Music Reviews and Music Demand: Evidence from Pitchfork and Last.fm

Luís Cabral
New York University

April Franco
University of Toronto

Matthew Mitchell
University of Toronto

April 2021

Abstract. Every business day since 2003, the website Pitchfork reviews five albums. Some also receive a badge as "Best New Music" (BNM). We examine the effect on music demand of this form of expert advice, especially the BNM badge. We measure demand by the number of listeners and number of players on the music site last.fm. Based on a Fuzzy Regression Discontinuity design, we estimate that BNM certification increases the number of listeners by three-fold and the number of plays by six-fold. A series of robustness checks suggest that these are causal effects.

Cabral: Stern School of Business, New York University; luis.cabral@nyu.edu. Franco: Rotman School of Management, University of Toronto; april.franco@rotman.utoronto.ca. Mitchell: Rotman School of Management, University of Toronto; matthew.mitchell@rotman.utoronto.ca
We are grateful to seminar audiences in Toronto and at the MIRA Conference for helpful comments. The usual disclaimer applies.

1. Introduction

With drastically lower production costs, the supply of media products has increased at a high rate — perhaps by orders of magnitude — since the beginning of the century (Waldfogel, 2015). The abundance of content available to consumers brings to life Simon's famous quip: a wealth of information creates a poverty of attention (Simon, 1971). It used to be difficult to get published; nowadays, it is difficult to be read, seen, listened to. In this context, the problem of discovery becomes particularly important. Consider the case of music, the industry we focus on in this paper. Listeners are offered a very large number of albums to choose from, most of which they are not even aware. How do the new hits get to each listener's ears? How are new artists discovered?

In the pre-digital era, radio play and label promotion fulfilled a central role. In addition to lower costs, the digital age brought two important new features: First, there are now multiple new platforms and web sites with abundant and free information, including in particular music reviews. To be sure, critics and reviews have always existed and have always played a role. However, considering the ease of access to webpages full of ratings, rankings and reviews, one may conjecture that reviews are particularly important. As Waldfogel (2015) points out,

> Along with many other effects of digitization, the Internet has led to an explosion of outlets providing critical assessment of new music. Since 1995 the number of outlets reviewing new music — and the number of reviews produced per year — has doubled (p 425).

A second important feature of the the digital world is the emergence of recommendation systems, algorithms which predict content a user is likely to enjoy based on past choices by similar users. Most current music listeners are familiar with Spotify's lists, but there are several other algorithmic-based music recommendations systems, some predating Spotify.

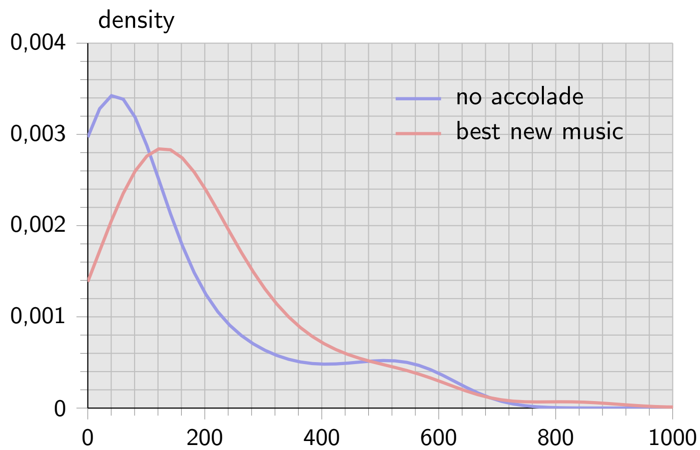
In this paper, we look at the role played by expert advice and by recommender systems. Specifically, we estimate the combined effect of influencer Pitchfork and Last.fm's algorithm. Pitchfork is an American online magazine launched in 1995, whereas Last.fm is a music website founded in 2002. In Section 2 we describe these in greater detail. In essence, Pitchfork is a source of critic-based recommendations, whereas Last.fm is a source of algorithm-based recommendations.

Since 2003, each day Pitchfork's website reviews five new albums. Some receive a classification "Best New Music" (BNM). Our goal is to determine whether this seal of quality has any effect on the number of listeners. Determining a causal effect of this sort is not an easy task. The problem is well known: how can we distinguish correlation from causality? A simple correlation analysis shows quite clearly that BNM titles are played more frequently. However, this largely follows from the simple fact that Pitchfork's critics and music listeners have similar tastes. In order to tease out correlation from causality, we follow a fuzzy regression discontinuity approach. Essentially, we estimate both the likelihood that Pitchfork assigns the BNM label as a function of the title's score; and then estimate the impact, in terms of number of listeners and number of plays, of receiving the BNM seal. Our discontinuity is motivated by the folklore that a score above 8.0 results in a BNM label, but in fact the significant discontinuity appears just above 8.0.

We estimate large effects arising from the BNM label. Consider for example a title receiving a score of 8.2, the score level above which titles are substantially more likely to

Figure 1

Pitchfork's Best New Music and last.fm listeners. Gaussian kernel density, Scott's (1992) rule.



be labeled as BNM. We estimate that, if the title does not receive the BNM seal, then it attracts an average 133,981 listeners, whereas with BNM that number increases to 453,722, a three-fold increase. For number of plays, we obtain even stronger average values: 1,786,439 for titles with no accolade, 11,306,639 for BNM albums, a six-fold increase.

Pitchfork rates all albums on a 1-10 scale. Moreover, as we will show later, Pitchfork's scores are highly correlated with those emanating from other critics. Further, there is no discontinuous impact of BNM on other critics' scores: that is, the effect of BNM is therefore over and above the information contained in those scores. The staggering effect of BNM begs the question of whether we are measuring real effects or rather a statistical anomaly. Much of the paper is devoted to discussing methodological issues as well as complementary empirical evidence to convince ourselves and the reader that BNM does indeed have a drastic effect on a music album's popularity. In short, we believe that the interaction of the importance of Pitchfork, and its ability to drive users receiving algorithmic suggestions, is a possible channel.

We do not have detailed data regarding the recommendations provided by last.fm. However, we observe a disparity between the scale of Pitchfork and the size of the effect of BNM. To put things in perspective, as of 2006 Pitchfork attracted about 240,000 unique visitors each day (Itzkoff, 2006). This is of the same order of magnitude as the 278,715 average increase in listeners (cf Table 2). However, we expect the number of Pitchfork visitors to be a weak upper bound of the listeners attracted to the recommended album. Moreover, our estimate of the number of listeners is a loose lower bound, since we only consider users affiliated with last.fm.

Figure 1 plots the density of number of listeners for albums with a Pitchfork score from 8.2–8.4, the score levels above which titles are most likely to be labeled as BNM. The blue line corresponds to albums not receiving any accolade, whereas the red line corresponds to BNM. As the figure shows, the distributions are highly skewed, though the distribution of BNM albums is considerably less skewed. This suggests that the increase in the density distribution's right tail can be observed at values as high as 600 and 800 thousand listeners, a value that is considerably higher than the number of Pitchfork visitors.

In sum, we estimate that the BNM label has a large and statistically significant effect on the number of listeners. This effect is likely greater than the number of listeners who observe the BNM label. This in turn suggests that the indirect effect through the last.fm recommendation system may play a role. All in all, the evidence suggests that new hits result from the combined effect of influencers such as Pitchfork and recommendation systems such as last.fm. We also note that the effect is stronger on the number of plays than on the number of listeners. Whereas listeners increase by a factor of 4, the number of plays increases by a factor of 12. In other words, the combined effect of Pitchfork and Last.fm is felt on the extensive margin but especially on the intensive margin.

Finally, considering that Spotify pays from $.003 to $.005 per stream, we estimate the difference between receiving and not a BNM badge from Pitchfork (conditional on a quality score of 8.2) to be in the order of $24,000 to $40,000. This may not be a lot for major acts, but it could be a very big deal for smaller and starting musicians. Moreover, this estimate is likely to be a lower bound, for two reasons. First, during the period we consider last.fm measure a fraction of the total number of listeners and plays. Second, musicians benefit from additional revenue streams which are also likely boosted by BNM: more concerts, bigger venues and more funds going their way. We will return to this later.

■ **Related literature.** The growth of the Internet and of online consumer reviews has led to an increased interest in the effect of reviews on demand. For example, Chevalier and Mayzlin (2006) use sales rank data from Amazon and Barnes and Noble websites to show that the addition of a favorable customer review of a book raises the sales rank. More recently, Reimers and Waldfogel (2021) also examine the impact of reviews on book sales, estimating that "the aggregate effect of star ratings on consumer surplus is ... more than ten times the effect of traditional review outlets."

Closer to our industry of interest, Newberry (2016) studies a music site (Amie Street Music) where consumers can buy new songs. Potential buyers can listen to a song and are provided two pieces of information: the number of previous users who listened to the song as well as the number of listeners who purchased the song, which together provide a measure of consumer valuation similar to consumer reviews. His results suggest that this observational learning leads to an increase in total welfare by 1.35%.

Also focused on the music industry, Aguiar and Waldfogel (2018) show that the inclusion on a Spotify playlist boosts streams by nearly 20 million on average. To some extent, our paper and Aguiar and Waldfogel's (2018) complement each other: whereas we focus primarily on the effect of an expert review (Pitchfork's) as it is then amplified by last.fm's recommendation system, Aguiar and Waldfogel (2018) focus primarily on a recommender system directly (Spotify's playlists).

A number of papers have estimated the causal effect of *expert* reviews on consumer demand, the focus of our paper. Eliashberg and Shugan (1997) and Reinstein and Snyder (2005) look at movie reviews. The former find no significant relation between the share of positive reviews and box-office revenue. The latter focus on two specific reviewers and report a marginal effect of reviews on opening weekend revenues.

Hilger, Rafert, and Villas-Boas (2011) and Friberg and Grönqvist (2012) examine the effect of wine expert reviews. The former set up a field experiment whereby quality ratings are posted on the shelf next to a random selection of wines, and find that positive reviews increase sales. They find that demand decreases for low-scoring wines and increases for wines scoring average or higher, which they interpret as indicating that expert opinion labels transmit quality information as opposed to simply visibility. Friberg and Grönqvist (2012) use a difference-in-differences analysis to show that the publication of a favorable review leads to a 6 percent demand increase.

Finally, Berger, Sorensen, and Rasmussen (2010) use a difference-in-differences approach to study the effect of *New York Times* book reviews. Like Hilger, Rafert, and Villas-Boas (2011), they are interested in the dual effects of review: quality information; and awareness. They show that a negative review hurts sales of books by well-known authors but increases sales of books that had lower prior awareness (so negative publicity can increase purchase likelihood).

2. Pitchfork and Last.fm

Our analysis is based on data from two different music sites: Pitchfork and last.fm. The former provides recommendations by humans, the latter recommendations by data-based algorithms.

■ **Pitchfork.** Pitchfork is an American online magazine based in Chicago, IL. Although it covers a variety of genres, its main focus is on independent music. Launched in 1995, Pitchfork is widely considered the most popular "independent-music focused" website. As of 2006, it attracted 240,000 readers per day, 1.5 million unique visitors per month (Itzkoff, 2006).

Pitchfork is widely recognized as an influencer in the music space. A number of bands, such as Arcade Fire, Modest Mouse and Broken Social Scene, owe their success at least in part, to Pitchfork's endorsement (du Lac, 2006).

Pitchfork typically concentrates on new music. New albums, typically five per business day, are scored on a 0-10 scale, in increments of 0.1, the result of the work of different reviewers. Moreover, since 2003, the site also highlights the finest music of the moment by assigning the seal "Best New Music" to some of the album titles. Although no rule exists for relating score to BNM, the BNM seal is often colloquially associated with ratings above 8. We explore the relationship between BNM and score below.

■ **Last.fm.** Last.fm is a music website founded in the United Kingdom in 2002. One of its distinguishing features is the use of a music recommender system, called "Audioscrobbler". As a music website puts it,

> Last.fm is the ultimate tool to track your ever-changing music tastes and introduce you to the next artist you love. All you need to do is create an account, connect it to your streaming service (or download their desktop app if you listen to music via iTunes), and start listening. Your profile will keep record of everything you spin. After listening for a decent period, Last.fm will have enough data to whip up some personal recommendations. The app also recommends new artists, reintroduces songs you've previously heard, and posts upcoming concerts of artists found in your library.

Specifically, last.fm builds for each user a detailed profile of musical tastes based on what they play. This information is transferred ("scrobbled") to last.fm's database and is then used to make specific recommendations based on similar users' previous plays. The recommendations are based on prior play or on similarity, as well as on popularity with other listeners. In this particular case, the system recommends Sony Stitt's "I Can't Give You Anything But Love" based on the observation that the artist is similar to Art Pepper, Paul Desmond and Lee Konitz.

3. Data and main results

Our main goal is to estimate the impact of Pitchfork's Best New Music (BNM) seal on an album's popularity, specifically its number of plays and listeners by users with a last.fm account. From a statistical point of view, this task is marred with the common problem of distinguishing correlation from causality: a simple correlation analysis shows quite clearly that BNM songs are played more frequently. However, this largely follows from the fact that Pitchfork's critics have similar tastes to the population of music listeners.

In this section, we first present our data sources and take a first look at the data. We then propose a strategy for estimating the causal impact of the BNM seal. Ideally, we would like to consider pairs of albums with the same characteristics and compare the actual demand of those with the BNM seal to those that were not considered BNM; and take the difference in demand as our estimate of the BNM demand effect. As a means to approximate this calculation, we propose a regression discontinuity approach. For reasons that will become clear, we are unable to apply a standard regression discontinuity design. Instead, we follow a fuzzy regression discontinuity design (Trochim, 1984; Angrist and Lavy, 1999; Hahn, Todd, and der Klaauw, 2001). We conclude the section by presenting the results and a series of robustness checks.

■ **Data.** We downloaded from Pitchfork's site data on all albums reviewed by the site from 2003 to 2014. For each title, we collected various title and artist characteristics. Most important, we observe Pitchfork's score, a value in the interval [0,10] in increments of 0.1; as well as whether the title received the Best New Music (BNM) seal.[1]

Regarding demand, our main variable is given by last.fm plays. Importantly for us, last.fm provides a convenient Application Programming Interface (API) which allows us to download demand statistics, both accumulated number of plays and accumulated number of listeners of a given album.

Our third source of data comes from the site metacritic.com. As the URL suggests, this site aggregates the reviews of multiple critics of a variety of media content, including music.[2] It provides a metascore on a scale from 0–100, a weighted average of reviews where certain publications are given a greater weight (the weights are not public information). The set of albums covered by Metacritic is a subset of the number of albums included in last.fm. For this reason, we restrict our sample to albums covered by Metacritic.

We measure plays and listeners through June 2018. This allows us to focus on the era when Spotify and other streaming services were less prevalent, and recommendations from

1. We also have the full text of each album's review, although we are currently not making use of it.
2. Metacritic has operated since 2000 and draws from over 100 sources of professional music criticism. Specifically, the site reports a "Metascore" for each title on a scale from 0 to 100 when at least three of its underlying sources review an album (Waldfogel, 2015). Metacritic also covers films, TV shows, video games (and, in the past, books).

Table 1
Summary statistics (N = 3806)

Statistic	Mean	St. Dev.	Min	Max
year	2009.043	3.236	2003	2014
score	6.841	1.354	0	10
plays	2,287,295	5,332,940	734	69,868,320
listeners	115,225	215,453	93	2,704,966
metacritic	73.088	7.523	34	97
bestnewmusic	0.0796	0.271	0	1

last.fm are more relevant. This results in a total of 13401 albums with play information. For our main results we focus on the "Best New Music" era after 2003, which is 11051 albums.

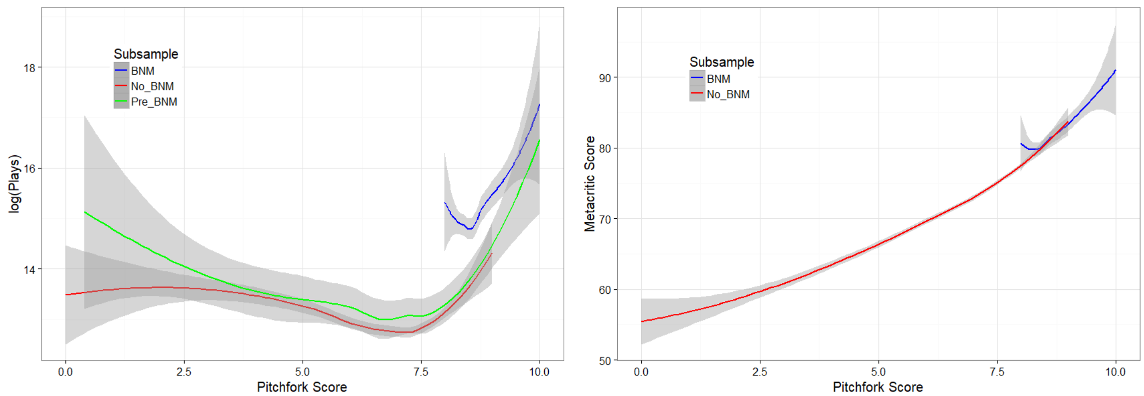
We exclude a small number of these albums. Because we want to compare to Metacritic, we restrict to albums that also appear there.[3] Because some (a small minority) of Pitchfork's reviews are for "old" albums, we remove any albums whose release year is more than 2 years before the review date. This does not catch all of the potentially not-new albums, though, because reissues sometimes mean a new release date, and some albums are collections of older songs. We therefore exclude manually any albums that include the words "anniversary," "greatest hits," "live," "soundtrack," "singles," or "edition." This is a small number of albums reviewed.

Table 1 presents some descriptive statistics of the data. The variable **score** corresponds to Pitchfork's score. As mentioned earlier, it varies on a scale from 0–10 with increments of 0.1. Given an average of 6.8 and a standard deviation of 1.3, we conclude that albums in the 8–10 range (the one we will consider) represent a relatively small fraction of the universe of albums considered by Pitchfork. In fact, **bestnewmusic**, the dummy variable indicating whether the album was certified by Pitchfork as BNM, has an average of close to 8%. **metacritic** corresponds to the metascore published by Metacritic.com. Unlike Pitchfork, **metacritic** varies from 0–100. **listeners** and **plays** corresponds to the accumulated number of listeners by album and accumulated number of plays per album, respectively, on last.fm. As often happens with media content, these distributions are rather right-skewed. For example, the average number of listeners of a given album is 115 thousand, but the higher value is close to 3 million.

■ **A first look at the data.** A natural first cut of the data is to plot last.fm plays for BNM titles vs non-BNM titles. This we do in the left panel of Figure 2, which allows us to make four points: First, for titles with a high Pitchfork score (about 7), there is a positive relationship between score and plays. Second, as mentioned above, BNM titles correspond to titles with higher Pitchfork's score. Third, the transition from non-BNM to BNM is not

3. This restriction cuts more than half of data; all our results are substantially more significant if we don't select on this. We believe this selection is useful, beyond allowing us to compare to any discontinuity in Metacritic score, because it also selects for genuinely new albums, which are therefore eligible for BNM, as we describe next.

Figure 2
Pitchfork score, BNM, and Metacritic score



based on a sharp threshold; that is, there is an interval of score values including both BNM and non-BNM titles. Finally, eyeballing the data suggest that the effect of BNM on last.fm plays is quite high (note the vertical axis is on a logarithmic scale).

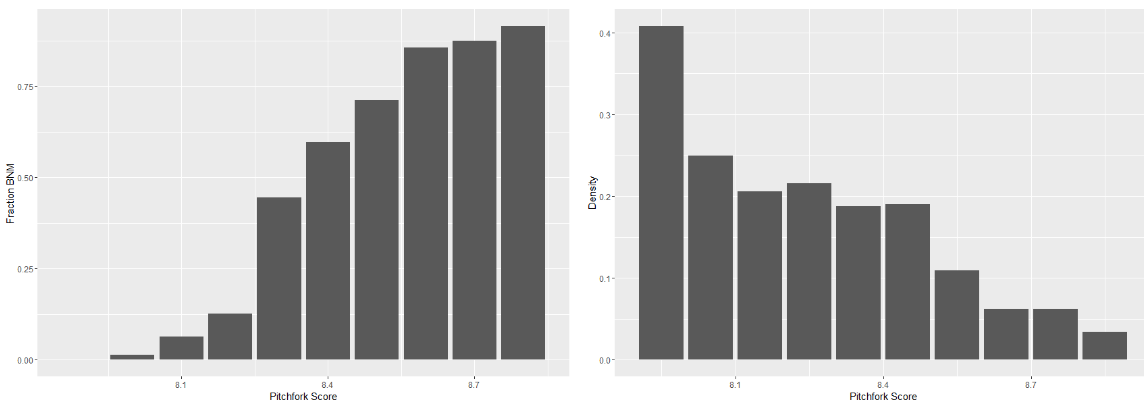
The right panel of Figure 2 complements the left panel. On the x axis, we continue to measure Pitchfork's score. On the y axis, instead of plays (left panel) we measure Metacritic's metascore. This panel allows us to make two points: First, Pitchfork's critics are remarkably mainstream among the set of critics aggregated by Metacritic: except for very low or very high score levels (where the number of observations is very small), the relation between Pitchfork and Metacritic score is remarkably correlated. Second, except for the left tail of BNM titles, the right panel of Figure 2 suggests that there is nothing special about the scoring of titles receiving the BNM seal. Specifically, the high correlation between Pitchfork scores and Metacritic scores is as close for these titles as for the remaining titles.

When we zoom in on the Pitchfork score interval where BNM and non-BNM titles coexist, our eyeball estimate indicates that BNM titles are associated with about twice as many plays as non-BNM. Although this comparison does not amount to proof of causality, we note that the difference between the BNM and non-BNM sets is quite substantial.

■ **Album score, BNM label, and regression discontinuity.** As mentioned earlier, the BNM seal is not based on a sharp threshold in terms of Pitchfork score, as one would require when applying standard Regression Discontinuity (RD). However, as Figure 3 illustrates, the data still allows us to infer the differential effect of BNM. Conventional wisdom puts the cutoff at a score of 8; we indeed find a significant discontinuty just above 8, although not at 8.0. Specifically, the left panel suggests that there is a significant increase in the probability of BNM as the score increases from 8.2 to 8.3. In other words, while 8.3 is not a sharp threshold in the sense of standard RD, there is a significant "discontinuity" at 8.3. In particular, we can show that the increase in probability is statistically significant. The p value for the equality test is lower than .001. No other score exhibits a significant discontinuity.

A condition for identification in RD design is that the density of the running variable be continuous at the discontinuity threshold (McCrary, 2008). The right panel of Figure 3 shows that the probability of receiving a given score declines with score level: an 8.7 score is

Figure 3
Fraction of titles with BNM label (right panel) and distribution of Pitchfork scores (left panel)



about 10 times less likely than an 8.0 score. However, the probability of an 8.2 score is very similar to the probability of an 8.3 score. This suggests that there is no “manipulation” of scores (which would result in “bunching” on one of the sides of the putative threshold). Specifically the equality test proposed by McCrary (2008) yields a coefficient of $\theta = -0.049$ (estimated size of jump) with a standard error of 0.085 and an implied p value of 0.562.

■ **Fuzzy regression discontinuity.** Formally, we estimate the structural equation

$$plays = g(score) + \gamma(score)\, BNM$$

where $g()$ and $\gamma()$ are general, *continuous* functions. Fuzzy RD uses the restriction that the discontinuity $\chi(score_i \geq 8.3)$ does not appear in the structural equation, but does appear in the determination of BNM to let us back out $\gamma(8.3)$, the effect of BNM at the discontinuity. Effectively, Fuzzy RD estimates

$$\begin{aligned} plays_i &= f_1(score_i) + \beta_1 \chi(score_i \geq 8.3) + \epsilon_{1,i} \\ BNM_i &= f_2(score_i) + \beta_2 \chi(score_i \geq 8.3) + \epsilon_{2,i} \end{aligned}$$

where f_1 and f_2 are smooth and then uses the ratio β_1/β_2 to recover $\gamma(8.3)$.

■ **Results.** We use the R code developed by Calonico, Cattaneo, and Titiunik (2015) for FRD estimation, bandwidth selection, and robust standard errors.[4] Table 2 presents the results of our FRD design. The point estimate of the BNM coefficient is equal to 9,520,200 for plays and 319,741 for listeners, with estimated p values of 0.011 and 0.005. The coefficients are quite large. Considering the base values (when the score value is 8.2) of 1,786,439 for number of plays and 133,981 for number of listeners at the threshold for the change in BNM, our estimates imply a six-fold increase in number of plays and a three-fold increase in the number of listeners resulting from the BNM certification.

■ **Robustness check: a placebo test.** The Best New Music label, introduced in 2001, is not as old as Pitchfork (founded in 1996). In other words, Pitchfork critics have been

4. We use the default MSE-optimal bandwidth selector. See https://cran.r-project.org/web/packages/rdrobust/rdrobust.pdf.

Table 2
Results

Dependent variable	plays	listeners
Best New Music	9,520,200 (0.011)	319,741 (0.005)
Observations	3,806	3,806
Effective observations below	330	330
Effective observations above	268	268
Mean(plays or listeners | score=8.2)	1,786,439	133,981

Note: p value in parentheses

scoring music for long before the BNM seal was introduced. This suggests a natural placebo test: if the discontinuity we observe at the 8.3 score level is due to the BNM seal, then such discontinuity should be absent during the time when the seal did not exist.

The left panel of Figure 2 illustrates this test. In addition to the BNM and non-BNM curves plotted before, we also plot the curve relating Pitchfork score to plays *before* the BNM label was introduced. The most remarkable feature of the latter curve is its continuous behavior: as predicted by our narrative, we observe no discontinuity around the score levels where the BNM discontinuity is observed.[5]

■ **Other tests.** Other available control variables do not alter the results. For instance, we considered year effects in the regression, as well as genre and label information. In particular we subdivided labels by "indie" and "major" in an attempt to see if there were differences in the patterns of the BNM seal. We did not find any.

Further research could leverage other plays data, but such data is difficult to come by. For instance, we do not have detailed history of play dates, so we cannot explore the temporal impact of Pitchfork reviews. Even to ascertain the characteristics of past experience of artists, one would like to know what their prior plays were from other works at the time of the review. This data is not available to us.[6]

4. Conclusion

Nate Patrin, former music critic at the *Rolling Stone*, argues that "professional music criticism is almost a thing of the past," as their opinion lost much of the relevancy it's had in the past (Singer, 2014). By contrast, our analysis suggests that, in a world of seemingly limitless supply, expert opinion plays an important role. Specifically, we show that the "best new music" seal given by the popular site Pitchfork has a significant effect on demand. We further argue that this effect takes place through increasing listener awareness of new

5. One caveat is that last.fm did not start tabulating plays until after this period started. Therefore this is an incomplete picture of plays and listeners for those albums.
6. We attempted to collect retrospective data from Google Trends to find out more about the artist's history, but found searches for various permutations of band names and narrowing words to not generate useful data for many of our artists. The complete data of Google Trends search information for our artists and album names is available upon request.

hits and operates both directly (listeners who visit Pitchfork) and indirectly (listeners who accept algorithm-based recommendations).

Our results suggest a series of questions which we do not address but could be addressed in future research. From a strategy point of view, one could consider the value of being an influencer and the monetization of reviews. As mentioned earlier, our estimates of the sales effect of BNM ranges from 24-40 thousand dollars (and this is likely a lower bound, as it excludes listeners we have no data for and other revenue streams beyond track plays). Conceivably, influencers might get a share of this pie. Related to this, from an antitrust point of view one might be concerned about the excessive power these reviewers / influencers might wield. This might be magnified to the extent that platforms like Spotify have, in some way, consolidated recommendations and delivery into the same platform.

References

Aguiar, Luis and Joel Waldfogel. 2018. "Platforms, Promotion, and Product Discovery: Evidence from Spotify Playlists." Tech. Rep. w24713, National Bureau of Economic Research.

Angrist, Joshua D. and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *The Quarterly Journal of Economics* 114 (2):533–575.

Berger, Jonah, Alan T. Sorensen, and Scott J. Rasmussen. 2010. "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales." *Marketing Science* 29 (5):815–827.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2015. "rdrobust: An r Package for Robust Nonparametric Inference in Regression-Discontinuity Designs." *R Journal* 7 (1).

Chevalier, Judith A. and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research* 43 (3):345–354.

du Lac, J. Freedom. 2006. "Giving Indie Acts A Plug, or Pulling It." *Washington Post* (Sunday, April 30).

Eliashberg, Jehoshua and Steven M. Shugan. 1997. "Film Critics: Influencers or Predictors?" *Journal of Marketing* 61 (2):68–78.

Friberg, Richard and Erik Grönqvist. 2012. "Do Expert Reviews Affect the Demand for Wine?" *American Economic Journal: Applied Economics* 4 (1):193–211.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1):201–209.

Hilger, James, Greg Rafert, and Sofia Villas-Boas. 2011. "Expert Opinion and the Demand for Experience Goods: An Experimental Approach in the Retail Wine Market." *The Review of Economics and Statistics* 93 (4):1289–1296.

Itzkoff, Dave. 2006. "The Pitchfork Effect." *Wired* (09.01.2006).

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2):698 – 714.

Newberry, Peter W. 2016. "An Empirical Study of Observational Learning." *The RAND Journal of Economics* 47 (2):394–432.

Reimers, Imke and Joel Waldfogel. 2021. "Digitization and Pre-Purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings." forthcoming in *American Economic Review* .

Reinstein, David A. and Christopher M. Snyder. 2005. "The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics." *Journal of Industrial Economics* 53 (1):27–51.

Scott, D.W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization.* John Wiley & Sons, New York, Chicester.

Simon, Herbert A. 1971. "Designing Organizations for an Information-Rich World." In *Computers, Communication, and the Public Interest*, edited by Martin Greenberger. Baltimore. MD: The Johns Hopkins Press, 40–41.

Singer, Dan. 2014. "Music Critics See Their Role and Influence Waning in The Era of Digital Music." *American Journalism Review* (November 13).

Trochim, William M. K. 1984. *Research Design for Program Evaluation: The Regression-Discontinuity Approach.* Sage Publications, Beverly Hills, CA.

Waldfogel, Joel. 2015. "Digitization and the Quality of New Media Products: The Case of Music." In *Economic Analysis of the Digital Economy*, edited by Avi Goldfarb, Shane M. Greenstein, and Catherine E. Tucker. The University of Chicago Press, 407–442.