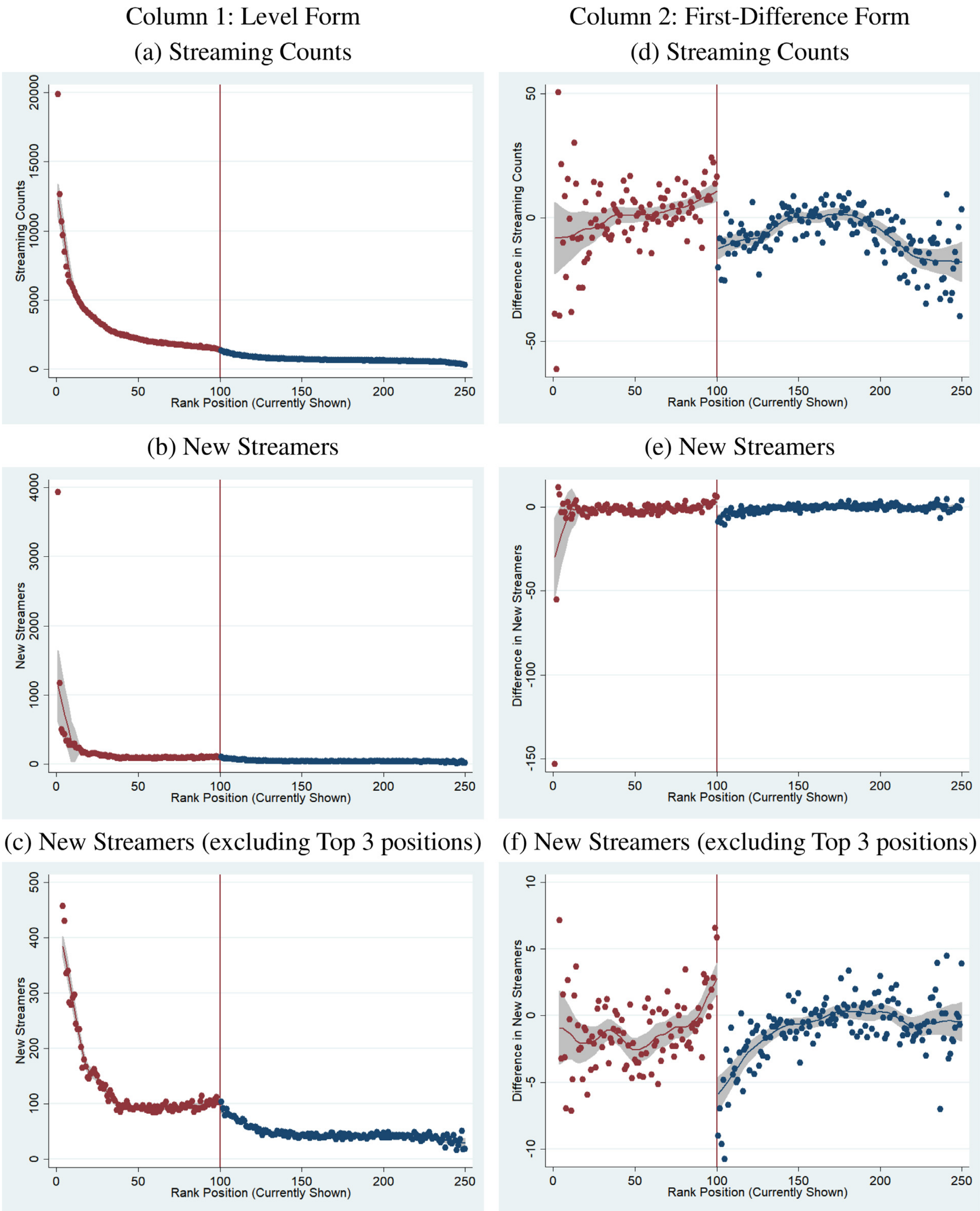

Fig. A2. Average streaming counts and new streamers throughout Top 250 positions.

Table A2
RDD estimates for repeat consumption.

Rank bandwidth:	[96, 105]	
Prior count bandwidth:	[−5%, +5%]	
Dependent variable:	Repeat consumption	
Specification:	Level form	First Difference
Top 100 listed (t)	0.0418*	0.0201*
	(0.0162)	(0.00889)
[Rank - 100] (t)	−8.77e-06	0.000164
	(0.00490)	(0.000541)
[Rank - 100] x top 100 listed (t)	−0.00367	−0.000375
	(0.00478)	(0.000759)
arcsinh(streaming counts) (t - 1)	0.748***	0.116*
	(0.0471)	(0.0571)
arcsinh(new streamers) (t - 1)	−0.0533**	−0.0459*
	(0.0199)	(0.0190)
Hour-of-day fixed effects	Yes	Yes
Day fixed effects	Yes	Yes
Song fixed effects	Yes	Yes
No. of songs	126	123
Observations	1401	1390
Within R-squared	0.927	0.674

Notes. Standard errors in parentheses are robust and clustered by song.
$^{+}p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table A3
RDD estimates with counting variables.

Rank bandwidth:	[96, 105]			
Prior count bandwidth:	[−5%, +5%]			
Specification:	Level form		First difference	
Dependent variable:	New streamers	Streaming counts	New streamers	Streaming counts*
Top 100 listed (t)	19.39***	73.61**	8.749***	37.17**
	(3.844)	(21.99)	(2.070)	(12.22)
[Rank - 100] (t)	1.321	2.333	−0.300+	0.0199
	(0.987)	(7.014)	(0.161)	(0.881)
[Rank - 100] x top 100 listed (t)	−0.102	−4.662	−0.127	−0.257
	(1.192)	(7.119)	(0.144)	(0.926)
Streaming counts (t - 1)	−0.00755	0.733***	0.0118	0.0755
	(0.00571)	(0.0461)	(0.0134)	(0.0564)
New streamers (t - 1)	0.766***	−0.0805	−0.190	−0.283
	(0.0519)	(0.0948)	(0.179)	(0.223)
Hour-of-day fixed effects	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes
Song fixed effects	Yes	Yes	Yes	Yes
No. of songs	126	126	126	126
Observations	1401	1401	1401	1401
Within R-squared	0.732	0.889	0.225	0.604

Notes. Robust standard errors are in parentheses. $^{+}p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

Table A4
RDD estimates excluding re-entering and re-exiting observations.

Rank bandwidth:	[91, 110]			
Prior count bandwidth:	Not considered			
Specification:	Level form		First difference	
Dependent variable:	New streamers	Streaming counts	New streamers	Streaming counts
Top 100 listed (t)	0.183***	−0.00939	0.123*	−0.0248
	(0.0518)	(0.0275)	(0.0575)	(0.0204)
[Rank - 100] (t)	−0.00715	0.00129	−0.00131	−0.00123*
	(0.00699)	(0.00293)	(0.00107)	(0.000486)
[Rank - 100] x top 100 listed (t)	0.00542	−0.00236	0.00130	0.00271*
	(0.00866)	(0.00384)	(0.00243)	(0.00114)
arcsinh(streaming counts) (t - 1)	0.353**	0.648***	0.473***	0.130*
	(0.114)	(0.0527)	(0.134)	(0.0659)
arcsinh(new streamers) (t - 1)	0.138+	−0.0717**	−0.598***	−0.0857***
	(0.0817)	(0.0247)	(0.0765)	(0.0209)
Hour-of-day fixed effects	Yes	Yes	Yes	Yes
Day fixed effects	Yes	Yes	Yes	Yes
Song fixed effects	Yes	Yes	Yes	Yes
No. of songs	226	226	207	207
Observations	1200	1200	1140	1140
Within R-squared	0.726	0.870	0.554	0.605

Notes. Standard errors in parentheses are robust and clustered by song. $^{+}p < 0.10$.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$. We utilize a wider rank bandwidth and do not use a prior count bandwidth to make the sample size comparable to Table 3.

References

Aguiar, L., Waldfogel, J., 2018a. As streaming reaches flood stage, does it stimulate or depress music sales? Int. J. Ind. Organ. 57, 278–307.
Aguiar, L., Waldfogel, J., 2018b. Platforms, promotion, and product discovery: evidence from spotify playlists. NBER Working Paper 24713.
Allcott, H., 2011. Social norms and energy conservation. J. Public Econ. 95, 1082–1095.
Anderson, M., Magruder, J., 2012. Learning from the crowd: regression discontinuity estimates of the effects of an online review database. Econ. J. 122, 957–989.
Bellemare, M.F., Wichman, C.J., 2020. Elasticities and the inverse hyperbolic sine transformation. Oxf. Bull. Econ. Stat. 82, 50–61.
Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? Q. J. Econ. 119, 249–275.
Besedeš, T., Deck, C., Sarangi, S., Shor, M., 2015. Reducing choice overload without reducing choices. Rev. Econ. Stat. 97, 793–802.
Burbidge, J.B., Magee, L., Robb, A.L., 1988. Alternative transformations to handle extreme values of the dependent variable. J. Am. Stat. Assoc. 83, 123–127.
Cai, H., Chen, Y., Fang, H., 2009. Observational learning: evidence from a randomized natural field experiment. Am. Econ. Rev. 99, 864–882.
Cao, J., Liang, H., Zhan, X., 2019. Peer effects of corporate social responsibility. Manage. Sci. 65, 5487–5503.
Carare, O., 2012. The impact of bestseller rank on demand: evidence from the app market. Int. Econ. Rev. 53, 717–742.
Chen, J., Wang, C., Wang, J., Philip, S.Y., 2016. Recommendation for repeat consumption from user implicit feedback. IEEE Trans. Knowl. Data Eng. 28, 3083–3097.
Datta, H., Knox, G., Bronnenberg, B.J., 2017. Changing their tune: how consumers' adoption of online streaming affects music consumption and discovery. Mark. Sci. 37, 5–21.
Dewan, S., Ramaprasad, J., 2012. Music blogging, online sampling, and the long tail. Inf. Syst. Res. 23, 1056–1067.
Eden, A., Tamborini, R., Grizzard, M., Lewis, R., Weber, R., Prabhu, S., 2014. Repeated exposure to narrative entertainment and the salience of moral intuitions. J. Commun. 64, 501–520.
Godinho de Matos, M., Ferreira, P., Smith, M.D., Telang, R., 2016. Culling the herd: using real-world randomized experiments to measure social bias with known costly goods. Manage. Sci. 62, 2457–2764.
Hahn, J., Todd, P., Van der Klaauw, W., 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69, 201–209.
Helmers, C., Krishnan, P., Patnam, M., 2019. Attention and saliency on the internet: evidence from an online recommendation system. J. Econ. Behav. Organ. 161, 216–242.
Hendricks, K., Sorensen, A., 2009. Information and the skewness of music sales. J. Polit. Econ. 117, 324–369.
IFPI, 2016. Global Music Report 2016: Music Consumption Exploding Worldwide. International Federation of the Phonographic Industry.
Imbens, G.W., Lemieux, T., 2008. Regression discontinuity designs: a guide to practice. J. Econom. 142, 615–635.
Isaac, M.S., Schindler, R.M., 2014. The top-ten effect: consumers' subjective categorization of ranked lists. J. Consum. Res. 40, 1181–1202.
Kumar, A., Smith, M.D., Telang, R., 2014. Information discovery and the long tail of motion picture content. MIS Q. 38, 1057–1078.
Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. J. Econ. Lit. 48, 281–355.
Luca, M., 2011. Reviews, reputation, and revenue: the case of Yelp.com. Working Paper No. 12-016. Harvard Business School.
MacKinnon, J.G., Magee, L., 1990. Transforming the dependent variable in regression models. Int. Econ. Rev. 31, 315–339.
Montoya, R.M., Horton, R.S., Vevea, J.L., Citkowicz, M., Lauber, E.A., 2017. A re-examination of the mere exposure effect: the influence of repeated exposure on recognition, familiarity, and liking. Psychol. Bull. 143, 459–498.
Moulton, B.R., 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. Rev. Econ. Stat. 72, 334–338.
Mrkva, K., Van Boven, L., 2020. Salience theory of mere exposure: relative exposure increases liking, extremity, and emotional intensity. J. Pers. Soc. Psychol. 118, 1118–1145.
Salganik, M.J., Dodds, P.S., Watts, D.J., 2006. Experimental study of inequality and unpredictability in an artificial cultural market. Science 311, 854–856.
Sorensen, A.T., 2007. Bestseller lists and product variety. J. Ind. Econ. 55, 715–738.

Sorensen, A.T., 2017. Bestseller lists and the economics of product discovery. Annu. Rev. Econom. 9, 87–101.
Thomes, T.P., 2013. An economic analysis of online streaming music services. Inf. Econ. Policy 25, 81–91.
Tucker, C., Zhang, J., 2011. How does popularity information affect choices? Manage. Sci. 57, 828–842.
Ursu, R.M., 2018. The power of rankings: quantifying the effect of rankings on online consumer search and purchase decisions. Mark. Sci. 37, 530–552.
Waldfogel, J., 2016. Cinematic explosion: new products, unpredictabilty and realized quality in the digital era. J. Ind. Econ. 64, 755–772.