The Effect of Information on the Charts:

Evidence from Billboard

Teni Adedeji

Advisor: Julia Lowell

March 2018

Special Thanks to Dr. Shelly Lundberg, Brendan Uyeshiro, and Ayo Adedeji

Abstract

With the introduction of digital intermediaries allowing people to access songs and artists more easily, there has been a shift in how listeners consume music. In this paper, I examine the effects of introducing Spotify in two countries – United States and Canada – on music diversity, using a Difference-in-Differences approach. In addition, I use a linear regression to analyze the change in the number of genres that have shown up in the Billboard Hot 100 since the beginning of the 21st century. Both these findings suggest that there is a negative effect on music diversity – defined by the number of *Unique Songs* and *Unique Genres* that show up on the Billboard Hot 100 charts – when there is more digital technology present. The effects on *Unique Songs* were negative at an insignificant level while the effects of *Unique Genres* were negative at a significant level.

A. Introduction

On November 2, 2017, Spotify, a popular music streaming service, released an article announcing that listening diversity, the number of unique artists each user streams per week, has increased nearly 40% on their platform since 2014 (Erlandsson and Perez, 2017). The company attributed most of this listening diversity to the recommended playlists and other programs they provide that encourage listeners to consume new songs and discover new artists. On the other hand, MIDIA, a media and technology analysis company, released a study a couple years ago that the top 1% of artists accounted for 77% of sales revenue in 2013 (Mulligan, 2014).

The statistics tell two different stories. One shines a light on the fact that the music industry has always been a Superstar economy with big gains for the winners. The other paints a more positive picture presenting the democratization of the music industry due to digital intermediaries.

This democratization comes from both the supply and demand side of the industry. Traditionally, artists have relied heavily on big record companies to help fund, manage, and market the production of their art. Now, with digital applications – such as, Spotify, Bandcamp, and SoundCloud – more independent artists are becoming successful without the help from traditional intermediaries (Music Business Worldwide, 2015). The ability to record a song, edit, produce, and upload to a global network can all be done from the comfort of a musician's bedroom. In the same regard, the way consumers discover and listen to music has been impacted through digital applications – increasing the capability to search and find songs right under their fingertips as opposed to relying heavily on radio stations and promotional services.

This paper aims to answer how the accessibility of information over time has impacted music diversity and consumer habits by examining the number of unique songs and the different genres of artists that appear on the Billboard Hot 100 charts. Sjolander (2016) and others have also examined the number of unique songs on the Billboard charts; although, this paper differentiates itself by looking specifically at how the number of unique songs differed between two different countries after the introduction of Spotify. In addition, this paper contributes and fills a gap in the literature by investigating the change in music diversity over time through genres.

It is important to understand how digital technologies impact the consumption of music for companies looking to create marketing and business strategies, potential investments, and artist contracts. In addition, being able to understand how the audience consumes music also allows us to address whether consumers are maximizing their utility by listening to what would better suit their specific tastes or if they are being incentivized to listen to general popular songs.

B. Literature Review

Most of the literature centers around two main theories when examining how the rise of digital technologies impacts the diversity in the music industry – the Superstar Effect and the Long Tail. Adler (1985) puts forth a model arguing that stardom is a phenomenon that exists where consumption requires knowledge. He explores a theory of "consumption capital" – first developed by George Stigler and Gary Becker (1977) – suggesting that people appreciate music more as they gain knowledge of the subject and as they discuss amongst other individuals. Therefore, consumers seek to minimize their search costs and maximize their shared knowledge by choosing the most popular artists. As a result, there is a group of musicians chosen to be the "stars" so that everyone can have knowledge of the same artists and make discussion possible.

In relation to this, another explanation for the Superstar Effect spoken about in the literature is that because online technologies make it easier for people to create and share songs, consumers are now overwhelmed with a paradox of too many choices. Due to the vast quantity of songs, listeners are better off choosing popular songs to lower search costs and increase the chance of good quality products much like the typical Bayesian signaling game. Lao and Nguyen (2016) found evidence that supported this theory after looking at the effect of the digital technology format on popularity characteristics. They found that there was a "strengthening of the bandwagon effect in the digital era" where established artists were crowding out the others due to their reputation. Their model observed this by looking at the breakout effect, comparing new artists to established artists. The result showed that the breakout effect diminished as music transitioned into the digital era – with a multitude of songs easily accessible – potentially displaying how consumers became drawn to artists that they already knew rather than picking new artists at random.

In opposition, the Long Tail theory suggests that instead of concentrating the demand on a select few, online technologies allow individuals to spread out their consumption across goods. Anderson (2006) examined this effect on different markets – for example, bookstores versus the online book market – and came out with the conclusion that the decreased search costs allowed consumers the chance to receive the right information appropriate for their specific tastes. The ease of the Internet and digital intermediaries allowed people to lower their search costs and discover new distinctive artists. This idea was supported further after Brynjolsson et al. (2011) provided evidence that sales made on the Internet were significantly less concentrated compared to physical sales even with supply-side factors held constant.

Additionally, the Internet allows for more select products to be available that can match consumers niche interests. Brynjolsson et al. (2003) looked at the Long Tail theory when they examined the difference in the proportion of niche products purchased on the Internet compared to more traditional channels. Their results found that the selection of products purchased was higher on the Internet, highlighting that technology's capacity to have a high amount of product information feeds into the Long Tail theory that lower search costs allows the demand for niche products to increase.

While a lot of the past literature explains the effect of the Internet and increasing information, few specifically isolate the effect of an online intermediary, such as Spotify, on the consumption of music. Hiller and Kim (2016) analyzed YouTube and how the availability of online music content impacted music sales and searches for artists. Datta, Knox, and Bronnenberg (2017) examined how the adoption of Spotify affects consumer listening habits by constructing a panel dataset of certain individuals and the time they spent listening to music as

well as other musical preferences. They provide evidence that Spotify increased consumer welfare by decreasing search frictions, and helped users discover new content.

Regarding past studies that have examined chart performance, there has been more support for the Superstar Effect than for the Long Tail theory. Bradlow & Fader (2001) estimated the expected lifetime for songs on the chart with results suggesting that popular artists had an extended survival on the charts. Bhattacharjee, et al. (2007) examined the impact of file-sharing on album survival and found that file-sharing had a negative impact on low-ranked albums and that top-ranked albums experienced no harm. Giles (2007) released another study on chart survival from the Billboard chart and found similar results to the first two – which was that the "stars" on average remained on the charts for a longer period. Lastly, Sjoldander (2016) looked at how the number of unique songs and the number of unique artists per year was changing over time. He found that the number of artists declined over time which again helps constitute towards the Superstar Effect.

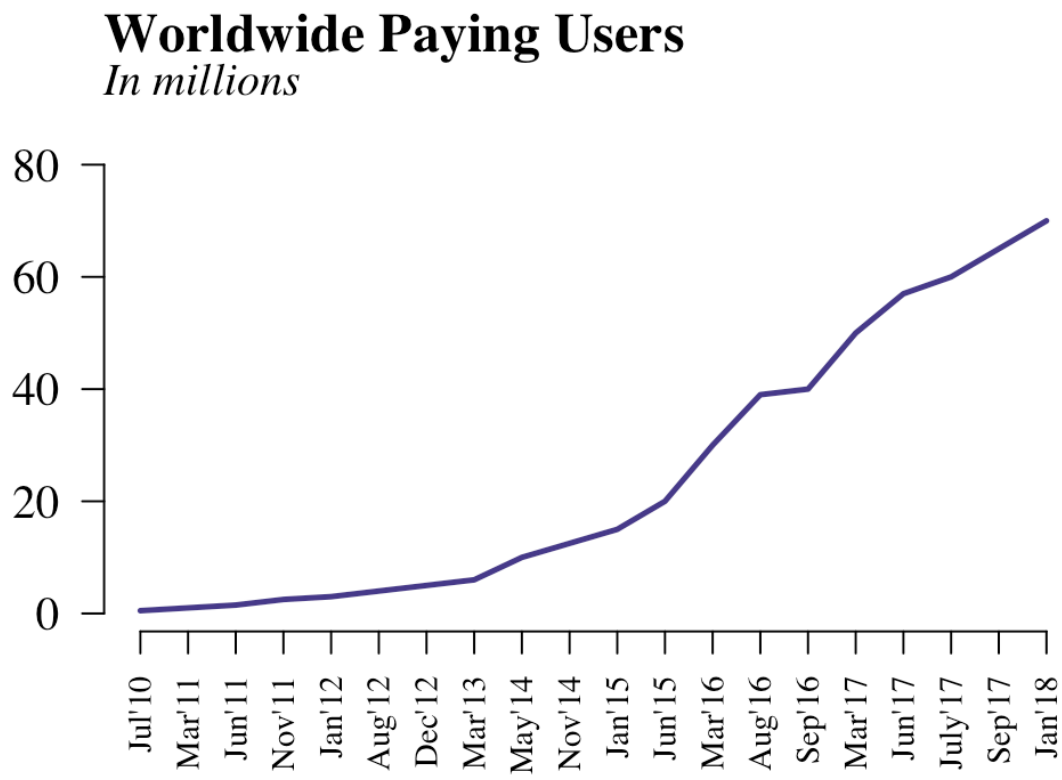
It is important to note that while most of the past literature is in support for the Superstar Effect, these results do not entirely discredit the Long Tail theory. There is a probable chance that both effects are taking place at the same time. This paper does not separate the two theories from each other, but aims to add another way in examining which force is currently prevailing through a different approach than past literature.

C. Empirical Strategy

While most of the literature focuses on the sales displacement effect of online intermediaries, the main outcome variable of this paper is *Unique Songs* – the number of different songs that showed up on the Billboard Hot 100 charts in a period. Another supplementary variable that is analyzed throughout this paper is *Unique Genres* – the number of different genres that appeared

on the Billboard Hot 100 chart in a year. This variable provides support for the main dependent variable and helps paint more of a picture on how the trend of music diversity has unraveled over time.

In efforts to estimate an increased accessibility of information, I use Spotify as a proxy of information. Spotify is a prime example of an online music intermediary that increased consumer's information and lowered the search costs for exploring different songs. It quickly emerged as the best quality music streaming service in the market and has continually held its position as the service with the biggest market share to date. Graph 1 displays the global user growth on the platform since July 2010.



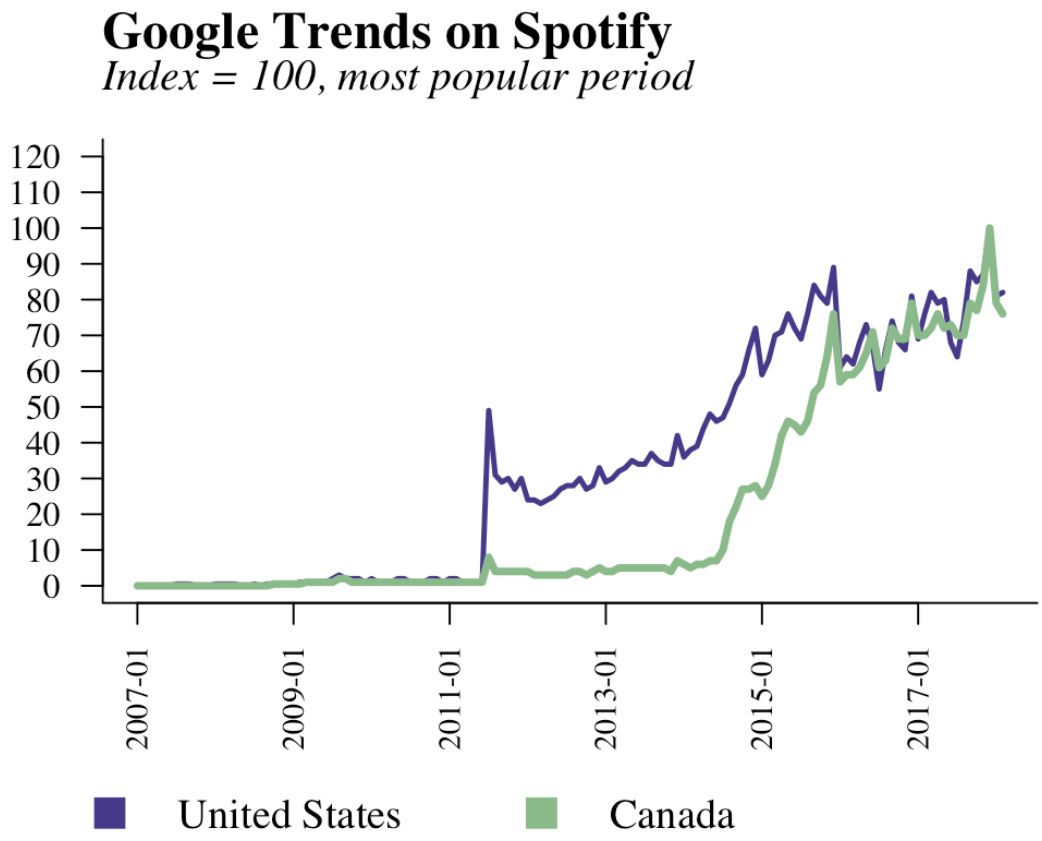
Graph 1 – Number of paying Spotify subscribers (2010-2018). Source: Statistica

Although this graph only shows the paid subscribers of Spotify, adding the number of users that have a free membership would provide additional support for Spotify's popularity. Initially,

Spotify was released on October 7, 2008 in Scandinavia, the United Kingdom, France and Spain by invitation only. It was not until mid-July 2011 that Spotify officially came to the United States and it took a few more years until it officially launched in Canada in late-September of 2014.

My methodology relies on the fact that Spotify was introduced to each country at different times in order estimate the effect of Spotify on *Unique Songs* using the Difference-in-Differences (D-in-D) approach. Because of this, it is important to consider the fact that some people in Canada may have had access to the United States version of Spotify before it officially launched in 2014; consequentially, since the two countries are in such proximity. After researching, there was no indication that people in Canada could easily access the U.S. Spotify at the time without downloading some additional hack software that changed their Virtual Private Network (VPN). On top of this, Spotify also had a geo-blocking software that would disable the app if the user was in another country. While there was still a few people that were using geo-locking services or using U.S. credit card detail to trick Spotify into thinking they were in the United States, there is no reason to believe that a material amount of the population did this.

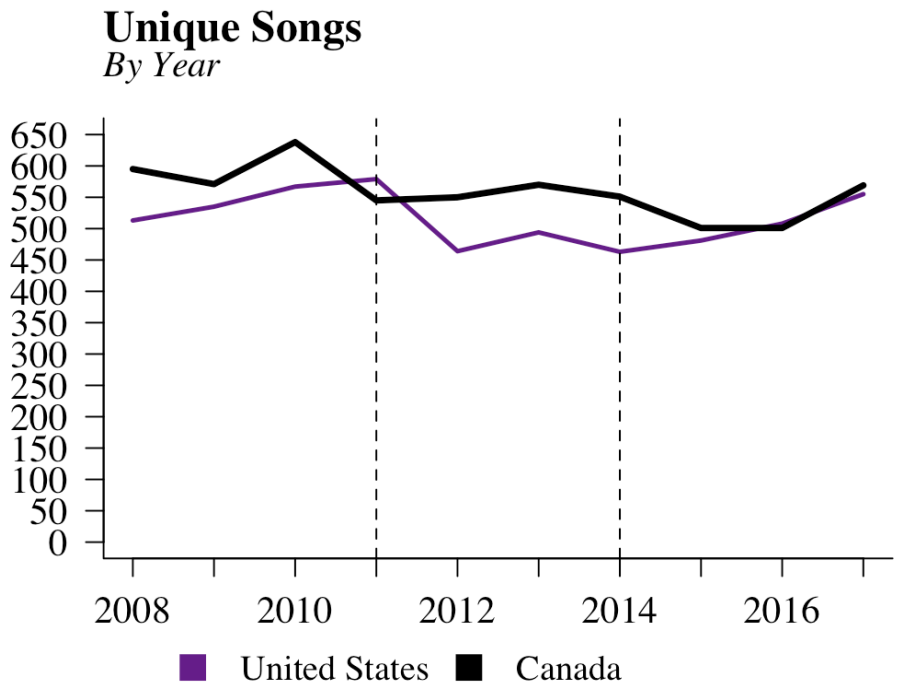
To further provide evidence that Canada was not easily able to access Spotify when it launched in the United States, Graphs 2 shows a Google Trends search of 'Spotify' in both countries. Google Trends indexes their data out of a 100, where 100 is the maximum search interest for the time and location selected. Using Google Trends data is a powerful way to see what Google users are searching and how people around the world react to different. When Spotify first launched in 2008 overseas, both Canada and the United States had immaterial amounts of activity. There was no real popularity of Spotify in the U.S. until July 2011 and in Canada until September 2014 – which are their official release dates.



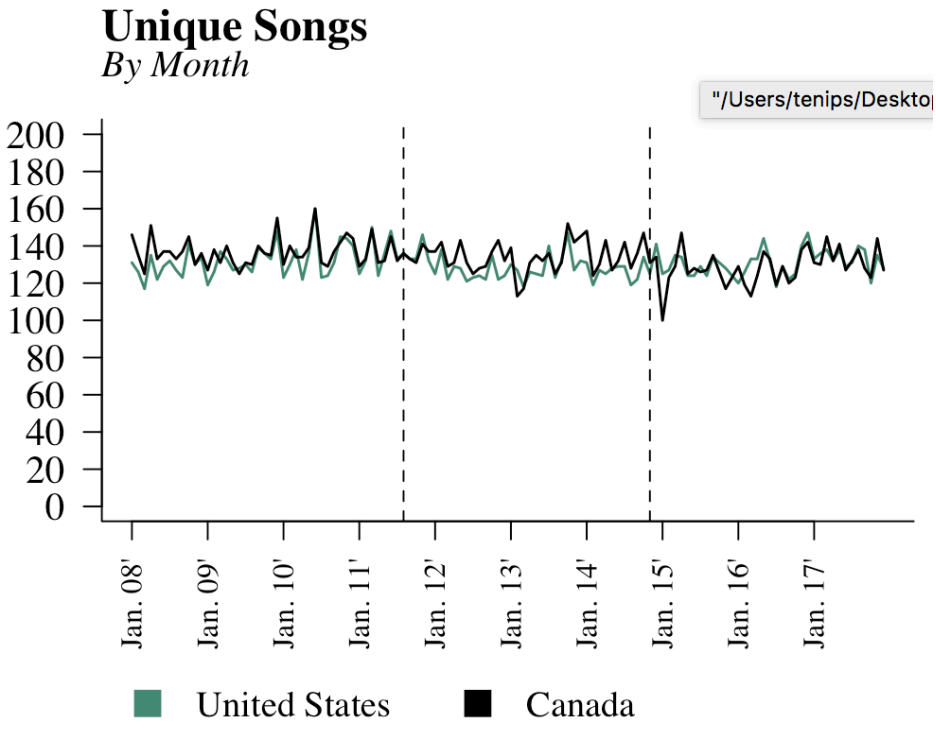
Graph 2 – Popularity of Google searches of "Spotify" for both countries.

After confirming that the time difference was valid, I used a python script to download and organize weekly data from Billboard.com on the United States and Canadian Billboard Hot 100 charts to make the variable *Unique Songs*. Next, I aggregated the weekly data into three different periods of time – 7 years, 28 quarters, and 81 months. There is further discussion about how Billboard data is collected and categorized in the Data Description below.

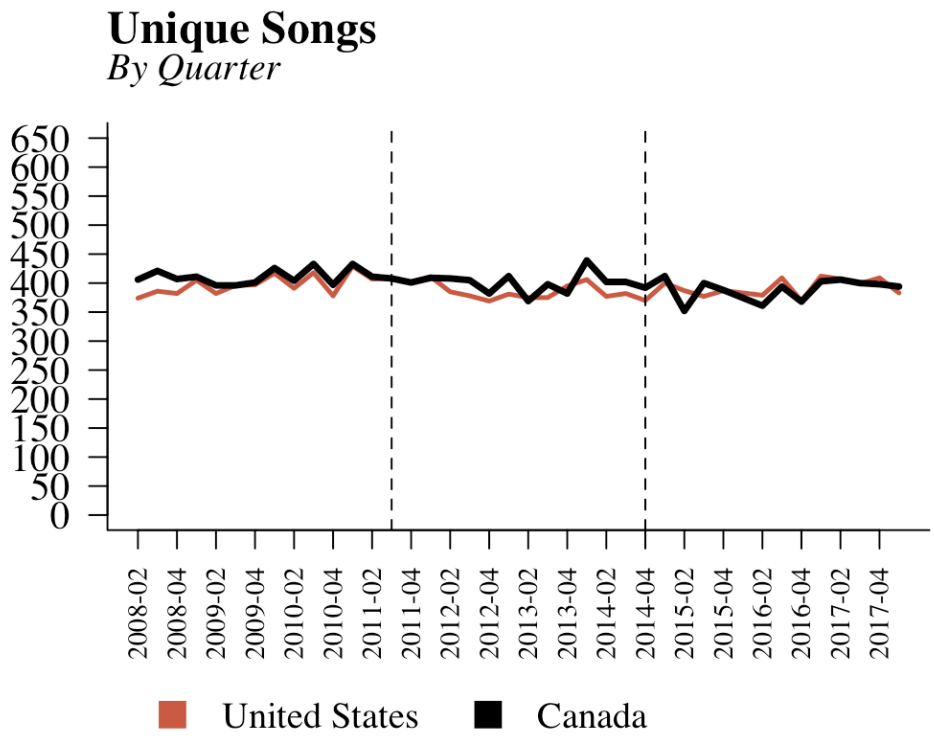
Graphs 3, 4, and 5 show the number of *Unique Songs* on the United States Hot 100 and on the Canadian Hot 100 between 2007 and 2016 using different time intervals. The dotted lines represent the introduction of Spotify in the United States and Canada, respectively.



Graph 3 – Number of unique songs, per year.



Graph 4 – Number of unique songs, per month.



Graph 5 – Number of unique songs, per quarter.

The reason why I choose to look at three different time intervals is because there is no established time for music turnover and it allowed me to capture the most perspective. Intuitively, the worst measure is the monthly interval because people easily get caught into songs for a week, three weeks, or even a couple months at a time. Also, since the Billboard Hot 100 comes out once per week it doesn't provide much opportunity for variation. The best measure would be annual turnover since it gives the most exposure to capture diversity. Since the Canadian Hot 100 chart came out in 2007, there was a limitation on how many yearly observations that could be analyzed. Looking at music turnover quarterly provides the next best option because it gives more observations within the dataset while still allowing for adequate time for variation.

As clearly shown in the yearly graph, there was a prominent declining trend in both the United States and Canada in terms of unique songs on their respective Billboard Hot 100 charts in between 2011 and 2014. For the quarter and monthly charts, the difference is not as noticeable visually because of how noisy the data gets at a granular level.

Also, shown by the graphs is that the charts follow closely together which raises another concern that Canada's Billboard Hot 100 may be heavily influenced by the United States. Due to the parallel trends assumption, this is a benefit for the D-in-D model because any differences with the introduction of Spotify can then theoretically be attributed solely to the introduction of Spotify. Looking at the data for *Unique Songs* on a yearly level does not exhibit a strong visual inspection of parallel trends which will lead to some bias in the estimates, but the quarterly and monthly are quite similar.

Datta, Knox, and Bronnenberg (2017) also used the D-in-D approach to look at how the adoption of Spotify affected individual music consumption. The model that I estimated is shown below:

$$\text{Unique_Songs} = \beta_0 + \beta_1 \text{US} + \beta_2 \text{Post2011} + \beta_3 \text{US*Post2011} + \varepsilon_t$$

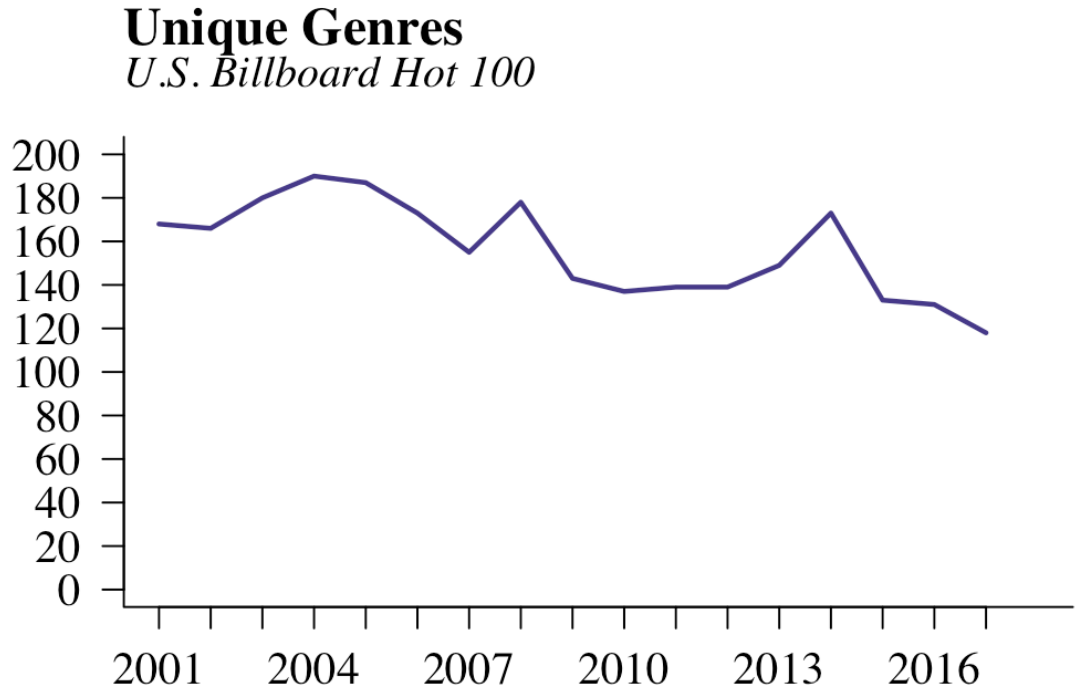
Unique Songs is the dependent variable, the number of unique songs that appear on the charts. US is an indicator variable that is 1 for the United States Hot 100 chart and 0 for the Canadian Hot 100 chart, separating the data by country. Post2011 is an indicator variable that is equal to 1 if it's a period after Spotify got introduced to the United States and 0 if it is a period before, allowing an isolation of the treatment. Additionally, ε_t is the error term. β_3 is the coefficient that shows the differences in differences between the two countries after the introduction of Spotify. Since the Superstar Effect and Long Tail theory may be happening at the same time, it is hard to isolate the effect of one or the other. My hypothesis is the Long Tail effect should be prevailing

due to the increased accessibility of music discovery that Spotify provides, allowing more people to fulfill their specific tastes.

For my supporting variable, I will be analyzing how genres have been impacted over time. The model will be focusing solely on the United States Hot 100 chart:

$$\text{Unique_Genres} = \beta_{0+}\beta_1 \text{Year} + \beta_3 \text{Unique_Genres}_{t-1+}\varepsilon_t$$

The dependent variable for the regression is *Unique Genres*, the number of different genres that appeared on the charts. *Unique Genres(t-1)* represents a lag variable on Unique Genres by one year. For this regression, my hypothesis is that the number of different genres appearing on the chart has increased over time because people have been able to use the Internet to explore more options. Graph 6 shows the trend for the number of different genres over time.



Graph 6 – Number of unique genres from 2001 to 2017.

D. Data Description

The dataset used in this study was collected from Billboards official website and includes the historical performance for the Billboard charts "The Hot 100" and "Canadian Hot 100". To properly organize the charts, I used a Python script to extract 936 weeks of data (2007 – 2016). Even though the United States Hot 100 chart has 58 years of historical data, the dataset begins in 2007 because the first Canadian Hot 100 chart did not come out until June 16, 2007. This limitation may be a problem because we're comparing the inception of the Canadian Hot 100 chart to the United States Hot 100 which had already been running since 1958.

Another limitation that comes with using the Billboard Hot 100 charts is that, frankly said, it is not the most accurate variable to capture music diversity. The Billboard Hot 100 has tried to keep up with changing consumer consumption over the years by changing the way they calculate the chart. Billboard started to include streaming and on-demand music in 2007 (Wikipedia) which falls in the line perfectly with the beginning of the sample period. The only other big change within the period of the sample is the addition of YouTube songs in 2013 (Wikipedia); but, this should not prove to be an issue since it would still allow us to analyze the impact of the United States having Spotify in 2012 compared to the absence of the platform in Canada.

Currently, according to Billboard.com, the Hot 100 formula "targets a ratio of sales (35-45%), airplay (30-40%) and streaming (20-30%)" to update the Hot 100 every week. Because of this, even if Spotify did allow people to discover and listen to more music from different artists, this effect would most likely not be captured in something as concentrated as the Billboard Hot 100 chart. For example, if two people started listening to ten new artists each week, but they

were both listening to different artists then these artists would technically never become "popular" and make it onto the charts.

Another concern is the classification of genres since they are not clearly defined for certain songs and artists and are frequently incorrect. For example, there was some controversy when Drake won a Grammy for his single "Hotline Bling" as the "best rap song" on whether this was the best category for a song that was fundamentally made as a Pop single. However, in the Grammy's defense, it is not uncommon to classify a song based on the artist – in this case, Drake is traditionally known as a rapper.

A songs genre can be classified in a few different ways such as the background of the artist, the demographic profile of the audience, the instruments played within, etc. In the dataset, the songs were categorized by the genre that artist that it was associated with. Therefore, if "Hotline Bling" showed up on the charts, it would have counted as a "Rap" single.

The following tables display the summary statistics for all four of the regressions discussed in the 'Empirical Strategy' section. Table 1 provides summary statistics of the number of unique songs that appeared in the Billboard charts in a year.

Table 1: Summary Statistics: Unique Songs on Hot 100 Chart by Year

Chart	Observations	Mean	Std. Dev	Min	Max
Canada	7	574.29	32.93	545	638
United States	7	516.43	46.43	463	579

Table 2 provides summary statistics of the number of unique songs that appeared in the Billboard charts in a month.

Table 2: Summary Statistics: Unique Songs on Hot 100 Chart by Month

Chart	Observations	Mean	Std. Dev	Min	Max
Canada	81	135.5	8.03	113	160
United States	81	130.4	8.33	117	160

Table 3 provides summary statistics of the number of unique songs that appeared in the Billboard charts in a quarter.

Table 3: Summary Statistics: Unique Songs on Hot 100 Chart by Quarter

Chart	Observations	Mean	Std. Dev	Min	Max
Canada	28	405.8	15.47	369	439
United States	28	392	16.25	369	429

Table 4 provides summary statistics of the number of different genres that appeared in the Billboard Charts in a year.

Table 4: Summary Statistics: Unique Genres on Hot 100 Chart by Year

Chart	Observations	Mean	Std. Dev	Min	Max
United States	18	156.9	21.45	118	190

E. Results

After validating the models and analyzing the data, I used STATA to regress the Differences-in-Differences model on *Unique Songs* with robust standard errors. Results for each of the three time intervals can be seen below in table 5.

Table 5: Unique Songs using Different Time Intervals

Unique Songs	Year	Quarter	Month
US	-30.25*	-12.86**	-4.52**
	(25.45)	(5.63)	(1.87)
Post2011	-38.75*	-9.79*	-2.93*
	(21.20)	(5.64)	(1.77)
US*Post2011	-44.58	-1.93	-0.95
	(27.99)	(8.11)	(2.50)
cons	587.25***	410.71***	136.90***
	(20.25)	(3.43)	(1.22)
R^2	0.7344	0.2647	0.1281

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

The results from Table 5 rejects my hypothesis for the Long Tail Theory at an insignificant level. US*Post2011 ended up being -44.58. -1.93, and -0.95 using the Yearly, Quarterly, and Monthly time intervals. To show evidence of the Long Tail Theory prevailing, the sign would have had to be positive which would represent that there's more unique songs appearing on the charts rather than the same songs continuously. Because the sign was negative for all time periods, it is safe to conclude that the Difference-in-Difference method between the two countries also supports the Superstar Effect in line with past literature, at an insignificant level.

While the US variable was not the one of interest for the D-in-D model, it shows a negative coefficient significant at the 90% and 95% level for different time intervals. Essentially, this means that the United States has had on average 12 – 13 songs less quarterly on the Billboard Hot 100 charts compared to Canada. This could be largely driven by Canada's content laws that requires a certain percentage of content that plays on the radio to be derived from Canadian producers, writers, and musicians.

After running the supporting regression to see the trend of genre over time, there was a clearer picture forming from the data. The results for the supporting regression is shown on Table 6.

Table 6: Unique Genres on Billboard Hot 100

Unique Genres	**Year**
Year	-3.37**
	(1.24)
Unique Genres t-1	.123
	(.272)
cons	6899.75**
	(2531.68)
R^2	0.5608
Adjusted R^2	0.4933

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 6 also rejects my Long Tail Theory hypothesis, but at a significant level. Looking at the Year variable which is tracking how genres has moved over time, one can interpret that for every year the number of genres that enter the Billboard Hot 100 chart is decreasing by 3 to 4 songs. The decreasing number of genres on the charts supports the Superstar Effect because the type of music that is being represented on the popular chart is going down.

F. Looking Ahead

To conclude, the results has shown support for the Superstar Effect with all four regressions. Although, this result could have been predicted visually because of the declining trend in all the graphs shown within the paper on *Unique Songs* and *Unique Genres.* As a

reminder, showing support for the Superstar Effect does not entirely discredit the Long Tail Theory because both forces may be working together at the same time.

As discussed previously in the 'Data Description' section, Billboard is not the best representation of music diversity and therefore may be the reason why the results are insignificant. For example, Billboard recently announced another weight change in the Billboard Hot 100 charts in 2018 where they will be giving YouTube videos less weight compared to paid or ad-supported streams. Because Billboard is a trade publication, it is not surprising that want to show more of where consumer's money is spent rather what they are spending their time listening to. One thought for Billboard moving forward is to decrease the weight on sales and airplay and increase the weight of streaming and YouTube views if they want to truly capture popularity instead of solely commercial tastes.

Individual listening habits most likely capture music diversity better than a trade publication which was seen in Datta, Knox, and Bronnenberg (2018) when they tested the effect of an individual's adoption of Spotify. Specifically, measures like the depth at which people listen to new artists would be beneficial to help explain the trend of diversity. For example, if someone found new artists and listened to one of their songs for three minutes it would be significantly different than if they found a new artist and continued to listen to them for several months. Another measure that would be good to analyze for diversity is popular music festivals and the range of genres that have appeared over time. Big music festivals, such as Coachella and Lollapalooza try to cater to as wide of an audience as possible so we could expect that their lineup would appropriately represent the changing music tastes of the population.

Overall, while better quality data would have been beneficial for this research project, these findings have helped support the prevalence of the Superstar Effect even as the

accessibility of information has increased – which could largely be attributed to what was mentioned in the “Literature Review” as the paradox of choice.

G. References

Adler M, 195. Stardom and talent. Amer. Economic. Rev. 75(1):208–212.

Anderson, C., 2006. The long tail: how endless choice is creating unlimited demand. London: Random House Business.

Brynjolfsson, E., Hu, Y. & MD Smith, 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. Management Science.

Brynjolfsson, E., Hu, Y. & Simester, D., 2011. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. Management Science, 57(8), pp. 1373-1386

Bhattacharjee, S. et al., 2007. The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts. Management Science, 53(9), pp. 1359- 1374.

Bradlow, E. T. & Fader, P. S., 2001. A Bayesian Lifetime Model for the "Hot 100" Billboard Songs. Journal of the American Statistical Association, 96(454), pp. 368-381.

Erlandsson, David and Perez, Jomar, 2017. https://insights.spotify.com/my-en/2017/11/02/listening-diversity-spotify/

Giles, D. E., 2007. Survival of the hippest: life at the top of the hot 100. Applied Economics, 39(15), pp. 1877-1887.

Hannes Datta, George Knox, Bart J. Bronnenberg (2017) Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery. Marketing Science

Hiller, Scott and Kim, Jin-Hyuk, 2017. Online Music, Sales Displacement, and Internet Search: Evidence from YouTube. CAPRI Publication 13-2.

Lao, Jerry and Hoan Nguyen, Kevin, 2016. One-Hit Wonder or Superstardom? The Role of Technology Format on Billboard's Hot 100 Performance.

Mulligan, Mark, 2014. The Death of the Long Tail: The Superstar Music Economy. https://www.midiaresearch.com/downloads/the-death-of-the-long-tail-the-superstar-music-economy/

Music Business Worldwide, 2015. Independent Music is a Growing Force in the Global Market. https://www.musicbusinessworldwide.com/independent-music-is-a-growing-force-in-the-global-market/

Sjolander, Simon, 2016. Did Modern Media Kill the Superstar? A contribution to the theory of consumer behaviour in the presence of increasing information.

Stigler, George and Becker, Gary, "De Gustibus Non Est Disuptandum," American Economic Review, March 1977, 67, 76-90.

Trust, Gary, 2013. Ask Billboard: How Does the Hot 100 Work? https://www.billboard.com/articles/columns/ask-billboard/5740625/ask-billboard-how-does-the-hot-100-work

Wikipedia, https://en.wikipedia.org/wiki/Spotify

Wikipedia, https://en.wikipedia.org/wiki/Billboard_Hot_100#cite_note-2