Hit Or Miss: What One Hit Wonders Show About The Production And Consumption Of Cultural Products

Jacob Derechin
Yale University

Abstract

We analyze One Hit Wonders on the Billboard Hot 100 from perspective of both cultural production and optimal differentiation to try and understand how One Hit Wonders emerge. One Hit Wonders as a category do not cleanly fit into either the cultural production or optimal differentiation framework, and as in intermediate case illustrates the interplay between both consumer taste and institutional power in markets for cultural products.

Keywords: music, one hit wonders, culture.

1. Introduction

The One Hit Wonder exists in the liminal space of a cultural system. In order for something to qualify as a One Hit Wonder, it must achieve a great deal of success and popularity but simultaneously never reach that scale of success again. If one believes that success begets success, one would expect the Music Industry to be dominated by a small number of Superstars (Rosen 1981; Krueger 2005). Notably, the One Hit Wonder is not the opposite of the Superstar, the completely unknown artist is; thus one could interpret the mixed success of a One Hit Wonder as having medium level talent.

Hamlen Jr (1991) has found some empirical evidence that sonic attributes of music can influence record sales, while Hamlen Jr (1991) interprets the magnitude of this effect as evidence against the existence of Superstardom, but Krueger (2005) shows that this evidence could be interpreted either way. Evidence from online music downloading experiments finds that overall demand for music has socially mediated components and that manipulating the perceived popularity of a song can influence the long run demand for it Salganik et al. (2006); Salganik and Watts (2008). Analysis of the Television industry has shown that the inability of institutional actors to predict whether a new product will be successful shaped their decision making processes. Bielby and Bielby (1994) If the music industry works similarly, then one would not expect record labels to contribute to musical success by shaping the sound itself, but through other means like advertising. One Hit Wonders exist both by virtue of being brought to market by a record label, but also through the choices of consumers who by the music. Thus the analysis of One Hit Wonders must consider both the factors involved in musical production as well as consumption.

2. Theory

The life cycle of a work of music from its inception in the mind of the artist to the point when consumers eventually listen is a complex process involving many steps and intermediaries. It is reasonable to suppose that factors from the talent of the artists, the institutional actions of record labels, and the taste of consumers all have a part to play in the ultimate success of a song. There are broadly two classes of models characterizing where the "action" is in the song life cycle: models that focus on the institutional conditions in which songs are produced, and models that focus on consumer choices and taste. These models broadly map onto the categories of either Cultural Production or Cultural Consumption. One Hit Wonders occupy a middle ground in between each of these models, as they are both highly successful and unsuccessful. If the emergence of One Hit Wonders is largely driven by gate keeping, one would expect to see more One Hit Wonders at independent labels than major label imprints. One would also expect to see similar behavior between One Hit Wonders and the Debut song of an artist that goes on to have more hits. If the emergence of One Hit Wonders is driven primarily by the tastes of consumers, one would expect to see One Hit Wonders to be more dissimilar to the songs on the chart when they start than Debut songs and songs by established artists. Additionally one might expect songs that are more popular to be less likely to be One Hit Wonders.

2.1. Cultural Production

We are governed, our minds are molded, our tastes formed, our ideas suggested, largely by men we have never heard of.

— Edward Bernays, [Bernays \(1928\)](#)

The perspective of Cultural Production has its origins in the scholarship of the Frankfurt School, with the key observation being that culture was produced and consumed like a product [Denisoff and Bridges \(1983\)](#). Unlike the Frankfurt school scholars, I do not view the rise of popular music with disdain or as a sign of the corruption of art but as an object itself worth of study. Further refinement of the Cultural Production model by [Hirsch \(1972\)](#) introduces the notion of surrogate consumers, who are intermediaries like critics, promoters and radio DJs. In the Cultural Production model gatekeeping is the name of the game and surrogate consumers hold the keys to accessing the masses, and without their support. There is significant evidence that record labels themselves view these kinds of cultural gatekeepers as incredibly important, for as [Rossman \(2012\)](#) shows, despite many governmental attempts to make Payola (paying for radio airplay) illegal, it always reemerges in a new form. There has also been research studying how the organization of the music industry impacts the types of music that ultimately becomes popular. For example [Peterson and Berger \(1975\)](#) find that the more concentrated the record labels are, the less musical diversity there is. It is not clear what the right level of analysis for measuring label concentration, since the major record labels have semi-autonomous subdivisions with decision making power [Peterson and Berger \(1996\)](#). Shared ownership in the music industry does not always correspond to synchronization in behavior, as [Rossman \(2012\)](#) demonstrates in the case of radio station ownership. [Alexander \(1996\)](#) finds evidence of a quadratic relationship between record label concentration and musical diversity, so while it seems likely that the industry-level factors influence what is ultimately produced, the precise relationship is unclear. It is also important to consider the role of Billboard charts themselves as an intermediary in this market. There is evidence that Billboard charts are not immune to the spectre of Payola that hangs over all aspects of the music industry as well as its role as a source of information for decision makers within record labels [Anand and Peterson \(2000\)](#).

A key assumption of these cultural production models is that there are significant barriers in pushing cultural products to the masses. In this kind of environment, for a cultural product to make it to the consumers at all is an achievement. Thus it is reasonable to study how to the set of cultural products consumers choose comes to be, instead of their choices. Given the historical dominance of radio as the channel for music discovery, these models probably worked quite well in the past, but, given changes in the music industry they may fit less well going forward [Rossman \(2012\)](#). The rise of online music blogs [Dewan and Ramaprasad \(2012\)](#) as well as online musical piracy [Bhattacharjee et al. \(2007\)](#), have changed the channels through which consumers can discover and consume music, which have impacted the types of music that ultimately become popular. Today, it is also easier than ever to produce and distribute music than ever before. For example, fully featured digital audio workstations are available at low cost and even free, allowing an individual to record, produce, mix and master their music all on their own. As for distribution all one has to do is upload their new music to a service like Youtube or Soundcloud where millions of users can then find it. Getting music in digital stores like iTunes or streaming services like Spotify is a little more complex, but it is still very doable as an individual. Even so, this does not mean that the age of record labels is over as almost all popular artists are signed to a record label, so it is likely that institutional

factors can still have considerable influence.

2.2. Optimal Differentiation

“We didn’t think it pertained to us because we didn’t think of ourselves as a Disco band. We were R&B, Jazz Funk band. We never had a Chic hit after 1979. Good Times was it we were done. Disco Sucks completely, shut us down.”

— Nile Rodgers, [Stevens \(2013\)](#)

The Nile Rodgers quote above illustrates two key points about the demand for music: first genre is not a static and rigid concept, and second, that shifts consumer tastes can be incredibly important. In his quote Nile Rodgers is referring to the disco demolition night, where people brought an burned disco records disco records during a White Sox game, in backlash to the genre’s popularity. [Stevens \(2013\)](#) A bunch of people getting together to burn disco records and riot is certainly quite an extreme event, but since consumers are ultimately the ones who actually purchases the music their preferences naturally influence what kinds of music becomes popular. For example [Rossman \(2012\)](#) demonstrates how conservative backlash the political statements made by The Dixie Chicks systematically cut them off from radio airplay, even as they were incredibly popular just, like Chic was before the Disco Sucks movement. These kinds of exogenous cultural events can drive the existence of One Hit Wonders, as songs from movies and other cultural domains can end up charting on the Hot 100 chart. For example a bagpipe rendition of Amazing Grace by the Royal Scots Dragoon Guards probably has less to do with a new consumer preference for bagpipe music and more to do with them being the successor to the Royal Scots Greys Band, which had existed for 300 years and recently disbanded [Jancik \(2008\)](#).

Nile Rodgers’ statement about genre also serves to illustrate some of the different ways in which Genre can mean different things to different people within the musical ecosystem. The concept of Genre itself is also fluid and fundamentally not stable over time [Lena \(2015\)](#). For example, the number one song on the Hot 100 in the week August 4th, 1958 is Poor Little Fool by Ricky Nelson and the number one song in the week November 18th, 2017 is Rockstar by Post Malone featuring 21 Savage. I invite the reader to listen to both songs and pay attention to the significant differences. Poor Little Fool is an acoustic song that could be characterized as Rock and Roll, while Rockstar has many electronic elements as well as significant auto tune and could be characterized as Rap or Hip Hop. The key point about the relational nature of genre is that both songs given the proper reference group could be considered pop music [Lena \(2015\)](#). Given that the boundaries of Genre are fluid, it is important to pin down how they are defined. [Lena and Peterson \(2008\)](#) find that musical Genres tend to exist in 4 types of categories: Avant-Garde, Scene Based, Industry Based and Traditionalist. For Avant-Garde, the genre is highly experimental can be confined to a few artists; for a scene based genre, the music exists within a spatially localized community; for Industry Based Genres, the recording industry is involved and traditionalist genres are genres that are past their peak with people constantly looking back to the music that was made in the “glory days” [Lena and Peterson \(2008\)](#).

It is important to look at genres not only as a set of musical conventions but also as communities of artists, fans, and members of the recording industry. All three of these separate

sets of actors play a role in defining the bounds of genres. For example studies of Rap music demonstrate how the artists use musical samples to define the bounds of their genre and communicate with each other [Lena \(2004\)](#); [Lena and Pachucki \(2013\)](#) Notably it was found that artists were more tolerant of musical innovation than the consumers were leading to mismatch [Lena and Pachucki \(2013\)](#). There is also evidence that the musical tastes of consumers systematically change. By analyzing listening behavior on Spotify, it was found that individuals have different musical preferences at different times of day and different musical preference for different seasons [Park et al. \(2019\)](#). Additionally there has been shifts over time in the lyrical composition of popular music over time, with negative words becoming more common and positive words becoming less common [Brand et al. \(2019\)](#). Consumers are also sensitive to the gender of artists, as charting behavior is different for male and female artists [Lafrance et al. \(2011\)](#).

Finally, it is essential to examine how the construction of categories themselves constrains the choices consumers face. A good model for addressing these concerns is the two stage valuation model. In this model there are two steps, first objects are grouped into categories and then consumers pick items from the categories [Zuckerman \(2016\)](#). In the case of popular music, the clear analogue is musical genre or a radio format. For example program managers at radio stations have a large range of autonomy in selecting the music that is played as long as it is reasonably within the bounds of the station’s format [Rossman \(2012\)](#). Under this model objects that are difficult to group into categories are expected to be less popular and this is supported by evidence in the radio market as well as in the stock market [Rossman \(2012\)](#); [Zuckerman \(1999\)](#). In this model, competition primarily happens within categories, so it is important for songs to stand out among their peers while not going too far. Previous analysis of the songs on the Hot 100 finds evidence for this phenomenon [Askin and Mauskopf \(2017\)](#).

3. Data

3.1. Data Sources

The primary metric of musical success used for this research is the Hot 100 chart from Magazine Billboard which attempts to measure the most popular songs based on consumer sales data ([Billboard 2018](#)). The Billboard Charts are a standard measure of the popularity of music and their information is also internalized by the music industry ([Burgess 2014](#)). Billboard releases weekly changes to the Hot 100 chart and the data used here spans from August 4th, 1958 to November 18th, 2017. I also track Radio Airplay using the Billboard Radio Songs chart, which spans November 3, 1990 to November 18th, 2017. I split the data into two main segments, August 4th, 1958 to May 18, 1991 and from May 25, 1991 to November 18th, 2017. In recent years Billboard has made significant changes to the algorithm that calculates the ranking for the Hot 100 chart. On May 25, 1991 Billboard switched to using point of sale data from Nielsen Soundscan to measure popularity instead of the record store surveys they had previously been using, and [Anand and Peterson \(2000\)](#) show that this switch does impact which songs chart going forward. The way users consume music has also changed significantly with the rise, digital downloads, music blogs, and streaming services. The rise of music blogs provided an alternative avenue for consumers to find music and this impacted which songs eventually charted [Dewan and Ramaprasad \(2012\)](#). To meet the changes in musical con-

sumption, Billboard changed the way it computed the Hot 100 chart in response, adding non single songs from Albums (Album Cuts) on December 5, 1998 [Giles \(2007\)](#), Digital Download Sales on February 12, 2005 [Trust \(2015\)](#), some streamed media on August 11, 2007 [bil \(2013\)](#), digital radio services like Spotify in January 2013 [Pietroluongo \(2013\)](#), and finally by incorporating Youtube views in February 23, 2013 [Sisario \(2013\)](#). Given it is all of these changes to how the Hot 100 is calculated, it is unlikely that the Billboard Hot 100 charts pre May 25, 1991 and post May 25, 1991 are comparable.

To get basic metadata about the songs in question like record label, whether the song was from a soundtrack, and type of artist, I drew from the music metadata database MusicBrainz.¹ Musicbrainz is a collection of user submitted music metadata and releases their database as public domain data [MusicBrainz \(2019\)](#). In order to measure the sonic attributes of the songs, I used the audio fingerprinting database run in partnership with the MusicBrainz called AcousticBrainz. The audio fingerprinting data is also generated by user submission and released to the public [AcousticBrainz \(2019\)](#). Records in the MusicBrainz database can be directly linked with corresponding fingerprinting data from AcousticBrainz making it simple to combine data from both sources. Since data from MusicBrainz and AcousticBrainz is released to the public and continuously updated it, it is straightforward to replicate research that draws from it.

3.2. Summary Statistics and Data Types

Summary Statistics

There are 27304 unique Song-Artist pairs in the entire data set. Of these 27304 songs I was able to link 21054 of them with entries in the MusicBrainz database. There is High Level Audio fingerprinting data available for 18704 of these songs and Low Level Audio fingerprinting data available for 18749 of the songs.² I classify a song as a One Hit Wonder, if the song is the only song that ever made the Billboard Hot 100 for every artist credited on the song. For example if a song is duet between two artists, where one of the artists has no other hits on the Hot 100 and one of the artists has other hits on the Hot 100, this song would not count as a One Hit Wonder by my metrics. There are 3115 songs that meet this criteria as One Hit Wonders. Thus about 11.4 % of the songs on the Billboard Hot 100 are One Hit Wonders. As an additional robustness check I also use the corpus of songs coded as One Hit Wonders by music journalist Wayne Jancik, One-Hit Wonders [Jancik \(2008\)](#). There are 681 songs that Jancik classified as One Hit Wonders, but Jancik only tracked songs until 1992. This corpus of songs will only be used as a robustness check for the pre May 25, 1991 subsample of the dataset.³

Data Types

The audio fingerprinting data from Acousticbrainz provides a wealth of under the hood data about sonic features of the songs as well as high level machine learning models to categorize attributes about the songs. The machine learning data includes both classification data (for example whether the song is bright or dark) as well as the probabilities it assigns to song being in each category. Inspired by the work done by [Kovács and Hannan \(2015\)](#) on concep-

tual distance, I will estimate my models using the discrete classifications and the continuous probability values separately. The high level machine learning models are trained using the low level data and to avoid multicollinearity and overfitting problems I use a subset of the audio features that I think are salient.

From the low level data I use the following features: song length, gain, average loudness, bpm, chords changes rate, key as estimated by the main chord progression, a measure of chord diversity, a measure of confidence in the chord estimation, danceability, sensory dissonance, dynamic complexity, musical key as estimated by a different key detection algorithm, and a measure of key detection confidence. More information about how these values are calculated can be found in the Essentia documentation here: [Essentia \(2019\)](#).

From the high level data I use the following classifiers: song brightness vs darkness [AcousticBrainz \(2019\)](#), whether the song is instrumental or vocal [AcousticBrainz \(2019\)](#), whether the vocals are female or male [AcousticBrainz \(2019\)](#), whether the music is electronic or not [Laurier et al. \(2009\)](#), whether the song is sad or not [Laurier et al. \(2009\)](#), whether the music is tonal or atonal [AcousticBrainz \(2019\)](#), whether the music is party music or not [Laurier et al. \(2009\)](#), whether the music is danceable or not [AcousticBrainz \(2019\)](#), whether the music is acoustic or not [Laurier et al. \(2009\)](#), whether the music is happy or not [Laurier et al. \(2009\)](#), whether the music is aggressive or not [Laurier et al. \(2009\)](#), whether the music is relaxed or not [Laurier et al. \(2009\)](#), an additional more general classifier of mood [Hu and Downie \(2007\)](#), and finally multiple different estimators of genre [Homburg et al. \(2005\)](#); [i Termens \(2009\)](#); [Tzanetakis and Cook \(2002\)](#); [Sturm \(2012\)](#). More information about the AcousticBrainz high level data can be found here: [Documentation \(2019\)](#).

4. Analysis

4.1. Summary Statistics: Fat Tails

In order to begin the analysis it is important to make sure the data is well behaved. Given the possibility that this could be a superstar market I checked many of the explanatory variables to see if they were Fat Tailed ([Rosen 1981](#); [Krueger 2005](#)). The presence of Fat Tails is important because standard statistical tools such as Ordinary Least Squares and the Pearson Correlation Coefficient are not robust to the presence of outliers. Table 1 shows the non Fisher adjusted Kurtosis values for some of the explanatory variables of interest. Despite the Kurtosis the distributions of some of the explanatory variables are bizarre looking. For example Figure 2 shows a histogram for the counts of maximum number of weeks on the Hot 100 Chart, and there are multiple breaks in monotonicity as well as significant lumping around various values. Figure 3 shows the Histogram for fraction of Songs at a given Debut Rank, which also exhibits some breaks in monotonicity but overall this is much more well behaved. Figure 4 shows the Histogram for fraction of Songs at a given Peak Rank and this distribution also has lumping and non monotonicity. Thus care is needed when analyzing this data.

Table 1: Kurtosis for Various Variables

Variable Name	Kurtosis
Max Weeks on the Hot 100	6.498853735915996
Peak Rank on the Hot 100	1.7077485138296822
Debut Rank on the Hot 100	7.258110753598563
Total Number of Songs by Imprint	21.16238099594883
Total Number of Songs by All Label Types	150.5836149139357

4.2. The Similarity Analysis

In this section I analyze the relationship between sonic properties of the songs and their success. As [Kovács and Hannan \(2015\)](#) demonstrate the distance/similarity between different types of categories can be quantified and measured. Using similar methods as [Askin and Mauskopf \(2017\)](#), I construct vectors of the audio fingerprinting elements described in Section 3 and compute their similarity with one another. One notable difference is the way I measure distance between musical keys, as [Askin and Mauskopf \(2017\)](#) use dummy variables for each of the different musical keys effectively measuring distance between the musical keys using the discrete metric. Different musical keys are more and less similar to each other based on the underlying pitches that make up the key. These relationships are represented in music theory as the Circle of Fifths and Figure 1 shows a visual representation of the Circle of Fifths. The Circle of Fifths forms a graph connecting all musical keys to each other. I call the geodesic distance between two Musical Keys within the Circle of Fifths graph Harmonic Distance. Working with Harmonic Distance presents some technical challenges, since the distance isn't reducible to coordinates so I don't see an obvious way to calculate inner products.⁴ Equation 1 shows the formula for calculating the Cosine Similarity between vectors A and B, this value depends on computing the inner product between A and B. Other work studying the impact of audio fingerprinting of songs on chart performance use Cosine Similarity as their similarity metric [Askin and Mauskopf \(2017\)](#), so I include this similarity function to be comparable but it cannot accommodate Harmonic Distance. All Cosine Similarity values calculated do not include Harmonic Distance. This motivated me to also use inverse Euclidean Distance as an additional similarity measure, since the calculation is function the square of distances allowing for more flexibility in the metrics that can be used. For my purposes I calculate the component level distances using the Euclidean metric for all components of the Audio Fingerprint but calculate distance with the Harmonic Distance metric for Musical Key. Equation 2 shows how Euclidean Similarity is normally calculated.

$$\text{Cosine Similarity} = \frac{A \cdot B}{\|A\| \|B\|} \quad (1)$$

$$\text{Euclidean Similarity} = \frac{1}{\|A - B\|} \quad (2)$$

Figure 1: Circle of Fifths Image by Watchduck (a.k.a. Tilman Piesk)

Table 2: Primary Variables

Variable Name	Description
One_Hit_Wonder	This is a dummy variable which is 1 if the song is a One Hit Wonder and 0 otherwise.
Peak_Position	This is the peak Rank of the song on the Hot 100.
major_label	If the record label that released the song is owned by a major record label.
Soundtrack	This is a dummy variable indicating whether the song was released as part of a soundtrack on its initial release.
christmas_season	This is a dummy variable indicating whether the song initially entered the Hot 100 in November or December.
summer_season	This is a dummy variable indicating whether the song initially entered the Hot 100 in June, July or August.
avg_similarity_prev	This represents the average distance in the vector space of audio features between the song and all songs charting the year before it enters the charts.
var_similarity_prev	This represents the variance in distance between the song and all songs charting the year before it enters the charts.
dummy_	These variables are time fixed effects for every 5 year interval in the dataset.

4.3. The Regressions

The dependant variable of interest in this analysis is the binary variable representing whether or not the song in question is a One Hit Wonder. I estimate a logistic regressions for the Pre May 25th, 1991 subset of songs, the Post Post May 25th, 1991 subset of songs, as well as doing a Regression Discontinuity in Time Analysis to assess the impact of the change how songs enter Billboard charts impacted production of One Hit Wonders. [Hausman and Rapson \(2018\)](#) Within each time period I estimate the model using all songs, as well as only comparing One Hit Wonders to Debut songs. I define debut songs as songs for which none of the performers have a previous song on the Hot 100. This accounts for collaborations, so that if a new artist's first song to make the charts in a collaboration with a more established artist, this song is not a debut song. The regression discontinuity models also include models calculated with the Imbens-Kalyanaraman Optimal Bandwidth and these models are denoted as Optimal Bandwidth. [Imbens and Kalyanaraman \(2012\)](#) The models which only include One Hit Wonders and Debut Songs are labeled comparison. The Kurtosis of many of the explanatory variables are well above 3, it is reasonable to expect that the tails are fat enough to cause problems in Logistic regressions. To compensate for outliers I use the robust generalized linear model regression outlined in [Cantoni and Ronchetti \(2001, 2006\)](#); [Stigler and Quast \(2015\)](#) and denote these models as Robust in the regression tables. Table 2 shows the primary variables used in the analysis.

For each period, I model the audio fingerprinting features using either Euclidean Similarity or Cosine Similarity. These values are calculated base on the similarity between the song of interest and the songs that were on the chart up to one year before the song of interest entered

the chart. To account for different ways that audio features could impact the success of a song, I use both the average similarity and the variance in similarity. The average similarity may measure how representative the song of interest is when compared to the songs on the chart up to a year before, while the variance may represent how effectively the song matches a niche. The next variable of interest is major label, which represents whether or the song is release by a label that is owned by a major record label. The list of Major Labels and their Subsidiaries is included in the appendix in Table 15. This is a stand in which represents the institutional advantages that being part of a major label may bring. I also consider the peak rank the song reached on the Billboard Hot 100, seasonal effects for the summer and winter, as well as time fixed effects for every five year interval.

Table 3 shows the regression models using Euclidean Similarity for the Pre May 25th, 1991 subset. The effects are largely consistent across models, with Peak Position having a statistically significant and positive coefficient across models and major label having a statistically significant and negative effect across models. This suggest that reaching a higher (closer to 1) peak position and being released by a major record label reduces makes it less likely that a song is a One Hit Wonder. Table 3 shows the models for Euclidean Similarity on the Post May 25th, 1991 subset and the story is quite similar. Peak Position is statistically significant and positive across all models while major label is only significant in the base models but not the comparison models. The major label coefficient is negative across all models. Average Similarity is statistically significant at the 10% and negative in the base models but not statistically significant in the comparison models. In the Pre May 25th, 1991 subset the average similarity effect is also negative so this is consistent across the different subsets. Table 5 shows the models for the Variance in Euclidean Similarity in the Pre May 25th, 1991 subset. The Peak Position and Major Label variables have the same sign as they do in previous models and are statistically significant across all models. Notably, Variance in Euclidean similarity is statistically significant and negative in the base models but not statistically significant in the comparison models. Table 6 shows the models for the Variance in Euclidean Similarity in the Post May 25th, 1991 subset. Once again the coefficients have the same sign as in all other Euclidean Similarity Models. Peak position is statistically significant across all models while major label is only significant in the base models but not the comparison models. Table 7 shows the models using Average Cosine Similarity using the Pre May 25th, 1991 subset. Across all models, peak position is statistically significant and positive, while major label is statistically significant and negative. Average Similarity is statistically significant at the 10% level and positive in the comparison models. This is in contrast to the sign of Euclidean Similarity, which is positive. Table 8 shows the models using Average Cosine Similarity using the Post May 25th, 1991 subset. The coefficient for Peak Position is statistically significant and positive across all models while the coefficient for Major Label is only statistically significant an negative in the base models and not the comparison models. Average similarity is not statistically significant in any models and the sign is negative in the base models but negative in the comparison models. Table 9 shows the models using Variance in Cosine Similarity using the Pre May 25th, 1991 subset. Peak Position and Major Label are statistically significant across all models, and have the same signs as previous models. Variance in Similarity is negative across all models, but only statistically significant at the 10% level in the Base Model but not the Robust Base Model. Table 10 shows the models using Variance in Cosine Similarity using the Post May 25th, 1991 subset. Peak Position is statistically significant and positive

across all models. Major Label is statistically significant and negative in the base models but not the comparison models. Variance in Cosine Similarity is statistically significant at the 10% level and negative in the comparison models but not in the base models. Table 11 and Table 12 show the Regression Discontinuity models for Average and Variance in Euclidean similarity respectively, while Table 13 and Table 12 show the Regression Discontinuity models for Average Cosine Similarity and Variance in Cosine Similarity. The effect of the discontinuity is not significant in any models, so there is not sufficient evidence to conclude that the change in the Billboard Charting algorithm had any impact on the production of One Hit Wonders.

5. Discussion

Of the covariates studied Peak Position on the Hot 100 had the most robust and consistent effect across models. These results are consistent with other results that suggest that record labels are responsive to sales and the Hot 100 charts are informative to record labels. ?? The next most consistent and robust effect is that of Major Label, showing that records released from labels that are owned by major labels tend to be less likely to be One Hit Wonders. The effect of Major Labels is only significant in the base models Post May 25th, 1991 but not in the comparison models. This may suggest that the changes in the industry structure since May 25th, 1991 may have reduced the advantage for records released from Major Labels, since the comparison model is a more direct comparison of new releases than the base model. Since the Post May 25th, 1991 subset is smaller than the Pre May 25th, 1991 subset, it is also possible that the analysis is simply under powered, especially in the case of the comparison models. The discontinuity effect of the Billboard Charting Algorithm is not significant in any models, so there isn't evidence of a direct effect on One Hit Wonders by the algorithm change and across models the 5 Year Time Fixed effects were largely not significant. Thus it is reasonable to conclude that any period driven institutional changes were driving responsible for the differences between the coefficients in the Pre and Post May 25th, 1991 models. The evidence for institutional advantage via Major Labels is clear in the Pre May 25th, 1991 period but it is not as clear in the Post Period. By and large neither the seasonal effects nor the soundtrack variables had robust effects across models. Thus this analysis cannot speak to the precise relationship between seasonality and the production of One Hit Wonders nor the impact of music that crosses over from other cultural domains.

As a whole the similarity measures did quite poorly at explaining variance in the One Hit Wonder rate. The average similarity measures sometimes switched sign across models within period for a given similarity measure. It is also important to note that for a given period average Euclidean Similarity and average Cosine Similarity sometimes had different signed coefficients in the models, suggesting that the differences in what they measure might be meaningful. Notably in the case of audio signals, the magnitude of the vector components may be the key sonic differentiating factor, not the angle between the fingerprint vectors. For example considering only the dimensions of loudness and bpm, the vectors $(-4, 80)$ and $(-8, 160)$ are parallel in this 2 dimensional space but would subjectively sound quite different. Vector two would be much quieter and much faster than vector one but their cosine similarity is still one. The results for both similarity metrics is quite inconsistent in this context, so it

I am cautious not to over interpret what the differences between these two metrics says in this context, but given that Euclidean Similarity can accommodate harmonic distance and seems to better map onto the perceptible differences between two given audio fingerprints it seems better suited to this context. The analysis for Variance in Similarity is much more illuminating. Unlike average similarity, variance in similarity in had a negative coefficient in all models for both Euclidean and Cosine similarity. It is also statistically significant in more models than any of the average similarity measures are. variance in similarity may represent niche matching better than average similarity, since higher variance would suggest that the song is closer to the songs it is close to and or further away than the songs it is far from. Without good genre data to sort the songs by, the measure of variance may be picking up the latent information of how well the song matches to its specific sonic niche. One could certainly compute the average distance of a song within its genre, but it is possible that variance may out preform even in that setting. Variance seems to be the clearer analogue of differentiation , since it picks up both local similarity as well as distinction from other surrounding songs and if optimal differentiation is a fractal process even within genres then using the average similarity would miss this second order differentiation information. This is not without drawbacks, as using variance as an explanatory variable requires Kurtosis to exist and be well defined. This is not guaranteed with fat tailed distributions, but with close to normal distributions and sufficiently large sample sizes this should not be too much of a problem. All things considered it is worth looking at variance in similarity as a measure of sonic differentiation.

6. Conclusion

The results of this analysis suggest that aspects of both the Cultural Production Model and the Optimal Differentiation are necessary for explaining the production of One Hit Wonders. As predicted by the Cultural Production Model, songs produced by a Major Label or its subsidiary do seem to be less likely to be a One Hit Wonder while songs that do not sell well as represented by the Peak Rank on the Billboard Hot 100 are more likely to be a One Hit Wonder. This suggests that both the institutional forces that shape which songs make it to market, and the consumers who choose which songs they like play a role in shaping the overall success of a record. The results from the audio fingerprint data are mostly inconclusive, but there is some evidence that greater variance in sonic similarity seems to reduce the chance of the song being a One Hit Wonder. This is consistent with the theory of Optimal Differentiation. Additionally when working with high dimensional data, it is important to consider theoretical motivation for choice of similarity metric as results may change depending on the choice of metric. Cultural products exist as a transaction between Cultural Producers and Cultural Consumers, as the analysis of One Hit Wonders shows, their success fundamentally depends on the forces that shape both Cultural Production and Cultural Consumption.

Notes

¹I also have discogs data downloaded and I may be able to get more reliable data on label imprints as well as assigned genres. I'll need to look into that going forward.

²I know that that only capturing two thirds of your sample space isn't great. I'm guessing a lot of the missing songs are actually in the database but Billboard and MusicBrainz use different naming conventions

and I have exhausted most of the low hanging fruit of programmatically matching the names and have to match the rest manually.

³I haven't done these robustness checks, yet but I will in a future draft.

⁴I'm not super sure about this one, but since these distances are defined by a graph the math gets really weird. Just because I don't know how to calculate inner products doesn't mean its impossible, but it didn't seem like a huge deal so I didn't want to look into it in a ton of depth.

References

(2013). Billboard hot 100 to include digital streams.

AcousticBrainz (2019). Acousticbrainz.

Alexander, P. J. (1996). Entropy and popular culture: product diversity in the popular music recording industry. *American Sociological Review*, 61(1):171–174.

Anand, N. and Peterson, R. A. (2000). When market information constitutes fields: Sense-making of markets in the commercial music industry. *Organization Science*, 11(3):270–284.

Askin, N. and Mauskapf, M. (2017). What makes popular culture popular? product features and optimal differentiation in music. *American Sociological Review*, 82(5):910–944.

Bernays, E. L. (1928). *Propaganda*. Ig publishing.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., and Telang, R. (2007). The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Science*, 53(9):1359–1374.

Bielby, W. T. and Bielby, D. D. (1994). "all hits are flukes": Institutionalized decision making and the rhetoric of network prime-time program development. *American Journal of Sociology*, 99(5):1287–1313.

Billboard (2018). Billboard hot 100. <https://www.billboard.com/charts/hot-100>.

Brand, C. O., Acerbi, A., and Mesoudi, A. (2019). Cultural evolution of emotional expression in 50 years of song lyrics.

Burgess, J. (2014). *“All Your Chocolate Rain Are Belong To Us?” Viral Video, YouTube and the Dynamics of Participatory Culture*, pages 86–96. UTSePress.

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Association*, 96(455):1022–1030.

Cantoni, E. and Ronchetti, E. (2006). A robust approach for skewed and heavy-tailed outcomes in the analysis of health care expenditures. *Journal of Health Economics*, 25(2):198–213.

Denisoff, R. S. and Bridges, J. (1983). The sociology of popular music: A review. *Popular Music Society*, 9(1):51–62.

Dewan, S. and Ramaprasad, J. (2012). Research note—music blogging, online sampling, and the long tail. *Information Systems Research*, 23(3-part-2):1056–1067.

- Documentation, A. (2019). Acousticbrainz.
- Essentia (2019). Music extractor.
- Giles, D. E. (2007). Survival of the hippest: life at the top of the hot 100. *Applied Economics*, 39(15):1877–1887.
- Hamlen Jr, W. A. (1991). Superstardom in popular music: Empirical evidence. *The Review of Economics and Statistics*, pages 729–733.
- Hausman, C. and Rapson, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics*, 10:533–552.
- Hirsch, P. M. (1972). Processing fads and fashions: An organization-set analysis of cultural industry systems. *American journal of sociology*, 77(4):639–659.
- Homburg, H., Mierswa, I., MÃller, B., Morik, K., and Wurst, M. (2005). A benchmark dataset for audio classification and clustering.
- Hu, X. and Downie, J. S. (2007). Exploring mood metadata: Relationships with genre, artist and usage metadata.
- i Termens, E. G. (2009). Audio content processing for automatic music genre classification: descriptors, databases, and classifiers.
- Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3):933–959.
- Jancik, W. (2008). *One-hit wonders*. Booksurge Pub.
- Kovács, B. and Hannan, M. T. (2015). Conceptual spaces and the consequences of category spanning. *Sociological science*, 2:252–286.
- Krueger, A. B. (2005). The economics of real superstars: The market for rock concerts in the material world. *Journal of Labor Economics*, 23(1):1–30.
- Lafrance, M., Worcester, L., and Burns, L. (2011). Gender and the billboard top 40 charts between 1997 and 2007. *Popular Music and Society*, 34(5):557–570.
- Laurier, C., Meyers, O., Serra, J., Blech, M., and Herrera, P. (2009). Music mood annotator design and integration.
- Leifeld, P. (2013). texreg: Conversion of statistical model output in R to L^AT_EX and HTML tables. *Journal of Statistical Software*, 55(8):1–24.
- Lena, J. C. (2004). Meaning and membership: samples in rap music, 1979–1995. *Poetics*, 32(3):297–310.
- Lena, J. C. (2015). *Genre: Relational Approaches to the Sociology of Music.*, pages 149–160. Routledge, New York.
- Lena, J. C. and Pachucki, M. C. (2013). The sincerest form of flattery: Innovation, repetition, and status in an art movement. *Poetics*, 41(3):236–264.

- Lena, J. C. and Peterson, R. A. (2008). Classification as culture: Types and trajectories of music genres. *American Sociological Review*, 73(5):697–718.
- MusicBrainz (2019). Musicbrainz.
- Park, M., Thom, J., Mennicken, S., Cramer, H., and Macy, M. (2019). Global music streaming data reveal diurnal and seasonal patterns of affective preference. *Nature Human Behaviour*.
- Peterson, R. A. and Berger, D. G. (1975). Cycles in symbol production: The case of popular music. *American sociological review*, pages 158–173.
- Peterson, R. A. and Berger, D. G. (1996). Measuring industry concentration, diversity, and innovation in popular music. *American Sociological Review*, 61(1):175–178.
- Pietroluongo, S. (2013). New dance/electronic songs chart launches with will.i.am britney at no. 1.
- Rosen, S. (1981). The economics of superstars. *The American economic review*, 71(5):845–858.
- Rossmann, G. (2012). *Climbing the charts: What radio airplay tells us about the diffusion of innovation*. Princeton University Press.
- Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856.
- Salganik, M. J. and Watts, D. J. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social psychology quarterly*, 71(4):338–355.
- Sisario, B. (2013). What’s billboard’s no. 1? now youtube has a say.
- Stevens, M. (2013). *Nile Rodgers: The Hitmaker*. BBC Wales.
- Stigler, M. and Quast, B. (2015). *rddtools: Toolbox for Regression Discontinuity Design (‘RDD’)*. R package version 0.4.0.
- Sturm, B. L. (2012). An analysis of the gtzan music genre dataset.
- Trust, G. (2015). Ten years ago, the digital download era began on the hot 100.
- Tzanetakis, G. and Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302.
- Zuckerman, E. W. (1999). The categorical imperative: Securities analysts and the illegitimacy discount. *American journal of sociology*, 104(5):1398–1438.
- Zuckerman, E. W. (2016). Optimal distinctiveness revisited. *The Oxford handbook of organizational identity*, page 183.

7. Appendix

Table 3: Average Euclidean Similarity Pre May 25th, 1991 Models

	Base Model	Robust Model	Base Model Comparison	Robust Model Comparison
(Intercept)	-1.8372336*** (0.5015017)	-1.7346280** (0.5372498)	-1.4586428* (0.6615380)	-1.2774248 (0.6641449)
christmas_season	-0.0627843 (0.1107724)	-0.0511108 (0.1202417)	0.1278573 (0.1502164)	0.1480489 (0.1507348)
summer_season	-0.0459005 (0.0945051)	-0.0274258 (0.1023400)	-0.0760320 (0.1229059)	-0.0690289 (0.1237278)
Soundtrack	0.4404533 (0.3450558)	0.5718035 (0.3481569)	0.4734200 (0.5044894)	0.4249198 (0.5043433)
Peak_Position	0.0181439*** (0.0013210)	0.0170348*** (0.0014217)	0.0278399*** (0.0017349)	0.0269178*** (0.0017366)
major_label	-0.6003508*** (0.0801953)	-0.5898758*** (0.0861244)	-0.2893111** (0.1069624)	-0.3050688** (0.1074333)
avg_similarity_prev	-1.0397400 (0.6043275)	-1.0985804 (0.6562544)	0.5568929 (0.8004481)	0.4915792 (0.8053683)
dummy__392	-0.5590514 (0.3286430)	-0.5423322 (0.3462953)	-0.5136857 (0.4414782)	-0.5923554 (0.4419588)
dummy__393	-0.7026286* (0.3289690)	-0.7018519* (0.3470808)	-0.5177134 (0.4431545)	-0.5797534 (0.4436161)
dummy__394	-0.6470132 (0.3313715)	-0.7420340* (0.3512503)	-0.5038830 (0.4445900)	-0.6105223 (0.4452532)
dummy__395	-0.3209201 (0.3298907)	-0.3461435 (0.3479810)	-0.3395873 (0.4435757)	-0.3781979 (0.4437721)
dummy__396	-0.4831849 (0.3347420)	-0.4924156 (0.3532888)	-0.3627514 (0.4493783)	-0.4339734 (0.4497784)
dummy__397	-0.3437266 (0.3342136)	-0.3126131 (0.3519780)	-0.3660307 (0.4485345)	-0.4245714 (0.4488815)
dummy__398	0.0208592 (0.3691840)	-0.0595824 (0.3923779)	-0.0176664 (0.4913043)	-0.0859148 (0.4914683)
AIC	4906.1250847		2263.2138143	
BIC	5006.3546908		2341.0403499	
Log Likelihood	-2439.0625423		-1117.6069072	
Deviance	4878.1250847		2235.2138143	
Num. obs.	9502	9502	1918	1918

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 4: Average Euclidean Similarity Post May 25th, 1991 Models

	Base Model	Robust Model	Base Model Comparison	Robust Model Comparison
(Intercept)	-0.6247531 (0.6621979)	-0.5332895 (0.7196493)	-0.6220949 (0.8386810)	-0.6142108 (0.8387598)
christmas_season	-0.2173178 (0.1385148)	-0.2634145 (0.1532805)	0.0486190 (0.1824410)	0.0362636 (0.1824157)
summer_season	-0.0396571 (0.1175648)	-0.0563037 (0.1276816)	-0.1581428 (0.1483795)	-0.1505449 (0.1484488)
Soundtrack	-0.0957885 (0.2980041)	0.0975882 (0.3017528)	0.3826121 (0.4526520)	0.3826866 (0.4525965)
Peak_Position	0.0120957*** (0.0017088)	0.0125964*** (0.0018647)	0.0159508*** (0.0021484)	0.0154869*** (0.0021473)
major_label	-0.2920023** (0.1045738)	-0.3717084*** (0.1123028)	-0.1722494 (0.1362552)	-0.1612792 (0.1362536)
avg_similarity_prev	-2.9302368* (1.1410881)	-2.9917775* (1.2425234)	-0.5069602 (1.4384552)	-0.4789470 (1.4386138)
dummy__399	0.1854267 (0.1520470)	0.1530215 (0.1595332)	0.2848369 (0.1945545)	0.2726943 (0.1944020)
dummy__400	-0.3499928* (0.1628140)	-0.3672652* (0.1723071)	-0.0868139 (0.2039934)	-0.0914421 (0.2040153)
dummy__401	-0.9090479*** (0.1715194)	-0.9426416*** (0.1839657)	-0.3221712 (0.2134210)	-0.3120907 (0.2136960)
dummy__402	-0.9665886*** (0.1711984)	-1.0656952*** (0.1871812)	0.0566612 (0.2218198)	0.0527414 (0.2217070)
dummy__403	-0.7950032** (0.2504296)	-0.7685930** (0.2670987)	0.1311763 (0.3355867)	0.1231070 (0.3352490)
AIC	3047.2116002		1439.8401576	
BIC	3126.4982006		1499.8222733	
Log Likelihood	-1511.6058001		-707.9200788	
Deviance	3023.2116002		1415.8401576	
Num. obs.	5471	5471	1095	1095

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 15: List of Major Record Labels and their Subsidiaries

Label Name
Universal Music Group
Sony Corporation of America
Warner Bros. Records
Capitol Records
Radio Corporation of America
Victor Red Seal
RCA Corporation
Warner Communications
Time Warner
Kinney National Company
EMI Records Group North America
MCA, Inc.
Syco Entertainment
Sony Music Latin
Sony Music
Mercury Record Corporation
Warner Bros. Pictures
BMG Rights Management
Fueled by Ramen
Polymedia Marketing Group GmbH
MCA Records
EMI Group
CBS
Capitol Records Ltd.
Island Records
Vagrant Records, LLC.
ARS Entertainment
Sparrow Label Group
Capitol-EMI Music, Inc.
Sub Pop Records
Irving Music, Inc.
EMI Music France
EMI Music Denmark
EMI Music Germany GmbH & Co. KG
CBS Corporation
Electric & Musical Industries Ltd.
Big Brother Recordings, Ltd.
Relentless Records Ltd.
American Recordings
Vale Music Spain, S.L.
DreamWorks Records
BMG Nederland BV
Monument Record Corp.
Warner Music Group Germany GmbH & Co. Holding OHG

Reunion Records, Inc.
Giant Records
Global Television Ltd.
The Rocket Record Company
EMI America Records, Inc.
The Echo Label Ltd.
Mute
Relativity Records
Rap-A-Lot Records
A&M Records, Inc.
CBS Records Canada Ltd.
Enigma Records
I.R.S. Records
Warner Bros.
EMI International
Blue Note Label Group
Windham Hill Records Inc.
De-Lite Recorded Sound Corp.
Ariola Eurodisc GmbH
Philadelphia International Records
Salsoul Records
PolyGram Nederland B.V.
T.K. Records
Big Life
Shelter Recording Company Inc.
Roulette Records
Sony Music UK
Kapp
EMI (Australia) Limited
Universal Music Publishing MGB Ltd.
Tooth & Nail Records
RCA Red Seal
PolyGram Group Distribution, Inc.
ABC-Paramount
The Compo Company Ltd.

Table 5: Variance in Euclidean Similarity Pre May 25th, 1991 Models

	Base Model	Robust Model	Base Model Comparison	Robust Model Comparison
(Intercept)	-2.2793726*** (0.3331526)	-2.1954678*** (0.3468924)	-1.0579271* (0.4412476)	-0.9164154* (0.4414336)
christmas_season	-0.0626024 (0.1108257)	-0.0575462 (0.1204445)	0.1251655 (0.1501125)	0.1453551 (0.1506401)
summer_season	-0.0452006 (0.0945409)	-0.0289308 (0.1023363)	-0.0806461 (0.1228954)	-0.0716669 (0.1237068)
Soundtrack	0.4266275 (0.3454152)	0.5494433 (0.3497818)	0.4610667 (0.5025669)	0.4182127 (0.5025539)
Peak_Position	0.0180335*** (0.0013219)	0.0168656*** (0.0014218)	0.0278688*** (0.0017353)	0.0269354*** (0.0017370)
major_label	-0.5901820*** (0.0801584)	-0.5751693*** (0.0860603)	-0.2919595** (0.1069065)	-0.3079696** (0.1073800)
var_similarity_prev	-10.7586849*** (3.1445470)	-9.8279990** (3.4207787)	-4.1368202 (4.1330536)	-4.0287156 (4.1596610)
dummy__392	-0.5488054 (0.3286987)	-0.5638148 (0.3422309)	-0.4841557 (0.4433751)	-0.5666474 (0.4438862)
dummy__393	-0.6547671* (0.3290600)	-0.6917405* (0.3431345)	-0.4838476 (0.4459031)	-0.5481553 (0.4463831)
dummy__394	-0.5975287 (0.3316622)	-0.7233045* (0.3474100)	-0.4584885 (0.4476726)	-0.5665200 (0.4483260)
dummy__395	-0.2728808 (0.3300612)	-0.3360233 (0.3440910)	-0.3009855 (0.4466340)	-0.3421157 (0.4468505)
dummy__396	-0.4273900 (0.3346213)	-0.4728306 (0.3491029)	-0.3333865 (0.4519598)	-0.4083162 (0.4523828)
dummy__397	-0.2540105 (0.3342812)	-0.2629238 (0.3480280)	-0.3298811 (0.4518854)	-0.3911841 (0.4522474)
dummy__398	0.1068693 (0.3683595)	-0.0031269 (0.3875309)	-0.0079416 (0.4924354)	-0.0768977 (0.4925837)
AIC	4896.8101661		2262.6918195	
BIC	4997.0397723		2340.5183550	
Log Likelihood	-2434.4050831		-1117.3459097	
Deviance	4868.8101661		2234.6918195	
Num. obs.	9502	9502	1918	1918

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 6: Variance in Euclidean Similarity Post May 25th, 1991 Models

	Base Model	Robust Model	Base Model Comparison	Robust Model Comparison
(Intercept)	-2.2102849*** (0.1749825)	-2.1631133*** (0.1854024)	-0.8512733*** (0.2216590)	-0.8272180*** (0.2215693)
christmas_season	-0.2220835 (0.1384449)	-0.2594511 (0.1525586)	0.0515682 (0.1825456)	0.0412922 (0.1824943)
summer_season	-0.0460762 (0.1174816)	-0.0701001 (0.1277453)	-0.1595056 (0.1484304)	-0.1551838 (0.1485237)
Soundtrack	-0.0778703 (0.2977333)	0.0810656 (0.3052658)	0.3790760 (0.4506025)	0.3790890 (0.4505032)
Peak_Position	0.0123807*** (0.0017034)	0.0128118*** (0.0018574)	0.0159881*** (0.0021408)	0.0155209*** (0.0021398)
major_label	-0.2942409** (0.1045663)	-0.3763843*** (0.1121989)	-0.1727604 (0.1362965)	-0.1627279 (0.1363002)
var_similarity_prev	-2.6686444 (2.2490631)	-2.1184146 (2.4276146)	-2.2628856 (2.7762121)	-2.2455561 (2.7815793)
dummy_399	0.1900729 (0.1520430)	0.1635825 (0.1594494)	0.2873116 (0.1945848)	0.2753872 (0.1944463)
dummy_400	-0.3278427* (0.1631033)	-0.3514172* (0.1728371)	-0.0702468 (0.2043761)	-0.0718597 (0.2043720)
dummy_401	-0.8915938*** (0.1718956)	-0.9275308*** (0.1844375)	-0.3104087 (0.2139529)	-0.2999208 (0.2142455)
dummy_402	-1.0188161*** (0.1705371)	-1.1033696*** (0.1859605)	0.0399307 (0.2213474)	0.0360358 (0.2212681)
dummy_403	-0.9241848*** (0.2467333)	-0.8727823*** (0.2609349)	0.0995511 (0.3312718)	0.0910589 (0.3309564)
AIC	3052.2789935		1439.2927870	
BIC	3131.5655938		1499.2749027	
Log Likelihood	-1514.1394967		-707.6463935	
Deviance	3028.2789935		1415.2927870	
Num. obs.	5471	5471	1095	1095

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 7: Average Cosine Similarity Pre May 25th, 1991 Models

	Base Model	Robust Model	Base Model Comparison	Robust Model Comparison
(Intercept)	-3.6212100*** (0.7471365)	-3.5411713*** (0.8103569)	-3.0984750** (0.9752396)	-2.9403390** (0.9832783)
christmas_season	-0.0552311 (0.1107666)	-0.0433387 (0.1200939)	0.1373347 (0.1505192)	0.1558190 (0.1511349)
summer_season	-0.0493760 (0.0945209)	-0.0394901 (0.1025282)	-0.0720463 (0.1230224)	-0.0672614 (0.1239012)
Soundtrack	0.4365266 (0.3455386)	0.5662005 (0.3491451)	0.4708571 (0.5075982)	0.4290882 (0.5084584)
Peak_Position	0.0180072*** (0.0013213)	0.0168618*** (0.0014212)	0.0277679*** (0.0017363)	0.0268763*** (0.0017390)
major_label	-0.5855271*** (0.0802417)	-0.5745134*** (0.0861495)	-0.2764523** (0.1072094)	-0.2906424** (0.1077210)
avg_similarity_prev	1.4383286 (0.8545369)	1.4101708 (0.9324977)	2.5387841* (1.1123716)	2.5251082* (1.1237033)
dummy__392	-0.5487510 (0.3283165)	-0.5239217 (0.3470924)	-0.5098713 (0.4419039)	-0.5966154 (0.4423957)
dummy__393	-0.6561553* (0.3290317)	-0.6487664 (0.3483461)	-0.4693552 (0.4441162)	-0.5410896 (0.4445780)
dummy__394	-0.6198249 (0.3313026)	-0.7029625* (0.3522963)	-0.4728195 (0.4452225)	-0.5855146 (0.4458768)
dummy__395	-0.2947666 (0.3295764)	-0.3033634 (0.3486946)	-0.3251757 (0.4441016)	-0.3723956 (0.4443052)
dummy__396	-0.4372486 (0.3343242)	-0.4329087 (0.3539392)	-0.3497715 (0.4498333)	-0.4287485 (0.4502209)
dummy__397	-0.3008344 (0.3335766)	-0.2527924 (0.3523085)	-0.3633560 (0.4487311)	-0.4318797 (0.4490812)
dummy__398	0.0965456 (0.3681732)	0.0366978 (0.3918829)	-0.0108072 (0.4912652)	-0.0859073 (0.4915271)
AIC	4906.1776320		2258.4154528	
BIC	5006.4072382		2336.2419884	
Log Likelihood	-2439.0888160		-1115.2077264	
Deviance	4878.1776320		2230.4154528	
Num. obs.	9502	9502	1918	1918

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 8: Average Cosine Similarity Post May 25th, 1991 Models

	Base Model	Robust Model	Base Model Comparison	Robust Model Comparison
(Intercept)	-1.6326282 (0.8456334)	-1.7636656 (0.9191603)	-1.4955761 (1.0890121)	-1.5438980 (1.0903369)
christmas_season	-0.2229530 (0.1384212)	-0.2633823 (0.1527327)	0.0491247 (0.1823921)	0.0384733 (0.1823596)
summer_season	-0.0431324 (0.1175146)	-0.0655028 (0.1277356)	-0.1608419 (0.1485187)	-0.1561504 (0.1486435)
Soundtrack	-0.0719718 (0.2977044)	0.1001723 (0.3035544)	0.4160054 (0.4518922)	0.4168334 (0.4517199)
Peak_Position	0.0123613*** (0.0017080)	0.0128856*** (0.0018623)	0.0161453*** (0.0021539)	0.0157070*** (0.0021537)
major_label	-0.2989758** (0.1045271)	-0.3856314*** (0.1121125)	-0.1730893 (0.1362974)	-0.1632330 (0.1363282)
avg_similarity_prev	-0.8500225 (1.0951656)	-0.5972943 (1.1914887)	0.7692300 (1.3989325)	0.8642635 (1.4004196)
dummy__399	0.1870649 (0.1519870)	0.1587857 (0.1594589)	0.2868127 (0.1945612)	0.2750103 (0.1944297)
dummy__400	-0.3520787* (0.1631609)	-0.3719721* (0.1730169)	-0.0686730 (0.2054612)	-0.0714679 (0.2054948)
dummy__401	-0.9179982*** (0.1718618)	-0.9491902*** (0.1844442)	-0.3096369 (0.2145710)	-0.2979059 (0.2148779)
dummy__402	-1.0003625*** (0.1707526)	-1.0912676*** (0.1861721)	0.0488091 (0.2209658)	0.0427525 (0.2209059)
dummy__403	-0.8815064*** (0.2493000)	-0.8415112** (0.2637155)	0.0890596 (0.3338699)	0.0730085 (0.3336892)
AIC	3053.1602812		1439.6615133	
BIC	3132.4468815		1499.6436290	
Log Likelihood	-1514.5801406		-707.8307566	
Deviance	3029.1602812		1415.6615133	
Num. obs.	5471	5471	1095	1095

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 9: Variance in Cosine Similarity Pre May 25th, 1991 Models

	Base Model	Robust Model	Base Model Comparison	Robust Model Comparison
(Intercept)	-2.3155252*** (0.3355506)	-2.2376486*** (0.3504516)	-0.9985091* (0.4434603)	-0.8620743 (0.4434563)
christmas_season	-0.0560948 (0.1107767)	-0.0502121 (0.1203145)	0.1366326 (0.1503444)	0.1528753 (0.1508702)
summer_season	-0.0465697 (0.0945137)	-0.0330163 (0.1023910)	-0.0785311 (0.1229459)	-0.0705168 (0.1237349)
Soundtrack	0.4354065 (0.3454558)	0.5603140 (0.3497992)	0.4650382 (0.5042512)	0.4239896 (0.5044004)
Peak_Position	0.0180223*** (0.0013208)	0.0168489*** (0.0014203)	0.0278392*** (0.0017355)	0.0268990*** (0.0017369)
major_label	-0.5849730*** (0.0802026)	-0.5747288*** (0.0860785)	-0.2840578** (0.1070354)	-0.2992555** (0.1074857)
var_similarity_prev	-24.7955299* (10.6068038)	-22.4464504 (11.5308838)	-22.2706967 (14.1020444)	-20.5398381 (14.1922370)
dummy__392	-0.5382789 (0.3282988)	-0.5417528 (0.3426496)	-0.4611376 (0.4435180)	-0.5467498 (0.4438034)
dummy__393	-0.6314233 (0.3292990)	-0.6618525 (0.3443170)	-0.4299633 (0.4478499)	-0.5048321 (0.4481391)
dummy__394	-0.5870385 (0.3317671)	-0.7061356* (0.3484720)	-0.4107224 (0.4491206)	-0.5251390 (0.4495399)
dummy__395	-0.2653143 (0.3299179)	-0.3181810 (0.3448392)	-0.2518689 (0.4481468)	-0.3020586 (0.4481694)
dummy__396	-0.4072837 (0.3347139)	-0.4464443 (0.3501764)	-0.2805759 (0.4537777)	-0.3627097 (0.4540024)
dummy__397	-0.2581564 (0.3342613)	-0.2562038 (0.3488983)	-0.2835325 (0.4530450)	-0.3508382 (0.4531914)
dummy__398	0.1325170 (0.3687205)	0.0327790 (0.3886159)	0.0493878 (0.4948610)	-0.0257267 (0.4948200)
AIC	4903.5305436		2261.1921236	
BIC	5003.7601498		2339.0186591	
Log Likelihood	-2437.7652718		-1116.5960618	
Deviance	4875.5305436		2233.1921236	
Num. obs.	9502	9502	1918	1918

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 10: Variance in Cosine Similarity Post May 25th, 1991 Models

	Base Model	Robust Model	Base Model Comparison	Robust Model Comparison
(Intercept)	-2.1352792*** (0.1979579)	-2.0822001*** (0.2109382)	-0.6037964* (0.2530221)	-0.5668711* (0.2530674)
christmas_season	-0.2224402 (0.1384482)	-0.2594520 (0.1526148)	0.0549187 (0.1827226)	0.0488439 (0.1828030)
summer_season	-0.0481404 (0.1175012)	-0.0711980 (0.1277603)	-0.1641596 (0.1487604)	-0.1605466 (0.1490391)
Soundtrack	-0.0696570 (0.2976894)	0.0837184 (0.3057979)	0.3831493 (0.4498809)	0.3758174 (0.4498172)
Peak_Position	0.0125000*** (0.0017028)	0.0128841*** (0.0018567)	0.0163171*** (0.0021523)	0.0159325*** (0.0021550)
major_label	-0.2940694** (0.1045802)	-0.3757910*** (0.1122382)	-0.1716104 (0.1366742)	-0.1689126 (0.1368075)
var_similarity_prev	-13.6665880 (10.5220973)	-12.8323849 (11.4288604)	-29.2386704* (13.6057963)	-30.4011015* (13.6547682)
dummy_399	0.1824921 (0.1520252)	0.1584232 (0.1594388)	0.2712839 (0.1949580)	0.2603611 (0.1949624)
dummy_400	-0.3227351* (0.1633588)	-0.3431270* (0.1730934)	-0.0471115 (0.2045954)	-0.0435789 (0.2047077)
dummy_401	-0.8882273*** (0.1720335)	-0.9227046*** (0.1846640)	-0.2866191 (0.2145939)	-0.2795031 (0.2151746)
dummy_402	-1.0222433*** (0.1706545)	-1.1093984*** (0.1862047)	0.0242404 (0.2215173)	0.0147554 (0.2216397)
dummy_403	-0.9437246*** (0.2477923)	-0.8897249*** (0.2618791)	0.0304929 (0.3336214)	0.0101406 (0.3336543)
AIC	3052.0512221		1435.3013134	
BIC	3131.3378225		1495.2834291	
Log Likelihood	-1514.0256111		-705.6506567	
Deviance	3028.0512221		1411.3013134	
Num. obs.	5471	5471	1095	1095

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 11: Regression Discontinuity Models Average Euclidean Similarity

	Optimal Bandwidth	No Bandwidth	Optimal Bandwidth Comparison	No Bandwidth Comparison
(Intercept)	-2.2170654** (0.8357702)	-0.1935349 (0.9894321)	0.3912861 (1.2426968)	1.8346295 (1.2791797)
D	0.0296096 (0.2311766)	0.0271624 (0.2309365)	-0.1448041 (0.2950849)	-0.1340000 (0.2957022)
x	0.0019714** (0.0006770)	0.0008626 (0.0005398)	0.0024564** (0.0007549)	0.0019060** (0.0006986)
x_right	-0.0027557** (0.0009789)	-0.0013511 (0.0008818)	-0.0038349** (0.0011916)	-0.0033160** (0.0011587)
Peak_Position	0.0158837*** (0.0011820)	0.0160687*** (0.0010430)	0.0229027*** (0.0013738)	0.0234877*** (0.0013439)
avg_similarity_prev	-0.8602910 (0.6220336)	-1.2584771* (0.5287991)	0.6360245 (0.7181538)	0.5796468 (0.6882823)
major_label	-0.3071002*** (0.0729200)	-0.4895217*** (0.0633452)	-0.2252380** (0.0859028)	-0.2520416** (0.0836818)
Soundtrack	0.0557788 (0.2414042)	0.0741561 (0.2256565)	0.3844102 (0.3385970)	0.3804023 (0.3393787)
christmas_season	-0.2732809** (0.1012743)	-0.1340446 (0.0872356)	0.0762490 (0.1202107)	0.0948704 (0.1165964)
summer_season	-0.1025343 (0.0829825)	-0.0434211 (0.0736481)	-0.1126418 (0.0973418)	-0.0969885 (0.0945436)
dummy__393	1.9499459 (1.0885300)	-1.0522292** (0.3933891)	0.8960442 (1.5005519)	-1.2243646* (0.5157163)
dummy__394	1.5762804 (0.9964450)	-1.2278228* (0.4877464)	0.2244752 (1.3874600)	-1.7438154** (0.6348773)
dummy__395	1.3671445 (0.8924176)	-1.1420280 (0.5976184)	-0.2597144 (1.2930489)	-2.0822698** (0.7804907)
dummy__396	0.6435366 (0.8131231)	-1.5558397* (0.7263273)	-0.9498894 (1.2261108)	-2.6224213** (0.9442786)
dummy__397	0.2892973 (0.7698872)	-1.6315192 (0.8484815)	-1.5613548 (1.1903041)	-3.0980969** (1.0993984)
dummy__398	0.3246054 (0.7295127)	-1.4245473 (0.9453700)	-1.6942126 (1.1491617)	-3.1304802* (1.2246406)
dummy__399	0.6930541 (0.5686298)	-1.1227497 (0.9725863)	-1.0714514 (0.9522925)	-2.4981611* (1.2594997)
dummy__400	0.3648453 (0.3997519)	-1.5245348 (1.0223499)	-1.0964267 (0.7410775)	-2.5153290 (1.3216804)
dummy__401	-0.0259720 (0.2483538)	-2.0243674 (1.1012425)	-1.0182298 (0.5362464)	-2.4353865 (1.4279283)
dummy__392		-0.6895382* (0.3377908)	1.3040732 (1.6026608)	-0.7558587 (0.4450098)
dummy__402		-1.9924911 (1.1947418)	-0.2665275 (0.3857491)	-1.6727777 (1.5670601)
dummy__403		-1.7241863 (1.2905391)		-1.3973728 (1.6904841)
AIC	6278.2126043	7959.8894953	3512.2109631	3704.6010849
BIC	6422.8786776	8127.3975802	3638.4354854	3836.8362988
Log Likelihood	-3120.1063022	-3957.9447477	-1735.1054815	-1830.3005425
Deviance	6240.2126043	7915.8894953	3470.2109631	3660.6010849
Num. obs.	11668	14973	2849	3013

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 12: Regression Discontinuity Models Variance in Euclidean Similarity

	Optimal Bandwidth	No Bandwidth	Optimal Bandwidth Comparison	No Bandwidth Comparison
(Intercept)	-2.7223385*** (0.7552128)	-0.7356028 (0.9498276)	0.7757676 (1.1724685)	2.1834193 (1.2307514)
D	0.0433827 (0.2309728)	0.0467572 (0.2307564)	-0.1422565 (0.2951317)	-0.1311909 (0.2957574)
x	0.0020358** (0.0006760)	0.0009599 (0.0005394)	0.0024320** (0.0007534)	0.0018785** (0.0006971)
x_right	-0.0027550** (0.0009785)	-0.0013791 (0.0008817)	-0.0037698** (0.0011893)	-0.0032488** (0.0011565)
Peak_Position	0.0158441*** (0.0011826)	0.0160167*** (0.0010434)	0.0228714*** (0.0013735)	0.0234579*** (0.0013437)
var_similarity_prev	-3.6653826* (1.7389805)	-4.5603812** (1.7080868)	-2.5189300 (2.0932976)	-2.3958982 (2.0754819)
major_label	-0.3014571*** (0.0729301)	-0.4822409*** (0.0633475)	-0.2259510** (0.0858983)	-0.2529445** (0.0836732)
Soundtrack	0.0502391 (0.2413838)	0.0689972 (0.2256378)	0.3578346 (0.3379794)	0.3553747 (0.3387926)
christmas_season	-0.2750470** (0.1013002)	-0.1357753 (0.0872489)	0.0789910 (0.1201971)	0.0955153 (0.1165818)
summer_season	-0.1043502 (0.0829955)	-0.0460787 (0.0736489)	-0.1166295 (0.0973403)	-0.1004316 (0.0945486)
dummy__393	2.0937249 (1.0900880)	-1.0579550** (0.3933015)	0.9312857 (1.5011225)	-1.1946124* (0.5162940)
dummy__394	1.6946757 (0.9979137)	-1.2671366** (0.4878567)	0.2765731 (1.3882912)	-1.6971437** (0.6344750)
dummy__395	1.4726790 (0.8942568)	-1.2011014* (0.5973131)	-0.2081854 (1.2937859)	-2.0346862** (0.7798961)
dummy__396	0.7400457 (0.8148987)	-1.6281748* (0.7264105)	-0.9002153 (1.2266641)	-2.5752450** (0.9434507)
dummy__397	0.3857642 (0.7720174)	-1.7083693* (0.8482215)	-1.5035679 (1.1908935)	-3.0419317** (1.0983544)
dummy__398	0.4159993 (0.7314147)	-1.5059562 (0.9455675)	-1.6550527 (1.1495300)	-3.0907613* (1.2239091)
dummy__399	0.7774871 (0.5704570)	-1.2104735 (0.9729237)	-1.0394085 (0.9527021)	-2.4653296 (1.2590697)
dummy__400	0.4481998 (0.4022707)	-1.6099649 (1.0225171)	-1.0614720 (0.7418961)	-2.4798526 (1.3214898)
dummy__401	0.0434631 (0.2490003)	-2.1259914 (1.1027026)	-0.9929636 (0.5365842)	-2.4095142 (1.4279987)
dummy__392		-0.6977631* (0.3375797)	1.3308306 (1.6029464)	-0.7322340 (0.4456790)
dummy__402		-2.1780275 (1.1963338)	-0.2740724 (0.3851904)	-1.6786421 (1.5671832)
dummy__403		-1.9625880 (1.2921516)		-1.3970373 (1.6902900)
AIC	6275.3813936	7957.8424887	3511.5301044	3703.9632485
BIC	6420.0474669	8125.3505736	3637.7546268	3836.1984624
Log Likelihood	-3118.6906968	-3956.9212444	-1734.7650522	-1829.9816243
Deviance	6237.3813936	7913.8424887	3469.5301044	3659.9632485
Num. obs.	11668	14973	2849	3013

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 13: Regression Discontinuity Models Average Cosine Similarity

	Optimal Bandwidth	No Bandwidth	Optimal Bandwidth Comparison	No Bandwidth Comparison
(Intercept)	-3.9540609*** (0.9730233)	-1.4299427 (1.0707844)	-1.0580230 (1.3704129)	0.5160825 (1.3901376)
D	0.0641290 (0.2312468)	0.0579950 (0.2308756)	-0.1222212 (0.2959911)	-0.1122331 (0.2965457)
x	0.0020409** (0.0006758)	0.0009587 (0.0005397)	0.0024595** (0.0007551)	0.0019153** (0.0006988)
x_right	-0.0028068** (0.0009776)	-0.0014604 (0.0008807)	-0.0038248** (0.0011899)	-0.0033143** (0.0011569)
Peak_Position	0.0159222*** (0.0011822)	0.0160831*** (0.0010429)	0.0229821*** (0.0013763)	0.0235541*** (0.0013460)
avg_similarity_prev	1.5440474* (0.7638496)	0.7585476 (0.6688241)	2.2765292* (0.8910448)	2.1628376* (0.8630741)
major_label	-0.2974006*** (0.0729973)	-0.4821271*** (0.0633898)	-0.2181191* (0.0860506)	-0.2448031** (0.0838247)
Soundtrack	0.0634751 (0.2414812)	0.0838682 (0.2256644)	0.4090262 (0.3392373)	0.4042996 (0.3399199)
christmas_season	-0.2719047** (0.1012957)	-0.1329718 (0.0872165)	0.0812753 (0.1203411)	0.1012385 (0.1167333)
summer_season	-0.1059763 (0.0830102)	-0.0470031 (0.0736489)	-0.1139990 (0.0974283)	-0.0985857 (0.0946219)
dummy__393	2.0599463 (1.0892016)	-1.0554193** (0.3934247)	1.0020948 (1.5020642)	-1.1890522* (0.5164783)
dummy__394	1.6453628 (0.9965144)	-1.2740007** (0.4879467)	0.3181952 (1.3884779)	-1.7223750** (0.6351382)
dummy__395	1.4177213 (0.8930376)	-1.2073297* (0.5974158)	-0.1846333 (1.2933860)	-2.0797919** (0.7812581)
dummy__396	0.6930594 (0.8137328)	-1.6280326* (0.7265426)	-0.8775405 (1.2261861)	-2.6241170** (0.9450898)
dummy__397	0.3196834 (0.7705081)	-1.7254223* (0.8484666)	-1.4997907 (1.1900464)	-3.1114850** (1.1005220)
dummy__398	0.3705857 (0.7302550)	-1.5051712 (0.9457272)	-1.6318306 (1.1487183)	-3.1436890* (1.2264504)
dummy__399	0.7307755 (0.5691191)	-1.2026705 (0.9730636)	-1.0123460 (0.9522322)	-2.5141749* (1.2614470)
dummy__400	0.4181377 (0.4010308)	-1.5893846 (1.0226141)	-1.0041314 (0.7419987)	-2.4992311 (1.3234396)
dummy__401	0.0246990 (0.2489740)	-2.0851459 (1.1016807)	-0.9319289 (0.5374818)	-2.4245908 (1.4295762)
dummy__392		-0.6965616* (0.3375273)	1.3749213 (1.6041591)	-0.7540253 (0.4455962)
dummy__402		-2.0832698 (1.1953356)	-0.2182811 (0.3866600)	-1.6980934 (1.5683849)
dummy__403		-1.8649150 (1.2915269)		-1.4688164 (1.6918580)
AIC	6275.9524135	7964.2227199	3506.3995880	3698.9684802
BIC	6420.6184868	8131.7308048	3632.6241103	3831.2036941
Log Likelihood	-3118.9762067	-3960.1113600	-1732.1997940	-1827.4842401
Deviance	6237.9524135	7920.2227199	3464.3995880	3654.9684802
Num. obs.	11668	14973	2849	3013

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 14: Regression Discontinuity Models Variance in Cosine Similarity

	Optimal Bandwidth	No Bandwidth	Optimal Bandwidth Comparison	No Bandwidth Comparison
(Intercept)	-2.6460986*** (0.7563700)	-0.6150870 (0.9534709)	0.8873151 (1.1732426)	2.4041152 (1.2363754)
D	0.0577801 (0.2310722)	0.0635883 (0.2308160)	-0.1061689 (0.2964665)	-0.0934514 (0.2971136)
x	0.0020515** (0.0006761)	0.0010055 (0.0005402)	0.0024905*** (0.0007556)	0.0019460** (0.0006993)
x_right	-0.0027786** (0.0009783)	-0.0014420 (0.0008816)	-0.0038139** (0.0011904)	-0.0033006** (0.0011574)
Peak_Position	0.0159012*** (0.0011822)	0.0160747*** (0.0010430)	0.0230002*** (0.0013768)	0.0235699*** (0.0013466)
var_similarity_prev	-16.6356130* (7.8342600)	-19.8041197** (7.3252767)	-27.2603943** (9.6619153)	-27.9028520** (9.5760898)
major_label	-0.2987008*** (0.0729691)	-0.4789721*** (0.0633840)	-0.2190748* (0.0860418)	-0.2463020** (0.0838113)
Soundtrack	0.0594539 (0.2413589)	0.0795186 (0.2256378)	0.3638198 (0.3382021)	0.3605329 (0.3389332)
christmas_season	-0.2732018** (0.1012957)	-0.1339598 (0.0872432)	0.0852443 (0.1203406)	0.1033266 (0.1167336)
summer_season	-0.1054920 (0.0830022)	-0.0472520 (0.0736522)	-0.1180874 (0.0974675)	-0.1020126 (0.0946793)
dummy__393	2.1076700 (1.0907449)	-1.0456959** (0.3933351)	1.0819149 (1.5036631)	-1.1322078* (0.5175941)
dummy__394	1.6960751 (0.9980941)	-1.2730141** (0.4881737)	0.4085318 (1.3900085)	-1.6557487** (0.6360059)
dummy__395	1.4659376 (0.8944703)	-1.2236484* (0.5975999)	-0.0916110 (1.2949137)	-2.0109565* (0.7817346)
dummy__396	0.7360984 (0.8151711)	-1.6546528* (0.7269369)	-0.7972213 (1.2274341)	-2.5676922** (0.9455439)
dummy__397	0.3711521 (0.7720840)	-1.7548749* (0.8487037)	-1.4152224 (1.1912178)	-3.0508595** (1.1007849)
dummy__398	0.4108994 (0.7317248)	-1.5448873 (0.9461815)	-1.5713538 (1.1496179)	-3.1066141* (1.2270038)
dummy__399	0.7619553 (0.5703796)	-1.2603737 (0.9736728)	-0.9771119 (0.9529828)	-2.5031888* (1.2621270)
dummy__400	0.4471096 (0.4025424)	-1.6409001 (1.0231283)	-0.9689244 (0.7427858)	-2.4859166 (1.3241337)
dummy__401	0.0432013 (0.2492785)	-2.1536618 (1.1033268)	-0.9041850 (0.5374469)	-2.4199018 (1.4307379)
dummy__392		-0.6946595* (0.3374213)	1.4536623 (1.6055496)	-0.6989048 (0.4463743)
dummy__402		-2.2024910 (1.1973106)	-0.2225962 (0.3858339)	-1.7294735 (1.5694634)
dummy__403		-2.0073687 (1.2937033)		-1.5008375 (1.6928029)
AIC	6275.5553021	7958.1070464	3504.9727211	3696.7528814
BIC	6420.2213754	8125.6151313	3631.1972434	3828.9880953
Log Likelihood	-3118.7776510	-3957.0535232	-1731.4863606	-1826.3764407
Deviance	6237.5553021	7914.1070464	3462.9727211	3652.7528814
Num. obs.	11668	14973	2849	3013

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Figure 2: Histogram of Max Weeks on The Hot 100

Figure 3: Histogram of Debut Rank on The Hot 100

Figure 4: Histogram of Max Rank on The Hot 100

Affiliation:

Jacob Derechin
Yale University
New Haven, CT
E-mail: jacob.derechin@yale.edu

SocArXiv Website
SocArXiv Preprints

Preprint
[URL/DOI GOES HERE](#)

<https://socopen.org/>
<https://osf.io/preprints/socarxiv>

Submitted: October 30, 2019
Accepted: October 30, 2019