MASTER THESIS MARKETING ANALYTICS AT TILBURG UNIVERSITY

Assessing the influence of major label content on Spotify playlist success

An empirical study of the 10,000 most popular playlists on Spotify

Name:	Robbert Oudelaar
SNR:	
Supervisor:	M.J. Pachali
Second assessor:	N.M. Alagöz
Date:	May 31, 2020

Management summary

The rise of digital alternatives impacted the way that music is consumed and the role that music labels have within this originally oligopolistic market structure. Music is mainly consumed via online streaming services, with Spotify as market leader. Two-thirds of time spent listening occurs through the backbone of how content is used on Spotify; playlists. Research based on the shares of content among the most popular music suggests that independent labels are closing the gap with the major labels (Universal, Sony, and Warner), but that these majors still dominate the market. Meanwhile, there is little to no research available on the drivers of playlist success, while it is referred to as the backbone of how music is consumed on Spotify. More specifically, there are yet no insights on how consumers value content from these major labels in the context of playlists. This study adds to existing literature by assessing to what extent the share of major label content influences playlist success in general. Playlist curators seek for more insights that elucidate the way that playlist characteristics drive playlist success. Because of the dominating status among popular content from these major labels, we expect this relationship to be positive for Universal, Sony and Warner.

This thesis uses panel data of the 10,000 most popular playlists on Spotify, including daily observations of playlist characteristics. We specify a log-log model with playlist fixed effects that allows to control for omitted time-invariant factors, and include a rich set of control variables that are time-variant. The coefficients of this model represent the elasticity of the number of followers with respect to the shares of major label content on a playlist. We find a significant and negative elasticity for the share of Sony content in general and fail to find an effect for the share of Universal and Warner content. We include a variety of robustness checks that show that the elasticities are stable and trustworthy. We discuss a couple of problems that might cause the estimates to be biased, for example in case of time-varying factors that influence playlist success that are not included in our model. The fact that we find one general elasticity for all popular playlists, including many different types of playlists, makes its practical implication obscure. Estimating these elasticities on playlist level would give more usable and generalizable results. Future research is therefore highly desired, also to find how the effects differ per genre and type of playlist owner. This study, however, presents a fair representation of the general effects.

Preface

This master thesis is written in order to complete the Master Marketing Analytics at Tilburg University. I decided to work on a project that personally interested me and offered the opportunity to develop new skills. This study interested me, because I have a lot of affection with music and use Spotify on a daily basis myself. Expanding my skillset by learning statistical programming in R and analyzing panel data is really valuable and is something I really enjoyed.

I would like to express my appreciation to everyone that supported me during the process of writing this thesis. In special, I would like to thank my supervisor Max Pachali for his critical view and directions. I am very thankful that our meetings could continue, despite of the inconveniences that came along with the corona virus. I would also like to thank dr. Hannes Datta for his support on the problems I encountered during the process of writing this thesis.

Finally, I would like to thank my parents, brother, and sister for their support during my entire educational career. Without their support I would not be in the position that I am today.

Robbert Oudelaar

Table of contents

1. Introduction

Online streaming services caused a revolutionary shift in the consumption of media in the past decade. Streaming has quickly gained market share in different media markets such as music, television and films (Amy Watson, 2019). Streaming accounted for 80% of revenues for the music industry of the United States in 2019 (Joshua P. Friedlander, 2020). The availability of streaming platforms and the mobility provided by smartphones cause consumers to stream music more often and in more places than ever before. Music streaming platforms such as Spotify, Apple Music and Amazon allow consumers to access a large variety of tracks on different devices.

With a 36% share of global paid music subscriptions in 2019, Spotify is the market leader among music streaming platforms (Amy Watson, 2019). Spotify offers more than 50 million tracks to 271 million users, of which 124 million subscribers, across 79 markets (Spotify Newsroom, 2020). Two-thirds of time spent listening occurs through the backbone of how content is used on Spotify; playlists (Goodwater, 2019). Discovering music through playlists occurs either via playlists created by an algorithm, often personalized and based on music you like, such as "Discover Weekly" and "Release Radar", or via general one-to-many playlists such as "Rock Classics" and "RapCaviar" (Spotify, 2020). This latter type of playlist is usually curated by humans and often contains artists that are widely known. Although many different playlists and playlist curators exist, popularity is highly concentrated among few curators. A quick view on chartmetric.com, a website providing reliable data, visuals and in-depth insights of the music industry, shows that out of the 100 most popular playlists, 98 are owned by Spotify itself (Chartmetric, 2019).

The rise of this digital type of music consumption has led to low-cost alternatives for music creators, resulting in substantial growth of independent labels (Waldfogel, 2017). These independent labels compete with the major labels (Warner, Universal, and Sony), which possess the majority of the market share in this oligopolistic market. The new digital environment allows independent labels to compete more effectively. With a change of independent labels' global market share of digital and physical recorded music revenue from 24% in 2012 to 33.8% in 2018, a significant change is witnessed in the consumption of these smaller labels (Amy Watson, 2020). Brynjolfsson, Hu and Smith (2003) and Anderson (2006) already presented the perspective of the mechanism in how the long tail of media markets is getting more relevant within the digital economy. Internet users have access to millions of products that they could not easily locate or purchase through brick and mortar retailers, leading to a substantial total welfare gain, that is negligibly small compared to the consumer welfare gain due to increased competition and lower prices (Brynjolfsson & Smith, 2000). Aguiar and Waldfogel (2016) confirmed the more relevant long tail for the industry of recorded music and found a significant rise of the share of independent label music consumption within the top 100 selling songs from 2006 to 2011 in the United States. Thus, Independent labels don't only have an increasing share of total consumption, they also have an increasing share among the most popular music that consumers listen to. Consequently, curators of playlists on digital streaming services such as Spotify can no longer ignore this stream of independent music success.

The growth of this digital type of music consumption did not only change the consumption distribution of major versus independent label music, it also increased the impact of how playlists are composed. Being included in an official playlist can help turn an artist into a superstar (The Economist, 2018), as it raises streams for prominent tracks substantially (Aguiar & Waldfogel, 2018). Being included in the Today's Top Hits list is worth almost 20 million extra streams, which translates to about $116,000 - $163,000 revenue from Spotify alone. But the impact of how playlists are composed are not only huge for artists, it might also impact the success of the playlist itself, and with that the curator, substantially. Knowing how playlists should be composed and maintained is therefore essential considering the success of a playlist. This is an increasingly important matter, since the impact of playlists curators' decisions keep on growing as the amount of Spotify users rise. These curators are confronted daily with a variety of new releases and already existing tracks that potentially deserve a spot in their playlists. For every track that is included in the playlist, a track already in the playlists is often removed. Both of these decisions have an effect on the characteristics of the playlist. Playlist curators might therefore seek for insights that elucidate the way that playlist characteristics drives playlist success and more specifically the share of major label content. Providing these insights helps curators understand and optimize their decisions. Without these insights, playlist curators can only guess what the effect of their decisions will be and might therefore have negative consequences for playlist success. Knowing how to manage the composition of a playlist might lead to higher playlist success and therefore a better position compared to competing playlists. Also, increasing playlist success might imply that users like the composition of a playlist better, leading to higher user satisfaction.

In conclusion, the long tail of media markets is getting more relevant within the digital economy and also within the music industry. Digitalization has led to alternatives that give independents the possibility to create and distribute music at substantially lower costs. Independent labels have an increasing share of total successful music consumption, but the major labels still seem to be in advantage. However, most studies looking into this matter examine shares of content among the most popular music, and no study captures how consumers respond directly to major label content.

It would be interesting to study how these major labels perform against each other, especially in the context of playlists. Playlists specifically are interesting because so much of total time spent listening occurs through them, but also because curators seek for insights for optimal playlist composition. Therefore, the purpose of this study is to obtain causal effects for each major labels' content share on playlist success. Obtaining these estimates for each separate label makes it possible to give an indication of which label has the most popular content compared to one another.

By using the Chartmetric API, we obtain static characteristics of the top 1.1 million playlists on Spotify, such as the playlist owner and number of listeners. For the top 10,000 playlists, we obtain dynamic playlist characteristics per day for a time span of approximately 4 years. This dataset contains characteristics such as the amount of followers, share of major label content and other track characteristics. Another dataset contains a list of 167,000 music labels that need to be classified into either a major or independent label. In terms of variables, the share of major label content is obtained as a proportion of the total content of a playlist. More specifically, this variable is obtained for each separate major label, allowing to compare the effects of these labels. This allows to give an indication of popularity of each label and more specifically their artists and contents.

2. Literature review

This chapter gives an overview of related research in this topic. There is little research available on how consumers respond to content from major labels. The drivers of playlist success is also not clearly defined in the literature. We therefore look at the topics that are available, and evaluate research in comparable contexts and markets. First, the role of major labels in the music industry is outlined in paragraph 2.1. Next, playlist success is documented in paragraph 2.2. Paragraph 2.3 covers the contribution to the literature, and the conceptual model and hypotheses are outlined in paragraph 2.4.

2.1 The role of major labels in the music industry

Although the composition of the major labels has changed over the years due to mergers and acquisitions of subsidiaries, the role of this group of record companies dominating the music industry has always been large and is often examined in research. The extent of success among popular artists is not equally distributed, but highly skewed, which is not surprising in a market where few “superstars” dominate (Strobl & Tucker, 2000). In their study, Strobl and Tucker (2000) state that the increasing concentration of industry activities in the hands of few international record companies was an important feature of the 1980’s and 1990’s, which was part of a general merger boom in the global economy from 1981 to 1989 with the intention to increase market share by deliberately seeking to diversify activities rather than relying on a single market. These record labels had a combined market share of over 80 percent, while they were directly involved in most activities in the industry, including recording, manufacturing, distributing, publishing, and the collection of royalties. Strobl and Tucker (2000) argue that the biggest problem the independent labels face are a lack of resources, which would enable them to attract leading artists to their label and a limited distribution network which prevents them from operating in some markets and also from exploiting economies of scale. Alexander (1994) predicted that digital distribution networks may promote greater competition in the industry, but only in case they would be non-exclusionary. He states that independents are unable to obtain national distribution, possibly reducing the diversity and variety of product offerings, since in part small firms tend to be product innovators. A digital distribution network potentially attenuates the effects of significant barriers to entry the music recording industry. It gives firms the opportunity to have their products distributed in a less-costly and non-exclusionary fashion.

2.1.1 Survival of major and independent content

Bhattacharjee, Gopal, Lertwachara, Marsen and Telan (2007) assessed the impact of the first digital developments related to the music industry, such as the mp3 player, peer-to-peer technologies, file sharing networks like Napster and online music stores on survival of music albums on the charts. They used data on the performance of music albums on the Billboard charts with file sharing data from a popular network. They find that albums promoted by minor labels tend to have survival duration 23% less than those promoted by major labels. They state that major labels generally have stronger capital than minor labels and can more easily promote and distribute music. However, independent albums have experienced a significant beneficial shift and are surviving longer than before. They therefore conclude that independent labels made a first step in closing the gap with the major labels. In a more recent study, Ren and Kauffman (2017) assess how track popularity and duration on the charts are determined with a dataset of 78,000+ track ranking observations from a streaming music service. Their results show that it is possible to explain chart popularity duration and the weekly ranking of music tracks. They concluded that whether an artist was signed at a major label or not did not have an effect on music track popularity duration, while it had a positive and significant effect on album sales. Another study by Im, Song and Jung (2018) investigated what factors are critical for music to succeed in download and streaming services, using the top 100 songs listed on the Korean music ranking charts between 2011 and 2014. For the download charts they find that songs from major labels survive longer than those of independent labels, in line with the finding by Bhattacharjee et al. (2007). However, the opposite was found for the streaming chart. They conclude that streaming services provide minor labels the opportunity to succeed, while they decrease the influence of major companies. Therefore, there are inconsistencies in the results regarding whether major label contents survive longer on the charts or not. One could possibly argue that there was still an advantage for major labels in time Battacharjee et al. (2007) conducted their work and that this effect disappeared in time others (Ren & Kauffman, 2017; Im et al., 2018) conducted their study.

Another possibility is that the effects differ for streaming services and the billboard charts, since all work that used data from streaming services (Ren & Kauffman, 2017; Im et al., 2018), which was not used in Bhattacharjee et al. (2007), point into the direction where major labels do not have an advantage relative to independent labels. This might suggest that the effects of major labels were present before streaming services made their introduction to the market, but don’t hold on data from streaming services. This would be in line with findings and a statement from Datta, Knox and Bronnenberg (2018), who studied how the adoption of music streaming affects listening behavior. The price of additional variety is set to zero by adopting digital streaming, while experimentation was expensive with music ownership. They found that adoption of streaming leads to very large increases in the quantity and diversity of consumption in the first months after adoption. Even after half a year, adopters play substantially more and more diverse music. They state that the shift from ownership to streaming is probably more favorable for smaller artists and labels (e.g. independent labels), since it levels the playing field, but investigating this was not included in their study.

2.1.2 The share of major and independent labels among popular content

The development of digital distribution networks enable low-cost digital distribution (Waldfogel, 2015). Not only distribution costs have lowered, costs for recording have also significantly dropped by low-cost equipment and software. The combination of both have allowed smaller music labels, but also individuals, to both release more music and bring this music to consumers' attention. Waldfogel (2015) finds that where in the past majors would dominate commercial success, independent labels now account for one-third of the artists appearing on the Billboard 200 each year. This would mean that consumers nowadays find much of the independent music more appealing than much of the diminished major-label fare. Aguiar and Waldfogel (2016) researched the independent share of the top 100 in the United States and Canada from 2006 to 2011. They found that independent record labels indeed have a rising share among the most popular content. In Canada, the independent share increased 0.9 to 1.9 percentage points per year, to about 10% in 2011. In the United States, it increased about 1 percentage point per year, to about 12% in 2011. These increases are statistically significant for the top 10,000 and top 50,000. Another study by Aguiar and Waldfogel (2018) analyzes the daily top 200 songs on Spotify in 2016 and 2017. They found that 19.0% of content of global streams is owned by independent labels. Also, they argue that Spotify has the power to influence consumption decisions. Artists that get their song on Today's Top Hits can expect almost 20 million additional streams on that song and platform only, which translates to a monetary value of $115,000 - $163,000. Spotify also has an immense influence on the success of new songs and artists, since getting on the top of the New Music Friday playlist in the United States is worth roughly 14 million streams, with a monetary value of $84,000 - $117,000. While the major global lists tend to promote major-label and United States-origin music, the New Music Friday lists provide a higher share of independent and domestic music.

The first stream of research on survival of content suggests that the gap between independent labels and the majors is closing, or might even be closed already. However, these studies all look at the survival of albums when they are already popular and ignore the fact that it might still be harder for independent labels to reach a place among the top content. The second school of thought shows that independent labels are closing the gap with the major labels, but own just a rather small share of the most popular content, implying it is still dominated by major labels. Another study that suggests that the major labels still dominate the market is given in Guichardaz, Bach and Penin (2019). These majors might not only dominate the intermediate functions of the new music markets, they are also dominating the metric race together with the music streaming services (Maasø &

Hagen, 2019). These powerful groups seem to benefit the increasing amount of data available the most.

2.1.3 Welfare gain

Another stream of research shows the results on consumer welfare that corresponds to the increasing amount of long tail products since the rise of digital assortments. In conventional retailer stores, limited shelf space constraints the type of products that a consumer can discover, evaluate and eventually purchase. Limits on shelf space are substantially lower for internet retailers, resulting in customers having access to millions of products that they didn't easily have access to in the conventional retailer stores. Brynjolfsson and Smith (2000) found that prices on the internet are 9-16% lower than prices in conventional outlets and that price adjustments are smaller online in the case of books and CDs. These lower prices, due to increased market efficiency, provide significant benefits to consumers. In later research, Brynjolfsson, Hu and Smith (2003) quantify the economic impact of product variety that came along with electronic markets in the case of online booksellers. They found that increased availability of products, that previously were hard to find, represents a positive impact of consumer welfare that is seven to ten times larger. Anderson (2006) claims in his book that the long list of products at the end of the distribution tail are of growing importance relative to the small number of products at the head.

A more recent study (Waldfogel, 2017) argues how digitization has ushered in a golden age of music, movies, books and television programming. The growth of digital music consumption emerged alternatives to create music that cost significantly less, resulting in substantial growth of independent labels. Bringing new products to these markets is much easier nowadays. Waldfogel (2017) confirms that the early view of digitization's effect on consumers in media markets is correct. Increasing access to the long tail of existing products is associated with a substantial welfare gain. The effects of digitization on production are even more substantial. But while independent labels all together might capture a larger share of popular content than before and benefit the low-cost alternatives of music creation, individual musicians must first overcome a new and dynamic range of barriers to success (Hracs, 2012). Barriers to enter the market have significantly lowered, but that market is loaded with competition and uncertainty.

2.2 Record labels and playlist success

Other research looks into the success of music and more specific playlists. The first subject that is covered is how success of playlists can be measured. As Paul Lamere (2014) discloses, listeners are telling us a little bit about their music taste every time they adjust the volume, skip a song, search for an artist, or abandon a listening session. Resembling this matter, the number of followers of a playlist indicates whether users like the composition of that playlist and can thus be used as measure for playlist success, as also used by Jenkins and Joven (2018). However, the addition of the estimated playlist listener count on chartmetric.com made Joven (2018) believe that the Follower-to-Estimated-Listener (FEL) ratio could be a more meaningful playlist engagement metric. In his study, Joven (2018) finds that the six playlists with the best FEL ratios are "This Is" playlists and all mega stars happened to be female (Madonna, Britney Spears, Beyoncé, Amy Winehouse, Katy Perry, and Taylor Swift). He then argues that fans of female A-list artists possibly show both interest by following and are more dedicated by actually listening. Such insights would have not been unveiled if only the number of followers were analyzed.

Furthermore, very little research that relates to the current study has been carried out on playlists. One of the few studies that did relate was that from Aguiar and Waldfogel (2018), whose findings were already discussed in the previous paragraph. Though, they rather looked into the effects of song success on popular playlists than on playlists success. Other interesting work was executed by Boughanmi, Ansari and Kohli (2019). They disclose how musical features among successful albums changed over time and which characteristics might influence album success nowadays. While they didn't look specifically at playlist success, their findings of how musical characteristics explain album success might be applicable to the case of playlists. They find that louder music has become more successful over time (which is also concluded by Serra, Corral, Boguñá, Haro and Acros (2012)) and that albums with slower tempo used to be more successful until 2010. Also, albums with songs that were recorded in lower keys seem to be more successful nowadays. These albums that are more successful nowadays have simpler compositions and combine a few complex songs with several simple ones. While albums with longer songs have always been less successful, albums with songs of different lengths have done better in recent years. Successful albums today have lower energy and low valence, while albums with high energy and low valence were more successful in the 60-80s. Constructing a balanced, harmonious and successful song compilation is important in both the case of albums and playlists. It is valid to assume that albums and playlists are similar in the way their curation affects their success, but certain factors logically don't hold up for both. Finally, Boughanmi et al. (2019) found that whether an album was housed by a major label had positive effects on album success in all years, but this effect diminished since the 1990's. This again indicates that the gap between independent and major labels is closing, but still suggests an advantage for major labels.

2.3 Contribution to the literature

To summarize, major labels' role in the music industry has always been huge and still is. Digitalization offered independent labels and artists the opportunity to compete with the major labels more effectively and had large impact on the consumption of long tail products in general. One stream of research (Bhattacharjee et al., 2007; Ren & Kauffman, 2017; Im et al., 2018) suggests that the gap between the majors and independents is closing, or might have already been closed, based on data about survival on the charts. However, these studies look at content that is already among the most popular and don't take into account that it might be harder for smaller music labels to actually reach the top. Another school of thought (Waldfogel, 2015; Aguiar & Waldfogel, 2016; Aguiar & Waldfogel, 2018; Guichardaz et al., 2019; Boughanmi et al., 2019) concludes that, even though independent labels are closing the gap, majors still dominate the market, based on the shares of content among the top music. While new alternatives to create and distribute music led to a significantly higher welfare gain for consumers, it made the market also loaded with competition and uncertainty. Another stream of research investigates which musical characteristics lead to album success and shows that different attributes play an important role.

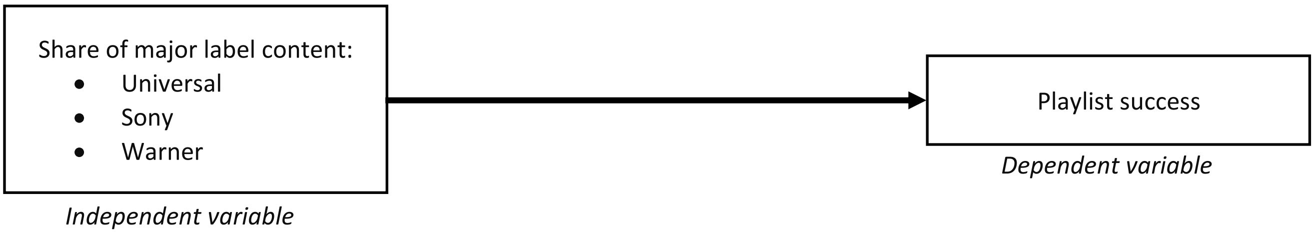
Most of the literature that look at and compare content from major and independent labels point into the direction that major labels still have an advantage over independent labels. Even though some research suggest that the gap is closing, and it is clear that independents benefit new digital types of consumption, it still implies that the gap is not fully closed yet. The literature that does suggest the gap is closed only looks at survival of content that is already popular, and don't take into account that independents might have more trouble reaching the top. Meanwhile, there is little to no research available on the factors that drive playlist success in the contemporary streaming era. Current literature encloses the musical attributes that can predict album success, but how this applies to the case of playlists is yet uncovered. While current literature investigate major labels' shares among the most popular content and survival on the charts, there are yet no insights on how listeners respond to major label content and how they value it in the context of playlists. This study therefore adds to existing literature by explaining to what extent the share of content of each major label influences playlist success, more specific in number of followers and listeners. We focus on providing insights that give a first implication of how the content from major labels is valued in general in the context of playlists.

2.4 Conceptual model and hypotheses

This study investigates the relationship between the share of major label content and playlist success. The share of major label content comprises how much of content on a playlist is owned by either Universal, Sony or Warner. We observe these shares on a set of playlists and investigate their effects on the number of followers and listeners of that playlist. The visual representation of the conceptual model is shown in Figure 1.

Figure 1

Conceptual model



Most research (Waldfogel, 2015; Aguiar & Waldfogel, 2016; Aguiar & Waldfogel, 2018; Guichardaz et al., 2019; Boughanmi et al., 2019) suggests that the gap between independent and major labels is closing, but they all conclude that the majors still have an advantage. Research on the survival of albums (Bhattacharjee et al., 2007; Ren & Kauffman, 2017; Im et al., 2018) argue that the gap between major and independent labels is getting smaller, or is closed already. Though, these studies only take albums into account that are already among the most popular content and therefore do not take into account that it might be harder for independents to acquire a spot among the top content. If major labels still perform best among the most popular content, and this type of content stimulates playlist success, it is valid to hypothesize that contents from major labels have a positive effect on playlist success. Existing literature does not include any perspective on how consumers respond to major label content. We therefore assume that the effects of all three of the major labels move into the same direction. We therefore hypothesize:

H1: The share of Universal content is positively related to playlist success.

H2: The share of Sony content is positively related to playlist success.

H3: The share of Warner content is positively related to playlist success.

3. Data

This chapter outlines the steps that we take in order to use the data to estimate our model. First we describe how the data is collected. Then we document data cleaning and preparation. Next, variable operationalization is discussed in paragraph 3.3. Finally, an overview of descriptive statistics is presented in paragraph 3.4.

3.1 Data collection

The analysis in this thesis draws on three underlying data sets. The first data set contains daily playlist characteristics for the top 10,000 playlists, measured at the end of the time frame, between 01-01-2016 and 25-11-2019. The 9,083,636 rows in this data set contain characteristics such as the share of major label content, number of tracks, and number of followers. The full list of variables in this list can be found in appendix 1. This data will be referred to as panel data. The second data set contains the top 1.1 million playlists and their static characteristics and will be used to compute the FEL ratio for each playlist and to determine by whom the playlists are owned. See appendix 2 for the full list of variables in this data set. This data set will be referred to as playlists data. The third and final data set obtained is a list of 167,000 music labels. The major labels in this list have to be classified into either Universal, Warner or Sony in order to assign the share of their content for each playlist on each day in the panel data. Therefore, the first step is to make an adequate classification of these labels.

3.2 Data cleaning and preparation

3.2.1 Classifying major labels

An existing code that classifies major labels, made available by Hannes Datta on GitHub[1], is used as basis for this challenge. This code already identified a large part of the major labels, but was not yet complete. The code has to be modified in order to make the classification more complete. In the base code, 737 labels are classified as Sony. 14 labels are added to this in order to classify 751 labels as Sony's in total. The base code identifies 1936 labels as Universal. 11 other music labels are identified to also belong to Universal. However, a major mistake in the base code wrongly classifies 461 labels to be Universal's. Correcting this mistake and adding the 11 other labels sets the total number of Universal labels to 1486. Another 116 music labels are added to the 380 Warner labels from the base code, setting the total number of Warner labels on 496. The exact modifications of the code can be found at the revisions of the gist on GitHub[2]. The changes in the classification of major labels leads to changes in the share of major label content in the panel data. We therefore update the panel data, considering these changes. Note that this method does not guarantee that all

[1] https://gist.github.com/hannesdatta/19508cba3ab80bf0b2a648bec2480d0e
[2] https://gist.github.com/hannesdatta/19508cba3ab80bf0b2a648bec2480d0e/revisions

major labels have been classified. Yet, the classification should contain an amount of labels that is large enough to be informative for this study. This method is similar to that of Aguiar and Waldfogel (2018).

3.2.2 Playlist and panel data

The first step of cleaning the playlists data is to check if and how many playlists appear multiple times. We observe that 94,175 playlists are duplicated in our data. For these duplicates we retain the first record and filter out the others, reducing our data set from 1,094,900 rows to 1,000,725 rows. The final 165,879 playlists of the data set appeared to have a follower count of -1 or NA, while it should be 0. For these playlists, the value is set to 0. A first calculation shows that the top 10,000 playlists account for 85% of all followers of playlists in our data.

As mentioned before, the panel data is gathered for the 10,000 most popular playlists only. Since success is highly concentrated at the most popular playlists, taking only the top 10,000 makes sure that there is enough variation in the dependent variable to measure effects. The panel data does not have any duplicate rows. When looking at the number of distinct playlists ids, we observe that 9,993 different playlists actually occur in the panel data. A quick look at the data shows that there are quite some rows that have no value for the followers variable. Removing all 1,939,526 rows where followers value is NA or 0 from the 9,083,636 rows we started with, resulted in 7,143,813 remaining rows[3]. The date now ranges from 03-11-2016 to 09-10-2019. If we look at the total number of different playlists now, we see that 9,992 are remaining, owned by 3,644 different users. A check of the ranges of each major label share shows that all are within the expected range of 0 and 1. Also, all other variables that later will be used in our model have a realistic range. 2,225 rows in the panel data seem to miss values for all musical characteristics. It appears that these rows belong to 13 playlists, which are all owned by Spotify. The potential bias that is associated with this nonrandom missing data is discussed in section 6.3.

3.3 Variable operationalization

In order to run the model, that we later specify in chapter 5, a few variables need to be added to the existing data. The FEL ratio is appended to the playlists data. This variable is required to compute the number of active listeners, used as dependent variable to check for robustness later in this study. The time trend and dummy variables for month-of-the-year and day-of-the-week will be added to the panel data.

[3] 298 rows appeared to have a follower count of 0. A quick check showed that these values were incorrectly measured.

The FEL ratio is computed by the static follower count divided by the number of monthly listeners. A high FEL ratio therefore means that there is a large difference between the amount of followers and actual listeners. This might imply that the list may seem a great idea at first, but listeners forget about it later. A FEL ratio of 1.0 would mean that all followers actually listen to the playlist. For the top 10,000 playlists, we observe that 3,842 have a FEL ratio which is NA. All these missing values are due to missing monthly listeners values in the playlists data. The number of listeners will not be used as initial dependent variable due to this large amount of missing data. It will be used as check for robustness, while the number of followers will be used as initial dependent variable. We next join the FEL ratio to the panel data. In the panel data, 4,555,973 out of the 7,144,110 rows contain this FEL ratio. Next, the amount of active followers is computed over time as the number of followers of that playlist on that date divided by the FEL ratio. The second variable that we compute in the playlist data and then join to the panel data is playlist ownership. This variables indicates whether a playlist is owned by Spotify, Universal, Sony, Warner, Independents or other. Next, the time trend is added to the panel data. This variable is computed by subtracting 02-11-2016 from each row's date, such that the time trend starts at 1 for the first observed date, which is 03-11-2016. The maximum value of this time trend is 1071 for date 09-10-2019. 11 dummies are created for the month-of-the-year effects with December as base level. 6 dummies are created for day-of-the-week effects with Sunday as base level. Finally, we want to include competitor shares of major label content in our model for all major labels. For each playlists and each day we therefore compute the share of major label content of all other playlists on that specific day. More precisely, the sum of share of major label content is taken for every day and every playlists. The value of the regarding playlist and day is then subtracted from that sum. This remaining value is then divided by the total observed playlists for that specific date. In this manner, three columns were added with the mean competitor content share of each major label.

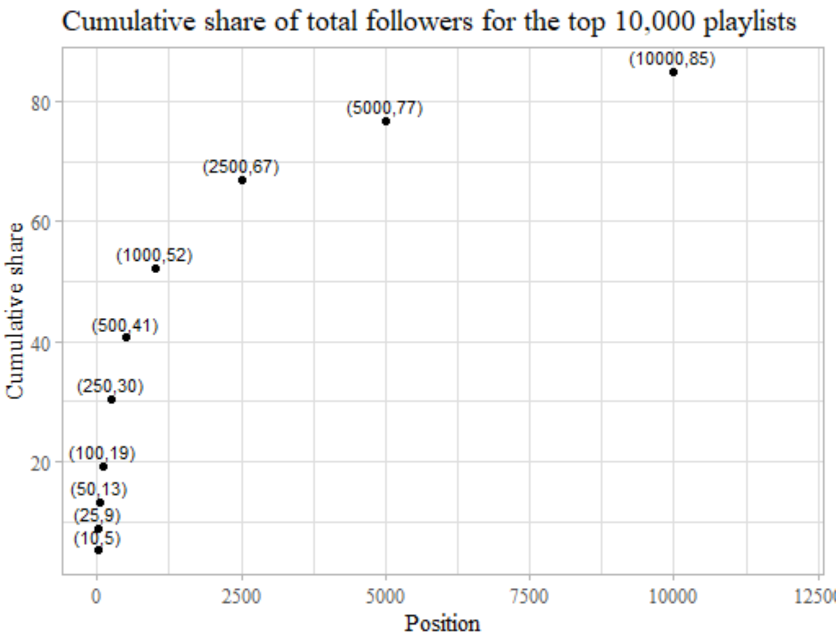
3.4 Descriptive statistics

Next, we describe the market for playlists on Spotify by means of documenting descriptive statistics. Of all playlists in the playlists data, 12,518 are owned by Spotify itself (1.25%), 7,812 are owned by major labels (0.78%) and 980,395 are classified as 'other' (97.97%). All these playlists are curated by 248,770 distinct curators. If we look just at the top 10,000 playlists, we see that 4,104 are owned by Spotify (41.04%), 839 by major labels (8.39%) and 5,057 by others (50.57%). For the top 10,000 playlists, we observe that they are owned by 3,646 distinct curators. The data shows how well Spotify's own playlists perform compared to those owned by major labels or others. If we take an even closer look at the performance per owner class on all playlists data, we observe that the mean followers of a playlist owned by Spotify is 106,933, compared to 13,371 for those owned by major labels and 596 for those owned by others. We also document the median of 2,608 followers for

Spotify, 536 for major labels and 1 for playlists owned by others. When looking at the mean FEL ratio per owner class, we observe that playlists owned by Spotify do not only have a higher follower count, they are also listened to more often. The mean FEL ratio for their playlists is 28, while the mean for major label owned playlists is 127 and 132 for those owned by others.

The number of followers of all playlists in the data lies between 0 and 24,129,529. There are 356,917 playlists with a followers count of 0. The average amount of followers is 2,026.709, and the median of 1 doesn't tell much, except that success is highly concentrated. How success is distributed among the playlists is shown in Figure 2, where the cumulative number of followers of the top 10,000 playlists is shown. The playlist on position 10,000 had 21,507 followers at the moment that the data was retrieved.

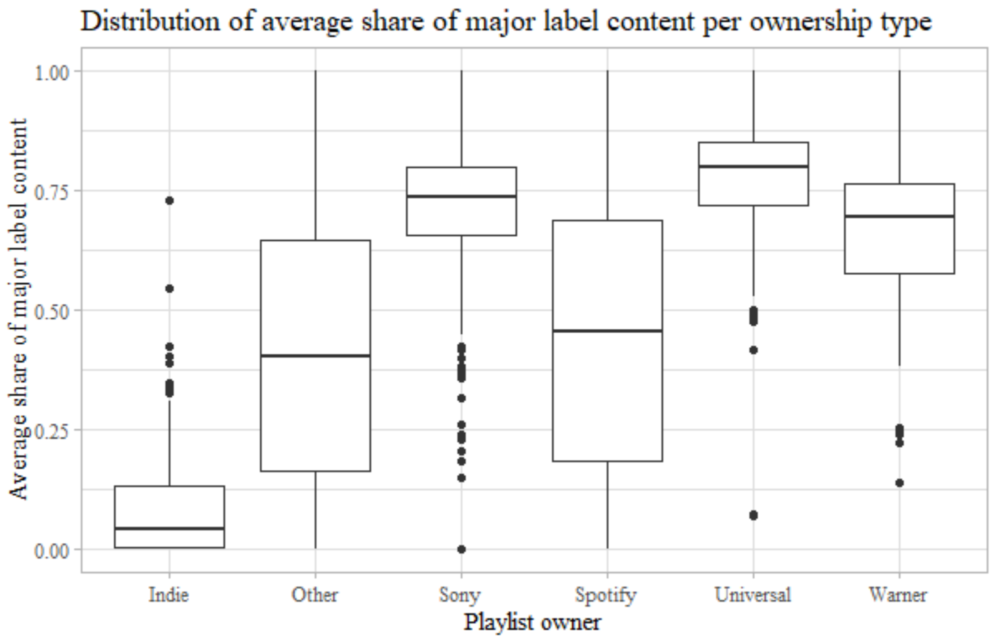
Figure 2



Note: the values within parentheses represent (x, y) values.

For a first understanding of how the share of major label content is distributed, we show the box plots of this distribution for the different ownership types in Figure 3. This figure shows how the average share of major label content at playlists level is distributed. We observe a lot of heterogeneity in the averages for major label owned playlists versus those owned by Spotify, Independent (Indie) labels or others. Playlists owned by Spotify or others seem to have a widely spread distribution, and playlists owned by Indie labels include a very low amount of major label

Figure 3

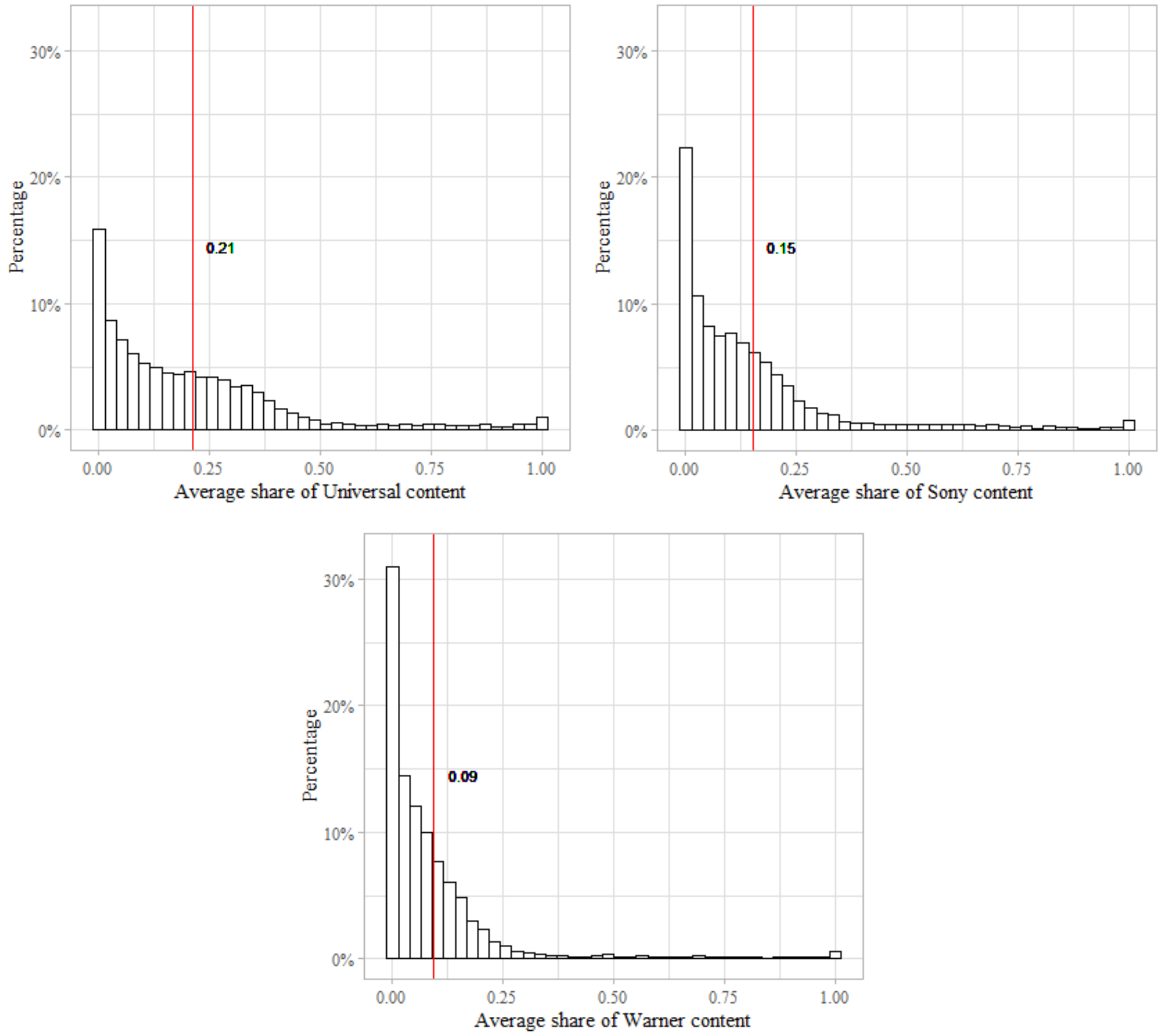


content[4]. The major labels clearly prefer to include major label content, but do leave some space for content released by independent labels. This is expected, since major labels get compensated for the amount of streams their content gets on the platform and therefore include a high level of own tracks in their playlists. Appendix 3 displays the shares of content for each specific major label per ownership type. These figures show that major labels are highly biased towards including their own content. Especially playlists owned by Universal seem to include their own content, while lists owned by Warner seem to do this a bit less. The same probably applies to independent labels, since their share of major label content is rather low. Eventually, the shares of content for each specific major label will be included in our model. Figure 4 therefore documents the variation in the average share of content across playlists for each major label more specifically. For all major labels applies that quite some playlists (15%-30%) have a share of their content which is 0, or close to 0. Universal appears to have the highest share of content over these playlists overall, with a mean of 0.21 and a median of 0.15. The distribution of this label is more spread, compared to Sony and Warner. Warner appears to have the lowest share overall, with a mean of 0.09, median of 0.05, and a distribution that is skewed more to 0. For Sony, the mean is 0.15 and the median is 0.09. Also, there is a higher percentage of playlists that include a lot (0.5-1) of Universal and Sony content compared to that of Warner. This is in line with the figure in appendix 3, which show that Warner content is included less intensive.

[4] Among the top 10,000 playlists, 110 were classified as ownership type that is Indie. Although probably not all independent lists are included, it should still give a representable view of the extent to which they include major label content.

Figure 4

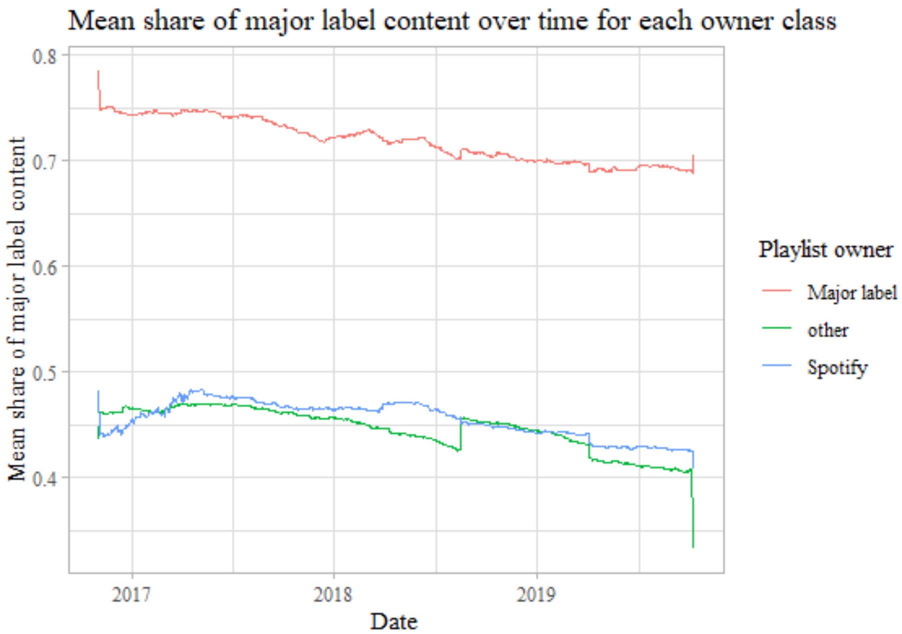
Distribution of the average share of content for each major label



Note: the average share of content for each major label is at playlist-level, i.e. the average of each playlist is taken and the graphs therefore show the variation across playlists.

Figure 5 provides an overview of the share of major label content among the most popular playlists for each owner class over time. It is necessary to say that one should be cautious with drawing implications from Figure 5 and Figure 6, since they contain content of the playlists that were most popular at the end of the period. Data from playlists that were somewhere in this time frame among the most popular, but were not at the end is missing. Also, the influence of Spotify's (new) playlists on the averages is substantial and changes the composition of the most popular playlists. For Spotify, we expect that success is not mainly driven by the characteristics of the playlists, but rather by intensive promotion. Though, we document some of the things we observe. What is observed in both Figure 5 and Figure 6 is that there are substantially higher peaks and lower drops in the first and last days of the data. It appeared that the first day and last three days have a substantially lower number of follower count observations, causing the averages to be unstable. These dates are

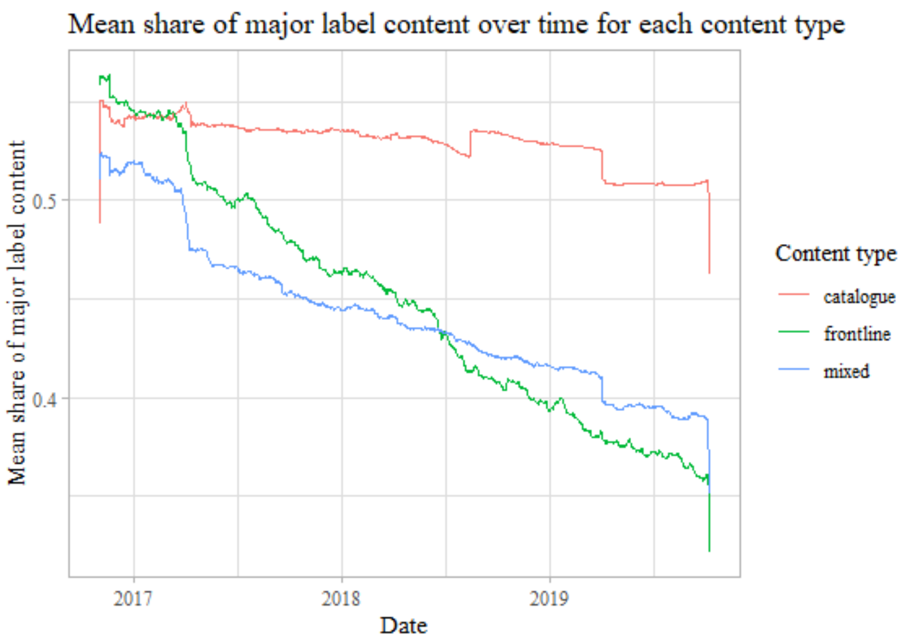
Figure 5



therefore removed from our data in the final analysis. Most noticeable about Figure 5, is that the mean share of major label content is again by far the highest for playlists that are owned by major labels. Also, it looks like the total share of major label content has decreased over the past years, possibly caused by independent labels occupying a larger share of content among the most popular playlists. A more in-depth view on the share of major labels over the past years can be found in appendix 4, where the shares of content for each specific major label is shown.

How the share of major label content changes over time for the most popular playlists is not only interesting to look at for different types of playlist owners, it is also interesting to look how it varies across the types of contents. The content type of a playlist can be frontline (75%-100% of content is younger than 18 months), catalogue (less than 0%-25% of content is younger than 18 months) or mixed (25%-50% of content is younger than 18 months). The share of major label content over time

Figure 6



for each of these content types is shown in Figure 6. When we look at this table, we see that the share of major label content is higher for playlists containing mainly older content and that it might have shifted downwards for playlists containing mainly new content. It might be the case that among the popular playlists, share of major label content went down over time. Meanwhile, the share of major label content for catalogue playlists does not change as much, possibly meaning older popular content is more often from major labels than new content is. Appendix 5 documents the shares of each major per content type.

Finally, summary statistics for all variables that will be included in the final model are provided in Table 1. We again observe that the share of major label content deviates quite much with a standard deviation of 0.310 and a mean of 0.479. Logically, the mean for the averages of major label content of other playlists is the same as the mean for the overall share of major label content. The standard deviation for the averages of the other playlists have a lower standard deviation though, since these values are already the mean of all other playlists at that specific time.

Table 1

Descriptive statistics for the final data set

Statistic	Mean	St. Dev.	Min	Max
Followers	144,301.100	460,964.800	1	24,215,260
Major label share	0.479	0.310	0.000	1.000
Universal share	0.220	0.241	0.000	1.000
Warner share	0.098	0.165	0.000	1.000
Sony share	0.160	0.214	0.000	1.000
Number of tracks	150.710	501.991	1	27,128
Energy	0.640	0.171	0.000	1.000
Speechiness	0.095	0.076	0.000	0.964
Acousticness	0.283	0.220	0.000	0.995
Instrumentalness	0.107	0.198	0.000	1.000
Liveness	0.206	0.073	0.000	0.991
Valence	0.493	0.153	0.000	0.986
Tempo	117.772	10.127	0.000	199.572
Loudness	-7.798	3.847	-49.165	0.000
Mean other playlists Universal	0.220	0.019	0.190	0.245
Mean other playlists Warner	0.098	0.004	0.090	0.113
Mean other playlists Sony	0.160	0.009	0.149	0.198
Active listeners	13,763.950	51,127.630	0.00003	2,523,721.000

Notes: excluded from this overview, but included in the final data set are the time trend and dummy variables for month-of-the-year and day-of-the-week, since their values are rather not informative.

4. Model

The goal of this study is to measure the causal effect of share of major label content on the amount of followers of a playlist. We expect that the relationship between share of major label content and amount of followers is nonlinear, and specify a log-log model. Using the log on both sides of the equation results in the beta's representing elasticities. In particular, the beta's indicate partial elasticity of the amount of followers with respect to the associated independent variable, assuming all other variables remain constant. This model is also known as the constant elasticity model (LaFrance, 1986). Without taking into account fixed variables and other covariates that we introduce later, the basic structure of this model is:

$$\log(Y_{pt}) = \beta^T log(1 + ML_{pt}) + \varepsilon_{pt} \tag{1}$$

where Y_{pt} is the number of followers of playlist p at time t, ML_{pt} is a vector containing the shares of major label content of Universal, Warner and Sony of playlist p at time t, and ε_{pt} is the error.

For a better understanding of how β should be interpreted, it is shown how it is equal to the elasticity of share of major label content. First, we define on the left hand side how the change in the dependent variable coheres with the change in the independent variable and β. Then on the right hand side it is shown how, after rearranging, the expression of β is equal to the elasticity:

$$\frac{\partial Y_{pt}}{Y_{pt}} = \beta \frac{\partial ML_{pt}}{ML_{pt}} \qquad \qquad \beta = \frac{\partial Y_{pt}}{\partial ML_{pt}} \frac{ML_{pt}}{Y_{pt}} = E$$

In our specific case, β thus represents the percentage change in followers from a one percent increase in share of content of the corresponding major label. Using the elasticity allows to give an indication of the popularity for each specific major label. Specifically, this represents the popularity of the contents and artists of the major labels.

4.1 Establishing causal interpretation

The problem with the basic model we specified is that it assumes that the share of major label content is randomly assigned. The share of major label content is not randomly assigned in general, and especially not in panel data. Also, playlists that are owned by major labels might have a higher share of major label content in general, but might also add/remove major labels tracks more often. There may also be unobserved and hence omitted factors that are correlated with both the share of major label content and the amount of followers of a playlist. In this case of unobserved heterogeneity, the estimated effect of share of major label content does not have a causal interpretation. We therefore seek for a method that makes it possible to control for omitted variables.

A model that allows unobserved heterogeneity to be correlated with the independent variable is the fixed effects model. This model is an effective tool to adjust for unobserved time-invariant confounders, but this only applies if the researcher has concerns about time-invariant confounders and there is absence of dynamic causal relationships (Imai & Kim, 2019). In our specific case, the amount of followers are likely more related to time-invariant factors, such as by whom the playlist is owned, rather than dynamic factors. The fixed effects model eliminates the time-invariant unobserved heterogeneity by demeaning the variables using the within transformation.

Therefore, we need a flexible model that includes a rich set of fixed effects to control for these confounding factors. We employ playlist fixed effects to control for persistent differences between playlists that are time-invariant. To control for aggregate trends in the amount of followers of a playlist, we employ a linear time trend. Also, regular and predictable changes in the amount of followers of a playlist might recur in specific parts of a calendar year. To control for this seasonality, dummy variables for eleven months of the year will be added. The same regular and predictable changes might recur on specific days of the week. Therefore, dummy variables for the days of the week will be added to the model. Also, following a playlist does not only depend on the share of major label content of playlist p, but also on this share of the other playlists. Following one playlist might mean that one does not follow certain other playlists. The share of major label content of other playlists will therefore be included in the model as the mean of all other playlists in the dataset for each t. Finally, other controls that vary for each playlist over time that might have an effect on the amount of followers are included to the model. Specifically, these are the musical characteristics of a playlist. To these characteristics belong: the number of tracks, energy, speechiness, acousticness, instrumentalness, liveness, valence, tempo, and loudness. These fixed effects and other controls should capture all predictable factors that influence the amount of followers of a playlist. The remaining variation in the share of major label content is therefore quasi-random with respect to the residual changes in amount of followers of a playlist. Including these fixed effects as discussed above, we derive to the following model:

$$\begin{aligned}\log\left(Y_{pt}\right) = {} & \beta^T log\left(1 + ML_{pt}\right) + \psi^T log\left(1 + ML_{-pt}\right) + \alpha_p + \delta_{time} t \\ & + \gamma_{M(t)} + \gamma_{D(t)} + \eta^T MC_{\mathrm{pt}} + \varepsilon_{pt}\end{aligned} \tag{2}$$

where ML_{-pt} is a vector including share of major label content of competing playlists, α_p is the time-invariant unobserved heterogeneity of playlist p, t is a general linear time trend, $\gamma_{M(t)}$ are month-of-year dummies that captures seasonal effects, $\gamma_{D(t)}$ are day-of-week dummies, MC_{pt} is a vector of controls containing music characteristics at playlist-day level, and ε_{pt} is the idiosyncratic error term. The intercept in this model is playlist specific and captures the playlist fixed effects.

One might argue that the amount of followers also depends on whether a playlist was on promotion at a certain point in time. Unfortunately, this data is not observed and cannot be included in the model. However, not including this variable does not essentially mean that the results will be biased. It is very likely that mostly playlists owned by Spotify are promoted on the platform. It is also very likely that the promotion of playlists does not change intensively over time. The same playlists, owned by Spotify, are therefore continuously promoted, making it a fixed effect that will be eliminated by demeaning the variables of this model. We discuss potential bias implications in section 6.3.

In the case that change of followers is sufficiently predictable, then curators could differentially adapt the share of major label content over time (time-specific endogenous adjustments). However, the lack of prior research in this field makes it plausible that curators do not choose major labels tracks over independent label tracks consciously at a certain point in time.

Because we use data of playlists over time, the observations within a playlist are related to each other. Since there are fixed effects included in our model, heterogeneity in the treatment effects is required in order to justify clustering of standard errors (Abadie, Athey, Imbens, & Wooldridge, 2017). We expect there to be heterogeneity in the treatment effects, so standard errors will be clustered for every playlist when estimating the regression models.

4.2 Other identification strategies

Other identification strategies have been examined and the applicability to our case has been considered. Using instruments are a potential way to estimate the causal effects, but there is a lack of variables that can be used in a manner to predict the share of major labor content. Therefore, there are no potential variables that can be used as an instrument. Difference-in-differences strategies are often used to establish causality in the case where there are e.g. policy changes and where a control and treatment group can be separated. Separating a treatment from a control group is not applicable in our case and this strategy is therefore not operable. Regression discontinuity is often used to establish causality in cases where characteristics of units are close to one another, but they are categorized in different groups by a random cut-off point. This also does not apply to our case, since such cut-off points do not exist in our data. We therefore use the model including fixed effects to estimate the causal effects of share of major label content on the amount of followers of a playlist. As mentioned before, the assumption is made that the share of major label content is not correlated with the idiosyncratic error term. If however, this endogeneity issue is present, the causal interpretation of our model should be reconsidered and one should be careful with generalizing the results.

The potential issues described and other identification strategies being non-applicable makes it very important to check for robustness. A check that will be executed is to run the regression with the amount of active listeners as dependent variable instead of the amount of followers, which might be a better measure according to (Joven, 2018). This variable is the amount of dynamic followers of a playlists divided by the FEL ratio. Also, we compare the results of the fixed effects model with a simpler pooled OLS model that includes dummies for playlist ownership.

5. Analysis and findings

This chapter covers the results of the analysis using the model specified in chapter 4[5]. The first step is to show the results of the restricted model, where we only include the shares of major label content. Next, this model is extended by including all fixed effects and control variables that were introduced last chapter. We finally check the robustness of the estimates as described in section 4.2. Two other robustness checks are added, because new issues are encountered that demand for extra checks.

5.1 The restricted model

Running the restricted log-log model (1), with the number of followers as dependent variable and only the shares of content for each major label as independent variables, leads to negative estimates with highly significant p-values ($p < .001$) for all independent variables. However, standard errors will be biased in case of serial correlation of the idiosyncratic component of the errors and might therefore cause these small p-values. The Durbin-Watson test for serial correlation in panel models shows that there is indeed serial correlation in the idiosyncratic errors ($p < .001$), as shown in appendix 6.1. The standard errors are therefore clustered at playlist level, leading to substantially higher and more realistic standard errors. Still, the estimates for the shares of all major label content are significant ($p < 0.001$ for Universal and Warner, $p = 0.043$ for Sony). The results of the restricted model with and without clustered standard errors are shown in appendix 6.2. When assessing model fit, we observe that this restricted model explains a very small part of the variance in the dependent variable ($R^2 = 0.011$, $R^2_{adj} = 0.010$, $F(3; 2{,}192{,}759) = 8{,}175.843$, $p < 0.001$). The estimates of $\beta^{Universal}$, $\beta^{Sony}$and Warner β^{Warner} suggest a negative elasticity of -0.862, -0.344, and -0.595 respectively, implying that an increase of the share of major label content is related to negative effects on the number of followers. However, we analyze the extended model first before actually drawing any valid conclusions.

5.2 The extended model

5.2.1 Overall model characteristics

The same steps are now executed for the extended model. A first look at the results, before clustering standard errors, shows that the estimates for the shares of major label content are all significant ($p < 0.001$), but smaller in the extended model compared to the restricted model. We see that the competitor share of major labels are significant ($p < 0.001$) for each major label. The linear

[5] Running the model using all available rows in the data set was not possible due to memory constraints. We therefore balanced the data by only keeping the playlists that had observations for all 1,067 days. The results are compared to those of a random sample in section 5.3.2 to make sure the estimates are not biased due to this sampling technique.

time trend (p <0.001) and all month dummies (p <0.001 for all except January and February (p <0.05)) are significant. Weekday doesn't seem to influence the number of followers of playlists, since all values for the day-of-the-week dummies are insignificant. Finally, all control variables that capture musical characteristics and the number of tracks are significant too (p <0.001). The full model results are displayed in appendix 7.1. However, the standard errors are biased and have to be clustered first. The Durbin-Watson test for the extended model suggests serial correlation of the error terms, as displayed in appendix 7.2. We add the results of the model including clustered standard errors at playlist level to appendix 7.1, including a column showing the p-values. These are the results of main interest and will be referred to as the main model or main results. Before analyzing these results, we report that this model predicts 34.5% of the variance in the dependent variable (R^2 = 0.345, R^2_{adj} = 0.344, F(33; 2,192,729) = 35,015.790, p <.001). The F test comparing the fixed effects model with the pooled OLS estimator model, as shown in appendix 7.3, supports the choice for the fixed effects model (p <0.001).

5.2.2 Results

In this section we discuss the results of the extended fixed effects model with clustered standard errors, as shown in appendix 7.1. We observe that clustering the standard errors results in insignificant effects for the share of Universal (p = 0.604) and Warner (p = 0.274) content. We therefore conclude that the elasticities for these record labels are not significantly different from zero. This elasticity is however significant for Sony (p = 0.038) and has a negative value, as the restricted model already indicated. The value of β^{Sony} (-0.283) is however smaller than the restricted model suggested. This makes sense, since the restricted model lacks control variables. We thus conclude a negative elasticity of number of followers in response to share of Sony content in general. The change in number of followers is less strong than the change in share of Sony content itself, because the elasticity is smaller than one. It is however remarkable that β^{Sony} in the restricted model was the only variable that had a substantially higher p-value after clustering compared to the before clustering case in the restricted model, while the p-values of Universal and Warner remained very low (p < 0.001), and that it is the only variable with a significant p-value in the extended model.

While we cannot confirm our hypothesis for Universal and Warner due to insignificant estimates, it cannot be confirmed for Sony either due to another reason. Our hypotheses predicted positive elasticities for the shares of major label content on number of followers, where we found a negative elasticity for Sony. The theory suggested that although the gap between independents and the major labels is closing, the majors still have an advantage. The direct effects of share of major label content has however never been researched before. Therefore, we cannot precisely compare our results to past research and participate in the discussion whether the role of major labels has changed. We can however conclude that on average, an increasing share of Sony content is associated with negative effects on the number of followers of a playlist.

5.2.3 Results of the other variables

Besides reporting the main variables of interest, it is valuable to discuss the estimates of the other variables included in the regression model. As appendix 7.1 shows, the absolute values of the estimates for the competitor shares of content of each major label are quite high, with positive values for $\psi^{Universal}$ (2.347) and ψ^{Warner} (7.867), while this value is negative for ψ^{Sony} (-9.099). All three of these estimates have very small p-values (p < 0.001). These variables are supposed to measure how changes in the share of major label content of other playlists affect the number of followers of the focal playlist. Given that β^{Sony} was negative, it is contrary to our expectations that ψ^{Sony} is also negative. These substantially higher estimates are assumedly a result of very little variation in this variable, because it takes the average content share of all other playlists in the data. We observe that this value has a minimum value of 0.17, 0.14, and 0.09 for Universal, Sony and Warner respectively, while their maximum values are 0.22, 0.18, and 0.11. It is thus not realistic to assume that this variable ever changes by 1% between *t* and *t-1*, which is the indication of the estimate. Also, the values might therefore be overestimated, making it meaningless to draw any implications from them. We run the model without the competitor shares of major label content in section 5.3.3 as extra check to see if the estimates of main interest are robust to these changes.

Next, we report the time trend of 0.001, significant with p < 0.001. This implies a positive linear time trend over the 1,067 observed days in the data. All dummies for month-of-the-year effects are significant at p < 0.05, except for February which has a p-value of 0.098. January, May, June, July, August, September, October and November seem to be months where the number of followers increase higher compared to December, which is the base level in this model. The remaining months seems to be associated with negative changes of the dependent variables compared to December. Recall that all estimates for day-of-the-week dummies were not significant in the case that standard errors were not clustered. After clustering however, some of these dummies are actually significant. More specifically, Tuesday (p < 0.001), Monday (p = 0.035), Wednesday (p < 0.001), and Saturday (p = 0.009) are now significant, and follower changes thus differ from the base level Sunday.

Finally, we observe that only two of the musical characteristics added to the model were significant. $\eta^{valence}$ is -0.362 with a p-value of 0.05. Higher valence therefore has negative influence on the number of followers, implying that lower valence works better for playlist success. This is in line with the finding by Boughanmi et al. (2019), who concluded that successful albums today have low valence. Also, $\eta^{loudness}$ is significant (p = 0.004) with a value of 0.030, implying louder music is associated with higher playlist success. This is in line with the findings by Boughanmi et al. (2019) and Serra et al. (2012), who both concluded that louder music has become more successful.

5.3 Robustness checks

This section outlines the robustness of our results to alternative methods with the purpose of showing how sensitive the results are to changes. The stability is assessed by comparing the estimates and significances of the variables. First, the exact same model is carried out with the number of active listeners as dependent variable instead. Second, we compare the results from last section with the same model on a random sample of the data. Next, the results of the model without competitor shares of major label content are documented. Finally, we compare the results with a simple pooled OLS model.

5.3.1 Active listeners model

Appendix 8 includes the coefficients of our main model and the model with active listeners as dependent variable in order to make comparing convenient. We first see that the main model explains 34,4% (R^2 = 0.345, R^2_{adj} = 0.344) of the variance, while the model with active listeners explains 38.2% (R^2 = 0.383, R^2_{adj} = 0.382) of the variance in the dependent variable. Next, we observe that the estimates of the variables of main interest are quite close to the values in our main results, and that the shares of Universal and Warner content are again not significant. However, this time the share of Sony content is not significant (p = 0.314) as well. The estimates for the competitor shares of major label content are again high and close to the values of the main model. The linear time trend is again significant (p < 0.001) and most of the month-of-the-year and day-of-the-week dummies are too. When looking at the musical characteristics, it is again loudness that seems to have an effect (p = 0.014). This time however, valence is not significant like all other musical characteristics.

Altogether, the results from both models have a lot of similarities. However, the share of Sony content being not significant raises questions. There are a couple of possible explanations that might cause this difference. First, the estimates might not be as stable as we had hoped, implying that the conclusions we draw in section 5.2.2 are not valid after all. Recall that the p-value of this variable was 0.038, indicating there is a present possibility of finding the effects due to chance. On the other hand, followers and listeners are still different concepts, and although they both might tell something about playlist success, it may be different variables that actually affect them. This could also explain the difference in variance explained by the two models. In that case, we can conclude that the share of Sony content affects the number of followers, while it doesn't affect how often the playlist is listened to. The question then becomes which of the two dependent variables best represents playlist success. Answering that question is not part of this study, since both could be valuable dependent on the corresponding purpose. Another approach to identifying why the estimates of both methods vary focusses on how the number of active listeners is measured. This variable was originally not time-varying, but measured as a static number for only a part of the playlists. This variable was operationalized to a varying variable by dividing the number of followers by the FEL ratio, as discussed in chapter 3. This outcome was thus dependent on both the number of followers and the number of listeners at only one point in time. Ideally, this robustness check would be executed with the true number of listeners for each day and playlists, but this information is yet not available. Eventually, it is not clear which option causes the estimates of these two models to differ. The presence of multiple options that potentially drives this difference might however suggest that this discrepancy is not enough evidence to reject the conclusions drawn before.

5.3.2 Random sample model

Recall that the main model included only the data from playlists that had observations for all 1,067 days in the time frame. Since this method of sampling might favor to include a prespecified group of playlists, estimates could be biased. We therefore check for robustness by estimating the same model on a random sample of 2,500 playlists from all available data, this time including also those that have less observations and again with follower count as dependent variable. The results of the random sample model are added to the table in appendix 8.

We first mention that this random sample model explains 39.4% of the variance (R^2 = 0.394, R^2_{adj} = 0.394). When comparing the estimates, we observe that the values for Universal and Warner are again not significantly different from zero. The evidence for the effect of Sony content share is however stronger this time, with an elasticity of -0.428 (p-value = 0.005) instead of -0.283 (p-value = 0.038) from the main model. The estimates for the competitor shares of major label content are all insignificant this time, supporting the decision to draw no conclusions about them prematurely. The linear time trend is again significant just like most of the other time-specific dummies. When we look at the estimates of the musical characteristics, we observe that it is again loudness that is significant, with similar estimates as before. Valence is this time however not significant, which is not surprising given that its p-value in our main model (0.05) is located on the boundary of significance.

Altogether, the results from the random sample model are very similar to those of the main model, confirming that the estimates are trustworthy. The small difference between the estimates is not enough to have doubts about the reliability of the results. By concluding that the estimates are robust, we assume that the differences between the main model and active listeners model are driven by the fact that those concepts are indeed different or that the active listeners amount is not measured the right way. This is a plausible assumption, as we discussed in last paragraph.

5.3.3 Model without competitor shares of major label content

As discussed in section 5.2.3, we run the model without competitor shares of major label content to check whether the estimates are robust to these changes. Appendix 9 includes the results of the main model and the model without competitor shares for comparability. An extensive elaboration of the comparison is not necessary here, because we observe that the estimates are very similar. Again, only β^{Sony} of the variables of main interest is significant (p = 0.018), this time with a coefficient of -0.325. The estimates for the time-specific and other control variables are also very similar, with a few small changes for the time dummies. We observe that the model without competitor shares explains 33.2% of the variance (R^2 = 0.333, R^2_{adj} = 0.332), a bit less than the 34.4% of the main model (R^2 = 0.345, R^2_{adj} = 0.344). We conclude that the estimates are also robust to removing the share of major label content from competitor playlists.

5.3.4 Pooled OLS model

Finally, the results of the main model are compared to a simple pooled OLS model with dummies for playlist ownership in order to give a better interpretation of the estimates. Appendix 10 includes the results of both of these models. One should be very careful with drawing implications from these results, because the standard errors and p-values are likely to be unrealistic. We can however take a look at what the estimates suggest. It is interesting to see how the results change in the pooled OLS model including ownership dummies, with Spotify as base level. The dummies for all major labels have negative estimates (Universal = -1.693, Sony = -1.714, and Warner = -2,067), which is expected due to Spotify's prominent position in the market. In addition, we observe that the estimates for the content shares of all major labels are now positive. This suggests that playlist success is associated with higher major label shares in general when ignoring the panel structure of the data. This makes sense, since we expect popular playlists to contain popular content. The fact that this effect becomes negative in the fixed effects model might raise questions. We explain this by the fact that a part of this effect is captured by the playlist fixed effects in our main model. This part is a certain threshold that popular playlists have for the major label share they include, and differs for each playlist. The variation of major label content share above this threshold seems to have a negative effect on playlist success. The pooled OLS model does not have a causal interpretation because it ignores this aspect, while our fixed effects model does if the assumptions are valid.

5.4 Conclusions

Table 2 gives an overview of the elasticities from all models that are estimated in this study. The pooled OLS model suggests a positive effect for all major labels. However, the coefficients of this model do not have a causal interpretation, because it ignores the panel structure of the data. This

Table 2

Elasticities for all models

	Model					
Label	Restricted	Main	Random sample	Without competitor shares	Pooled OLS	Listeners
Universal	-0.862***	-0.067	-0.183	0.016	0.315***	0.031
Sony	-0.344**	-0.283**	-0.428***	-0.325**	0.170***	-0.162
Warner	-0.595***	-0.141	0.143	-0.148	0.591***	-0.132

Note: recall that the active listeners model included the number of listeners as dependent variable, while all other models had log(followers) as dependent variable.
*p<0.1;**p<0.05;***p<0.01

method suggests that successful playlists in general have higher shares of major label content. This is what might be expected, because the popular content is what makes these playlists successful.

However, the pooled OLS model does not take into that the major label share of a playlists starts at a certain level, different for each playlist. The fixed effect model acknowledges the panel structure and estimates how the remaining variation in the share of major label content influences playlist success. All fixed effects models with followers as dependent variable show stable negative and significant elasticities for Sony, implying that this estimate in our main model is indeed robust. It appears that this effect only applies to Sony, since the elasticities for Universal and Warner are insignificant in all cases, except for the restricted model. These conclusions apply to the playlists that had observations for all days in the time frame and should also apply to the other playlists among the top 10,000, since we find the same results for the random sample. The estimated effect of Sony content has a causal interpretation in case there are no unmeasured confounders. The estimates will be biased in case of unmeasured time-varying variables that have an influence. These potential bias implications are therefore discussed in section 6.3. Future research in this topic is required for more usable implications, and therefore included in section 6.2

6. Discussion

This thesis obtains insights into the general effects that the shares of Universal, Sony, and Warner content have on playlist success among the most popular playlists. More specifically, this thesis aims to provide elasticities that indicate how sensitive the change in number of followers is to changes in the share of major label content. We conclude that this elasticity is not different from zero for Universal and Warner. There is however a significant negative elasticity for the share of Sony content, indicating that adding their content is associated with negative consequences for the number of followers in general. This finding is contrary to the flow of research indicating that major labels still dominate among the most popular content. The current study however does not focus on the popularity of content, but rather on how this content affects playlist success, and therefore captures a different concept that has not been captured in previous research and presumably cannot be directly compared with previous research. We performed multiple robustness checks confirming that our findings in the fixed effects model are indeed valid. However, we do not observe a couple of potential variables that might cause this relationship to be negative. Also, the fact that this study captures the effects over all types of playlists (e.g., genres, owners) make its practical implications obscure, as we discuss next.

6.1 Theoretical and managerial implications

One stream of research suggests that the gap between majors and independent labels might be closed already, based on data about survival on the charts. Another school of thought concludes that even though independent labels are closing the gap, the majors still dominate, based on shares of content among the most popular music. Based on this latter point of view, we expected that the content of each major label would have a positive effect on the success of playlists. Although the current study does not measure the same concepts as previous research, we add to the literature with the finding that the shares of two of the major labels don't affect playlist success and that the share of Sony content negatively affects playlist success in general. We can therefore not conclude that any of our expectations are true. We cannot accept two hypotheses because the effects are insignificant, and cannot accept the remaining hypothesis due to the effects being negative instead of positive. The findings provided in this study do not help to determine whether the gap between independent and major labels is closed or not, but it does provide a first view on how major labels contribute to playlist success, and that this contribution is in general not positive for Sony. Also, we confirm that the musical characteristics that explain album success are applicable to the case of playlists. In line with Boughanmi et al. (2019) and Serra et al. (2012), both loudness and valence were significant, although valence was just on the edge of significance in our main model and was insignificant in the robustness check. However, it still shows that albums and playlists have similar musical characteristics influencing them.

From a practical point of view, you can state that curators should be careful with including Sony content. It is however not that straightforward. It definitely does not mean that curators should not include Sony content in their playlists. Stating that well-considered choices have to be made for including this type of content, although this is always important, might also be too premature. In this study, we focus on the general effects that are present and lack insights that might be more case-specific. There might be huge differences between genres and different playlist owners that we did not look into. We found that major labels are biased towards including their own content for the playlists they possess. There are enough reasons to believe that such differences exist per genre, especially when they become more niche. We did not look into the differences between these types of playlists, and more research is necessary to distinguish the effects of these different types. Therefore, we discuss future research in the next paragraph. It is very feasible for playlist owners to conduct this type of research for their own lists, or let someone help them with it, making potential findings even more relevant.

6.2 Future research

Future research can build upon this study in two different fields. The first one could be of value for the discussion about the gap between major and independent labels. If a better classification of major and independent labels than we discussed in section 3.2.1 can be made, the estimates of these labels can be compared and define what content works best for playlist success. The other direction that future research could possibly go is one that describes the drivers of playlist success better. Before this study, no prior research covered the drivers of playlist success. In both fields, there are a couple of interesting aspects that could be included. First, it would be interesting to see how the effects of major label shares differ per genre. As discussed before, certain genres are likely to include a lot of major label content and others specifically do not include this type of content. Another option would be to see how the effects differ per owner type (e.g. Spotify human curated, Spotify automatically curated, major labels, or independent labels), or even at playlist level. These studies could identify why the general effect is negative and would give more usable and generalizable results.

Another application focusses more on user specific data analysis. This is however only applicable for Spotify, since it is likely that they are the only one owning this type of data. They can perform the method used in this study to find out personal preferences even better and eventually optimize automatically curated personalized playlists.

6.3 Limitations

Finally, we discuss the potential limitations of this study. First of all, there is a lack of previous research in this topic, which makes it likely that there are factors influencing the number of followers that we do not capture. We used a fixed effects model to take care of the time-invariant factors that are part of this problem. However, if there are still unmeasured confounders, the estimates do not have a causal interpretation. The estimates are biased in case there are time-varying factors that influence playlist success we do not measure. Whether a playlist was on promotion or not is one of the factors we do not control for. We assumed that the variation in this variable is limited, as discussed in section 4. Though, it is not ruled out that this variable causes our estimates to be biased. The significant negative relation of Sony and number of followers might be driven by significantly fewer promotion of playlists that are high in Sony content. In this case the estimates do not have a causal interpretation. Moreover, it could also be that playlists owned by Sony, and therefore have higher shares of Sony content, are not performing well on the platform. This can be a consequence of their own management, but also by Spotify treating these playlists in a negative manner. This emphasizes the relevance of future research that is required to find how the effects differ per category. Next, our conclusions apply to the top 10,000 most popular playlists only. There is a large probability that less popular playlists have different characteristics and respond different to changes. Therefore, our conclusions do not apply to those playlists. Furthermore, values for musical characteristics were missing for 13 playlists, as discussed in section 3.2.2. This missing data is not missing at random, since the owner is in all cases Spotify. Because this occurs on only a small fraction of the data, we assume that it does not influence the results. Moreover, the competitor share of major label content was included in our analysis as the mean of all other playlists at that point in time. A lack of variety in this number led to estimates that are hard to interpret and potentially overestimated. We did confirm that the estimates of main interest did not change by removing these competitor shares. However, these values should optimally be included for only a few comparable playlists, which we were not able to realize. Also, we measured playlist success with two different concepts, i.e. the number of followers and the number of listeners. We found different results for both concepts. This raised the question which concept better captures playlist success. We discussed that we did not obtain the time-varying number of listeners, but computed the number of listeners by using the static FEL ratio at the time the data was collected. We therefore kept the number of followers as main measure for playlist success. A final potential for future research would therefore be to obtain estimates for the actual time-varying number of listeners, and investigate the relationship between the number of listeners and followers.

Reference list

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). When Should You Adjust Standard Errors for Clustering. *The National Bureau of Economic Research*, 15:17.

Aguiar, L., & Waldfogel, J. (2016). Even the Losers Get Lucky Sometimes: New Products and the Evolution of Music Quality since Napster. *Information Economics and Policy*, 19, 28.

Aguiar, L., & Waldfogel, J. (2018). Platforms, Promotion, and Product Discovery: Evidence from Spotify Playlists. *The National Bureau of Economic Research*, 26.

Aguiar, L., & Waldfogel, J. (2018). Platforms, Promotion, and Product Discovery: Evidence from Spotify Playlists. *The National Bureau of Economic Research*, 1-26.

Alexander, P. J. (1994). New Technology and Market Structure: Evidence from the Music Recording Industry . *Journal of Cultural Economics*, 113-122.

Amy Watson. (2019, October 18). *Streaming - Statistics & Facts*. Opgehaald van Statista: https://www.statista.com/topics/1594/streaming/

Amy Watson. (2019, December 9). *Subscriber share of music streaming services worldwide 2019*. Opgehaald van Statista: https://www.statista.com/statistics/653926/music-streaming-service-subscriber-share/

Amy Watson. (2020, Februari 13). *Digital and physical revenue market share of the largest record companies worldwide from 2012 to 2018*. Opgehaald van Statista: https://www.statista.com/statistics/422926/record-companies-market-share-worldwide-physical-digital-revenues/

Anderson, C. (2006). *Why The Future Of Business Is Selling Less Of More.* Berkely: Hyperion.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., & Telan, R. (2007). The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts. *Management Science*, 1359-1374.

Boughanmi, K., Ansari, A., & Kohli, R. (2019). Dynamics of Musical Success: A Bayesian Nonparametric Approach. 8-43.

Brynjolfsson, E., & Smith, M. D. (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science*, 580.

Brynjolfsson, E., Hu, Y., & Smith, M. (2003). Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. *Management Science*, 25, 26.

Chartmetric. (2019). *Playlists*. Opgehaald van chartmetric.com: https://app.chartmetric.com/playlist-list/spotify?code2=xx&indie=false&limit=100&offset=0&sortColumn=followers&sortOrderDesc=true

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery. *Marketing Science*, 5-21.

Goodwater. (2019, March 15). *Understanding Spotify: Making Music Through Innovation*. Opgehaald van Goodwatercap: https://www.goodwatercap.com/thesis/understanding-spotify

Guichardaz, R., Bach, L., & Penin, J. (2019). Music industry intermediation in the digital era and the resilience of the Majors' oligopoly: the role of transactional capability. *Industry and Innovation*, 843-869.

Hracs, B. J. (2012). A Creative Industry in Transition: The Rise of Digitally Driven Independent Music Productio. *Growth and Change*, 442-461.

Im, H., Song, H., & Jung, J. (2018). A survival analysis of songs on digital music platform . *Telematics and Informatics*, 1675-1686.

Imai, K., & Kim, I. S. (2019). When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data? *American Journal of Political Science*, 481:484.

Jenkins, C., & Joven, J. (2018, March 13). *Spotify: The Rise of the Contextual Playlist*. Opgehaald van blog.chartmetric.io: https://blog.chartmetric.io/spotify-the-rise-of-the-contextual-playlist-c6f2c26900f4

Joshua P. Friedlander. (2020, Januari 5). *Mid-Year 2019 RIAA Music Revenues Report*. Opgehaald van Recording Industry Association of America (RIAA): http://www.riaa.com/wp-content/uploads/2019/09/Mid-Year-2019-RIAA-Music-Revenues-Report.pdf

Joven, J. (2018, May 22). *blog.chartmetric.io*. Opgehaald van Measuring Attention on Spotify Playlists: Followers vs. Listeners: https://blog.chartmetric.io/measuring-attention-on-spotify-followers-vs-listeners-da0383d90796

LaFrance, J. T. (1986). The structure of Constant Elasticity Demand Models. *American Journal of Agricultural Economics*, August 1.

Maasø, A., & Hagen, A. N. (2019). Metrics and decision-making in music streaming. *The International Journal of Media and Culture*, 18-31.

Paul Lamere . (2014, January 14). *The Zero Button Music Player*. Opgehaald van Music Machinery: https://musicmachinery.com/2014/01/14/the-zero-button-music-player-2/

Ren, J., & Kauffman, R. (2017). UNDERSTANDING MUSIC TRACK POPULARITY IN A SOCIAL NETWORK. *Association for Information Systems*, 374-388.

Serra, J., Corral, Á., Boguñá, M., Haro, M., & Acros, J. L. (2012). Measuring the Evolution of Contemporary Western Popular Music. *Scientific Reports*, 1-6.

Spotify. (2020). *Playlists*. Opgehaald van Artists.spotify.com: https://artists.spotify.com/guide/playlists

Spotify Newsroom. (2020). *Company info*. Opgehaald van newsroom.spotify.com: https://newsroom.spotify.com/company-info/

Strobl, E. A., & Tucker, C. (2000). The Dynamics of Chart Success in the U.K. Pre-Recorded Popular Music Industry. *Journal of Cultural Economics*, 113-130.

The Economist. (2018, May 17). *Spotify kicks off a cultural shift in the music business*. Opgehaald van economist.com: https://www.economist.com/business/2018/05/17/spotify-kicks-off-a-cultural-shift-in-the-music-business

Waldfogel, J. (2015). Digitization and the Quality of New Media Products, The Case of Music. *National Bureau of Economic Research*, 407-442.

Waldfogel, J. (2017). How Digitization Has Created a Golden Age of Music, Movies, Books, and Television. *Journal of Economic Perspectives*, 198-199.

Appendices

Appendix 1: list of variables in the panel data

Variable	Operationalization	Static/dynamic
ID	Unique playlist identifier by Chartmetric.	Static
Date	Date of observation (YYYY-MM-DD).	Dynamic
Followers	Number of followers.	Dynamic
ML share	Share of major label content on playlist (Universal, Warner, and Sony).	Dynamic
Universal share	Share of content on playlist that is owned by Universal.	Dynamic
Warner share	Share of content on playlist that is owned by Warner.	Dynamic
Sony share	Share of content on playlist that is owned by Sony.	Dynamic
Ntracks	Number of tracks on the playlists.	Dynamic
Ntracksotherpl	Average number of playlists that tracks on focal playlist are listed on.	Dynamic
Trackage	Average age of tracks on playlist (number of days).	Dynamic
Energy	Acoustic attribute: average energy of tracks on playlist.	Dynamic
Speechiness	Acoustic attribute: average speechiness of tracks on playlist.	Dynamic
Acousticness	Acoustic attribute: average acousticness of tracks on playlist.	Dynamic
Instrumentalness	Acoustic attribute: average instrumentalness of tracks on playlist.	Dynamic
Liveness	Acoustic attribute: average liveness of tracks on playlist.	Dynamic
Valence	Acoustic attribute: average valence of tracks on playlist.	Dynamic
Tempo	Acoustic attribute: average tempo of tracks on playlist.	Dynamic
loudness	Acoustic attribute: average loudness of tracks on playlist.	Dynamic
Playlist ID	Spotify playlist identifier.	Static
Name	Name of playlist.	Static
Followers	Number of followers of the playlist at snapshot date.	Static
Owner name	Name of the playlist owner.	Static
Owner ID	Identifier of the playlist owner by Chartmetric.	Static
User ID	Spotify user identifier of playlist owner.	Static
From Spotify	Whether the playlist is owned by Spotify or not (TRUE/FALSE).	Static
Content type	Type of content: if 75%-100% is new (younger than 18 months) it is frontline, if 25%-75% it is mixed, if 0%-25% it is catalogue.	Static
Active ratio	Number of listeners divided by followers at snapshot date.	Static
Owner class	The owner of the playlist, classified as Spotify, major label or other.	Static
Rank	Rank of playlists at snapshot date, based on number of followers.	Static

Note: “Active ratio” seems to be incorrect and is ignored in this study.

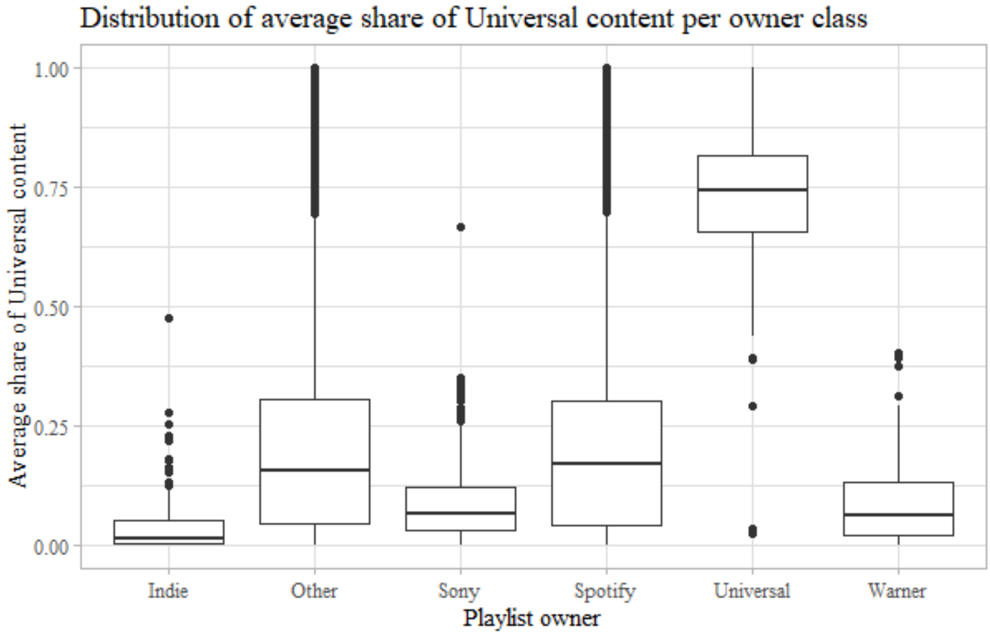
Appendix 2: list of variables in the playlists data

Variable	Operationalization
Retrieval unix	Timestamp (Unix) of data retrieval.
Position	Rank at time of retrieval, based on number of followers.
ID	Unique playlist identifier by Chartmetric.
Location	Locale of playlist, only available for playlists made by Spotify.
Playlists ID	Spotify playlist identifier.
Name	Name of the playlist.
Personalized	Whether the playlist can be personalized for a user (TRUE/FALSE).
Followers	The amount of followers at time of data retrieval.
Owner name	Name of the playlist owner.
Owner ID	ID of the playlist owner (Chartmetric).
User id	Spotify user identifier of playlist owner.
From Spotify	Whether the playlist is owned by Spotify or not (TRUE/FALSE).
Genre	Collection of genres that playlist belongs to.
Monthly listeners	Number of monthly listeners of the playlist at data retrieval.
Listeners-to-followers ratio	Number of listeners divided by the number of followers, both at data retrieval.
Content type	Type of content: if 75%-100% is new (younger than 18 months) it is frontline, if 25%-75% it is mixed, if 0%-25% it is catalogue.
Active ratio	Number of listeners divided by followers at snapshot date.
From major label	Whether the playlists is owned by a major label or not (TRUE/FALSE).
Owner class	The owner of the playlist, classified as Spotify, major label or other.

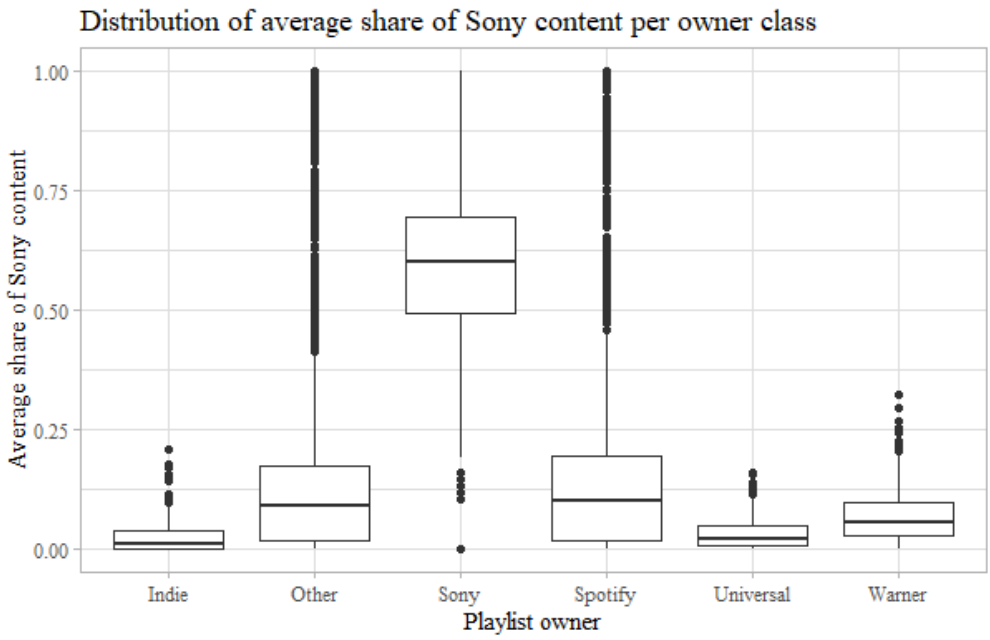
Note: both “listeners-to-followers ratio” and “active ratio” should represent the same thing, but both values are different and do not correspond to the number of listeners divided by the number of followers and will therefore be ignored.

Appendix 3: distribution of average share of Universal, Sony and Warner content per ownership type

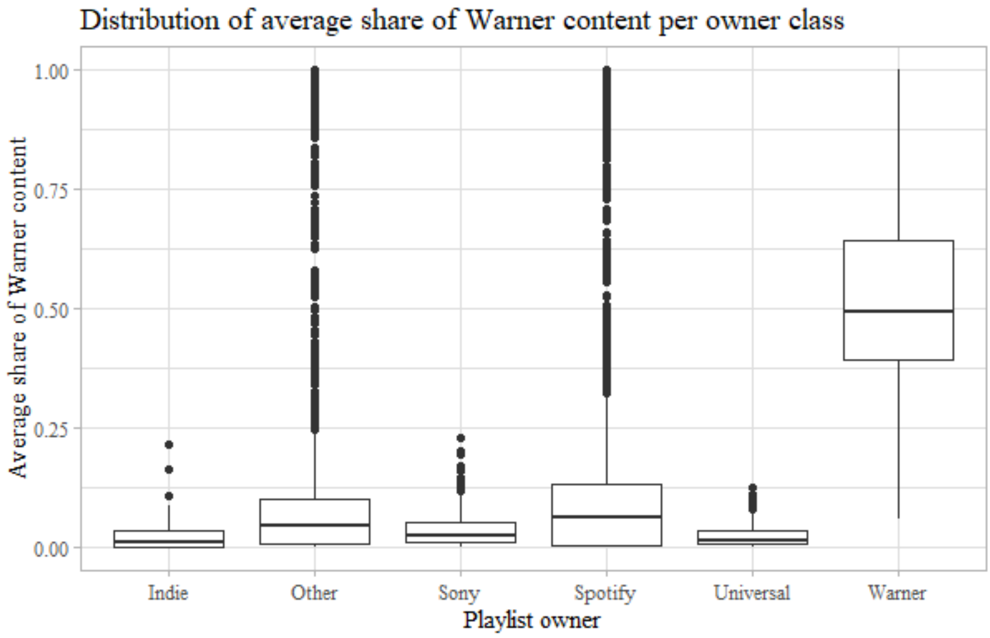
Appendix 3.1: Universal



Appendix 3.2: Sony

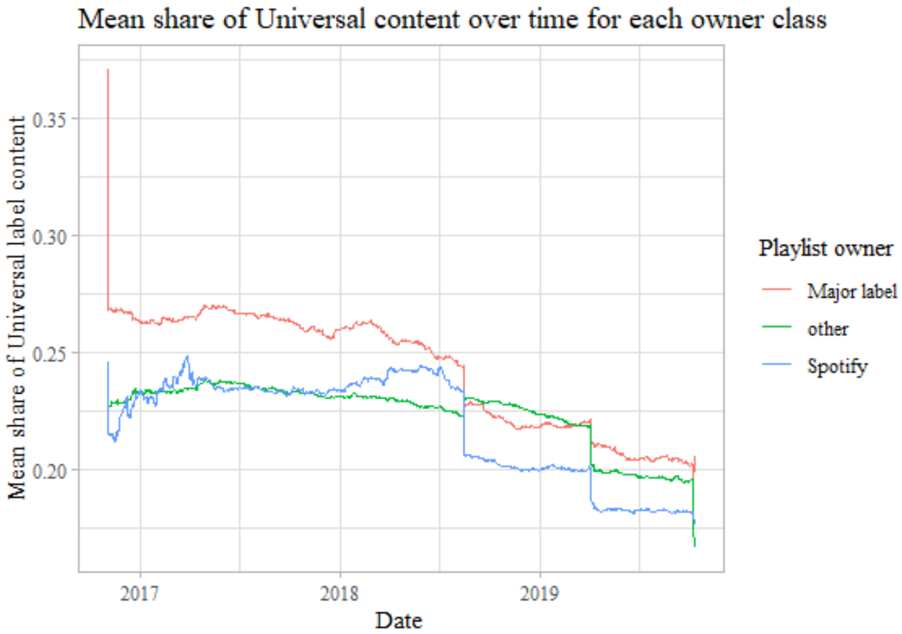


Appendix 3.3: Warner

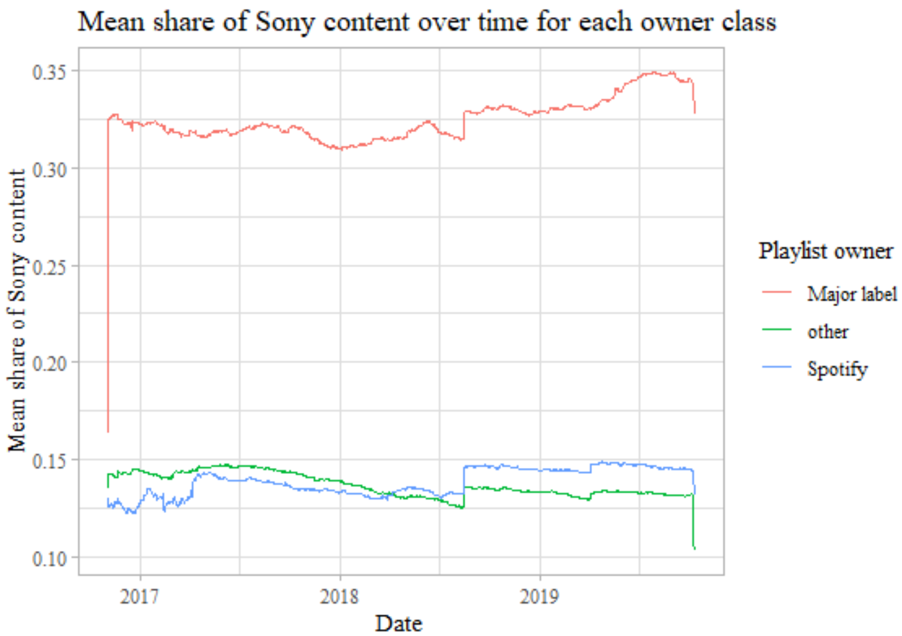


Appendix 4: share of each major label's content per owner class

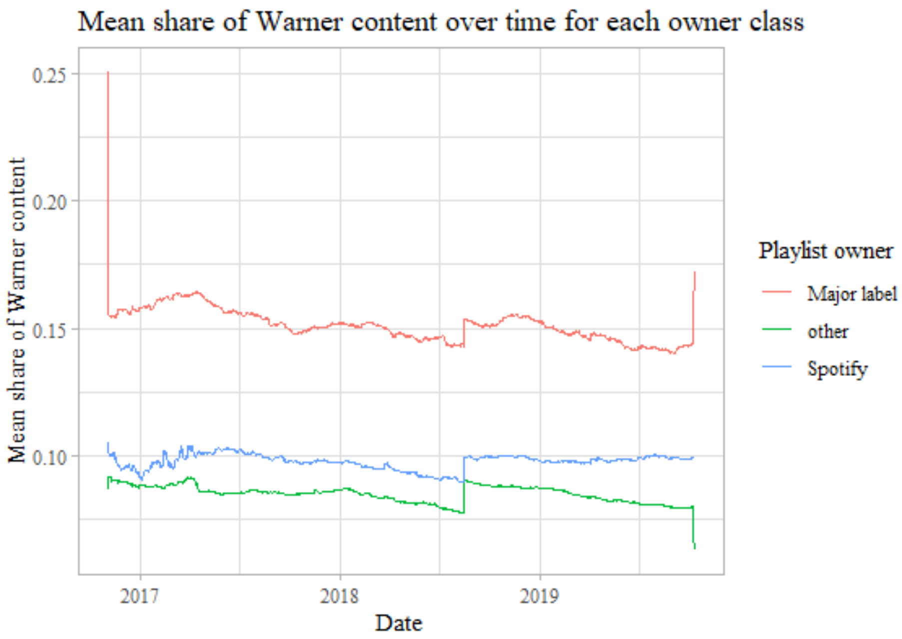
Appendix 4.1: Universal



Appendix 4.2: Sony

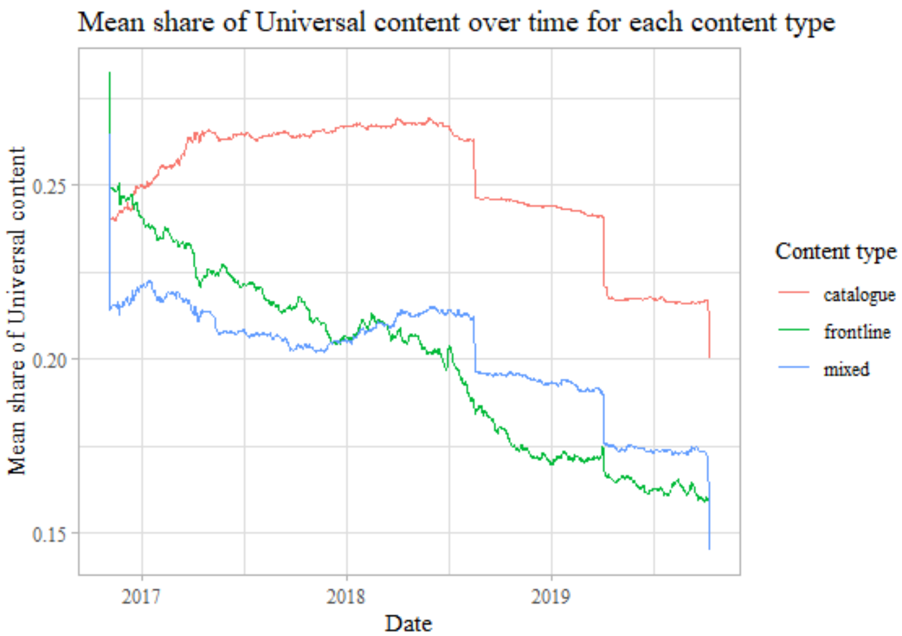


Appendix 4.3: Warner

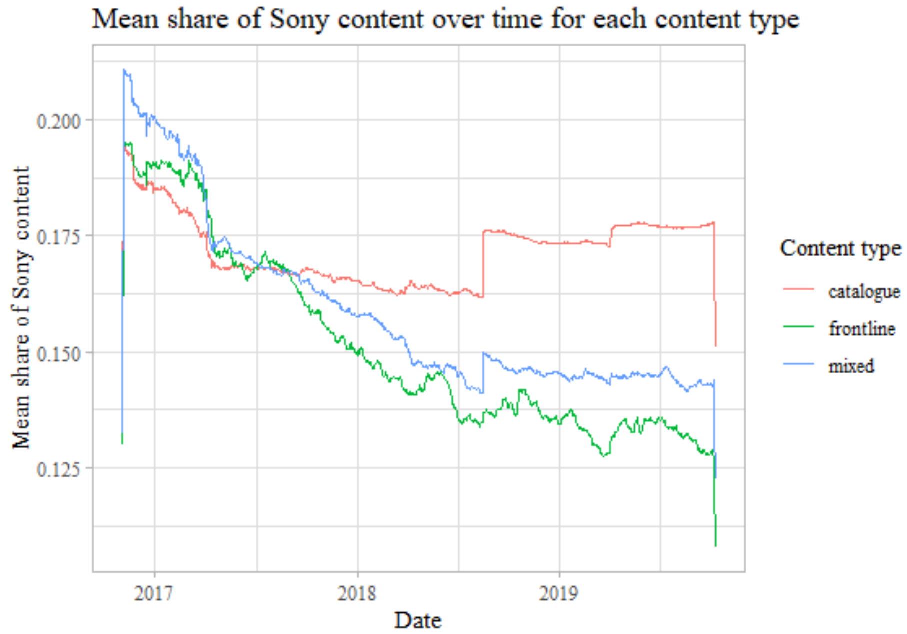


Appendix 5: share of each major label's content per content type

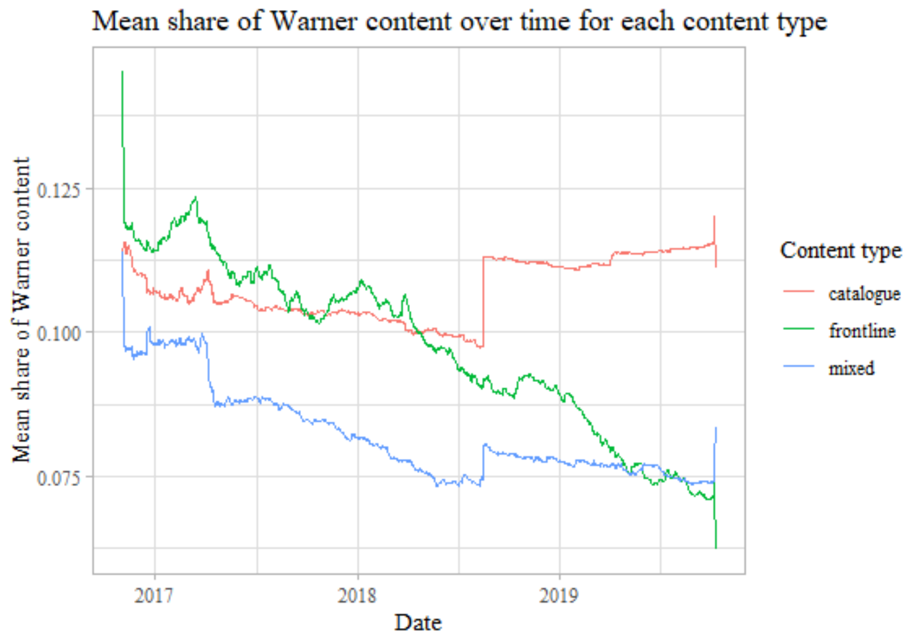
Appendix 5.1: Universal



Appendix 5.2: Sony



Appendix 5.3: Warner



Appendix 6: restricted model

Appendix 6.1: Durbin-Watson test for the restricted model

Durbin-Watson test for serial correlation in panel models

data: log(followers) ~ log(1 + Universal share) + log(1 + Sony share) + log(1 + Warner share)
DW = 0.011556, p-value < 2.2e-16

alternative hypothesis: serial correlation in idiosyncratic errors

Appendix 6.2: Results of the restricted model

	Dependent variable:		
	Log(Followers)		
	Without clustered SEs	With clustered SEs	With clustered SEs and p-values
Log(1 + Universal share)	-0.862^{***}	-0.862^{***}	-0.862^{***}
	(0.006)	(0.160)	p = 0.00000
Log(1 + Sony share)	-0.344^{***}	-0.344^{**}	-0.344^{**}
	(0.007)	(0.170)	p = 0.043
Log(1 + Warner share)	-0.595^{***}	-0.595^{***}	-0.595^{***}
	(0.008)	(0.166)	p = 0.0004
Observations	2,194,819		
R^2	0.011		
Adjusted R^2	0.010		
F Statistic	$8,175.843^{***}$ (df = 3; 2192759)		

Note: in the columns "Without clustered SEs" and "With clustered SEs" the value between brackets contains the standard error. In the final column "With clustered SEs and p-values" the standard errors are left out and the p-values are shown.

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

Appendix 7: the extended model
Appendix 7.1: results of the extended model

	Dependent variable:		
	Log(Followers)		
	Without clustered SEs	With clustered SEs	With clustered SEs and p-values shown.
Log(1 + Universal share)	-0.067^{***}	-0.067	-0.067
	(0.005)	(0.129)	p = 0.604
Log(1 + Sony share)	-0.283^{***}	-0.283^{**}	-0.283^{**}
	(0.006)	(0.136)	p = 0.038
Log(1 + Warner share)	-0.141^{***}	-0.141	-0.141
	(0.007)	(0.128)	p = 0.274
Log(1 + Universal share other playlists)	2.347^{***}	2.347^{***}	2.347^{***}
	(0.099)	(0.431)	p = 0.00000
Log(1 + Sony share other playlists)	-9.099^{***}	-9.099^{***}	-9.099^{***}
	(0.114)	(0.697)	p = 0.000
Log(1 + Warner share other playlists)	7.867^{***}	7.867^{***}	7.867^{***}
	(0.194)	(0.579)	p = 0.000
Linear time trend	0.001^{***}	0.001^{***}	0.001^{***}
	(0.00001)	(0.00004)	p = 0.000
Month April	-0.007^{***}	-0.007^{***}	-0.007^{***}
	(0.001)	(0.003)	p = 0.010
Month August	0.012^{***}	0.012^{***}	0.012^{***}
	(0.001)	(0.003)	p = 0.0003
Month February	-0.003^{**}	-0.003^{*}	-0.003^{*}
	(0.001)	(0.002)	p = 0.098
Month January	0.003^{**}	0.003^{**}	0.003^{**}
	(0.001)	(0.001)	p = 0.046

Month July	0.008***	0.008**	0.008**
	(0.001)	(0.003)	p = 0.028
Month June	0.008***	0.008**	0.008**
	(0.001)	(0.003)	p = 0.020
Month March	-0.010***	-0.010***	-0.010***
	(0.001)	(0.002)	p = 0.00002
Month May	0.007***	0.007**	0.007**
	(0.001)	(0.003)	p = 0.041
Month November	0.006***	0.006***	0.006***
	(0.001)	(0.001)	p = 0.00001
Month October	0.022***	0.022***	0.022***
	(0.001)	(0.002)	p = 0.000
Month September	0.012***	0.012***	0.012***
	(0.001)	(0.003)	p = 0.0001
Tuesday	-0.0003	-0.0003***	-0.0003***
	(0.001)	(0.0001)	p = 0.00000
Thursday	0.00003	0.00003	0.00003
	(0.001)	(0.0001)	p = 0.712
Monday	0.0001	0.0001**	0.0001**
	(0.001)	(0.00003)	p = 0.035
Friday	0.00000	0.00000	0.00000
	(0.001)	(0.0001)	p = 0.998
Wednesday	-0.0004	-0.0004***	-0.0004***
	(0.001)	(0.0001)	p = 0.000
Saturday	-0.0002	-0.0002***	-0.0002***
	(0.001)	(0.0001)	p = 0.009
Number of tracks	0.00001***	0.00001	0.00001
	(0.00000)	(0.00002)	p = 0.371

Energy	-0.274***	-0.274	-0.274
	(0.012)	(0.280)	p = 0.330
Speechiness	-0.179***	-0.179	-0.179
	(0.016)	(0.367)	p = 0.626
Acousticness	-0.116***	-0.116	-0.116
	(0.008)	(0.180)	p = 0.520
Instrumentalness	-0.110***	-0.110	-0.110
	(0.007)	(0.168)	p = 0.514
Liveness	-0.352***	-0.352	-0.352
	(0.010)	(0.236)	p = 0.137
Valence	-0.362***	-0.362**	-0.362**
	(0.008)	(0.185)	p = 0.050
Tempo	-0.001***	-0.001	-0.001
	(0.0001)	(0.001)	p = 0.590
Loudness	0.030***	0.030***	0.030***
	(0.0005)	(0.010)	p = 0.004
Observations	2,194,819		
R^2	0.345		
Adjusted R^2	0.344		
F Statistic	35,015.790*** (df = 33; 2192729)		

Note: in the columns "Without clustered SEs" and "With clustered SEs" the value between brackets contains the standard error. In the final column "With clustered SEs and p-values" the standard errors are left out and the p-values are shown.

*p<0.1; **p<0.05; ***p<0.01

Appendix 7.2: Durbin-Watson test for the extended model

Durbin-Watson test for serial correlation in panel models

data: log(followers) ~ log(1 + Universal share) + log(1+ Sony share) + log(1+ Warner share) + log(1+ Universal share other playlists) + log(1+ Sony share other playlists) + (1+ Warner share othe r playlists) + Time trend + Month + Weekday + Number of tracks + Energy + Speechiness + Aco usticness + Instrumentalness + Liveness + Valence + Tempo + Loudness
DW = 0.011465, p-value < 2.2e-16

alternative hypothesis: serial correlation in idiosyncratic errors

Appendix 7.3: F test comparing the fixed effects model and the pooled OLS model

F test for individual effects

log(followers) ~ log(1 + Universal share) + log(1+ Sony share) + log(1+ Warner share) + log(1+ Universal share other playlists) + log(1+ Sony share other playlists) + (1+ Warner share other playlists) + Time trend + Month + Weekday + Number of tracks + Energy + Speechiness + Acousticness + Instrumentalness + Liveness + Valence + Tempo + Loudness
F = 11013, df1 = 2056, df2 = 2192729, p-value < 2.2e-16

alternative hypothesis: significant effects

Appendix 8: results of the main model, active listeners model, and random sample models

	Dependent variable:		
	Log(followers)	log(Active listeners)	Log(followers)
	Main model	Active listeners model	Random sample model
Log(1 + Universal share)	-0.067	0.031	-0.183
	p = 0.604	p = 0.836	p = 0.136
Log(1 + Sony share)	-0.283**	-0.162	-0.428***
	p = 0.038	p = 0.314	p = 0.005
Log(1 + Warner share)	-0.141	-0.132	0.143
	p = 0.274	p = 0.453	p = 0.372
Log(1 + Universal share other playlists)	2.347***	2.642***	1.877*
	p = 0.00000	p = 0.00001	p = 0.053
Log(1 + Sony share other playlists)	-9.099***	-8.676***	0.046
	p = 0.000	p = 0.000	p = 0.980
Log(1 + Warner share other playlists)	7.867***	7.786***	-0.099
	p = 0.000	p = 0.000	p = 0.950
Linear time trend	0.001***	0.001***	0.001***
	p = 0.000	p = 0.000	p = 0.000
Month April	-0.007***	-0.006	0.009**
	p = 0.010	p = 0.121	p = 0.036
Month August	0.012***	0.017***	0.023***
	p = 0.0003	p = 0.00005	p = 0.00002
Month February	-0.003*	-0.004	0.018***
	p = 0.098	p = 0.124	p = 0.001
Month January	0.003**	0.001	0.019***
	p = 0.046	p = 0.768	p = 0.001
Month July	0.008**	0.013***	0.0001
	p = 0.028	p = 0.005	p = 0.984
Month June	0.008**	0.013***	-0.007

	p = 0.020	p = 0.004	p = 0.114
Month March	-0.010***	-0.010***	0.011**
	p = 0.00002	p = 0.001	p = 0.028
Month May	0.007**	0.010**	-0.003
	p = 0.041	p = 0.022	p = 0.393
Month November	0.006***	0.010***	0.027***
	p = 0.00001	p = 0.000	p = 0.00001
Month October	0.022***	0.026***	0.016***
	p = 0.000	p = 0.000	p = 0.008
Month September	0.012***	0.016***	0.009
	p = 0.0001	p = 0.00002	p = 0.108
Tuesday	-0.0003***	-0.0003***	-0.00004
	p = 0.00000	p = 0.0002	p = 0.792
Thursday	0.00003	0.0001	0.0003***
	p = 0.712	p = 0.591	p = 0.010
Monday	0.0001**	0.0001	-0.0002
	p = 0.035	p = 0.141	p = 0.329
Friday	0.00000	0.00004	-0.0002**
	p = 0.998	p = 0.744	p = 0.040
Wednesday	-0.0004***	-0.0004***	-0.001***
	p = 0.000	p = 0.00001	p = 0.0001
Saturday	-0.0002***	-0.0004***	-0.0002
	p = 0.009	p = 0.003	p = 0.165
Number of tracks	0.00001	0.00003*	-0.00000
	p = 0.371	p = 0.075	p = 0.915
Energy	-0.274	-0.516	-0.403
	p = 0.330	p = 0.148	p = 0.104
Speechiness	-0.179	-0.264	0.178
	p = 0.626	p = 0.539	p = 0.583

Acousticness	-0.116	0.129	0.323*
	p = 0.520	p = 0.560	p = 0.074
Instrumentalness	-0.110	-0.049	-0.219
	p = 0.514	p = 0.811	p = 0.146
Liveness	-0.352	-0.368	-0.005
	p = 0.137	p = 0.275	p = 0.984
Valence	-0.362**	-0.285	-0.337*
	p = 0.050	p = 0.271	p = 0.086
Tempo	-0.001	0.0002	0.0001
	p = 0.590	p = 0.886	p = 0.953
Loudness	0.030***	0.038***	0.038***
	p = 0.004	p = 0.008	p = 0.0002
Observations	2,194,819	1,410,574	1,774,950
R^2	0.345	0.383	0.394
Adjusted R^2	0.344	0.382	0.394
F Statistic	35,015.790*** (df = 33; 2192729)	26,477.200*** (df = 33; 1409219)	34,986.050*** (df = 33; 1772420)

Note: all p-values in this table are after clustering the standard errors.
*p<0.1; **p<0.05; ***p<0.01

Appendix 9: results of the main model and model without competitor shares of major label content

	Dependent variable:	
	Log(followers)	
	Main model	Model without competitor shares
Log(1 + Universal share)	-0.067	0.016
	p = 0.604	p = 0.904
Log(1 + Sony share)	-0.283**	-0.325**
	p = 0.038	p = 0.018
Log(1 + Warner share)	-0.141	-0.148
	p = 0.274	p = 0.255
Log(1 + Universal share other playlists)	2.347***	
	p = 0.00000	
Log(1 + Sony share other playlists)	-9.099***	
	p = 0.000	
Log(1 + Warner share other playlists)	7.867***	
	p = 0.000	
Linear time trend	0.001***	0.001***
	p = 0.000	p = 0.000
Month April	-0.007***	0.017***
	p = 0.010	p = 0.000
Month August	0.012***	0.020***
	p = 0.0003	p = 0.000
Month February	-0.003*	0.012***
	p = 0.098	p = 0.000
Month January	0.003**	0.006***
	p = 0.046	p = 0.00000
Month July	0.008**	0.023***
	p = 0.028	p = 0.000
Month June	0.008**	0.023***
	p = 0.020	p = 0.000
Month March	-0.010***	0.016***
	p = 0.00002	p = 0.000
Month May	0.007**	0.021***
	p = 0.041	p = 0.000
Month November	0.006***	-0.001

	p = 0.00001	p = 0.278
Month October	0.022***	0.042***
	p = 0.000	p = 0.000
Month September	0.012***	0.013***
	p = 0.0001	p = 0.00001
Tuesday	-0.0003***	0.0003***
	p = 0.00000	p = 0.000
Thursday	0.00003	0.001***
	p = 0.712	p = 0.000
Monday	0.0001**	0.001***
	p = 0.035	p = 0.000
Friday	0.00000	-0.00000
	p = 0.998	p = 0.940
Wednesday	-0.0004***	0.001***
	p = 0.000	p = 0.000
Saturday	-0.0002***	-0.0003***
	p = 0.009	p = 0.0002
Number of tracks	0.00001	0.00001
	p = 0.371	p = 0.683
Energy	-0.274	-0.568**
	p = 0.330	p = 0.045
Speechiness	-0.179	-0.233
	p = 0.626	p = 0.524
Acousticness	-0.116	-0.241
	p = 0.520	p = 0.196
Instrumentalness	-0.110	-0.059
	p = 0.514	p = 0.730
Liveness	-0.352	-0.407*
	p = 0.137	p = 0.092
Valence	-0.362**	-0.381**
	p = 0.050	p = 0.044
Tempo	-0.001	-0.001
	p = 0.590	p = 0.264
Loudness	0.030***	0.044***
	p = 0.004	p = 0.0001
Observations	2,194,819	2,194,819
R^2	0.345	0.333
Adjusted R^2	0.344	0.332

F Statistic	35,015.790*** (df = 33; 2192729)	36,466.760*** (df = 30; 2192732)

Note: all p-values in this table are after clustering the standard errors.
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Appendix 10: Results of the main model and pooled OLS model with playlist ownership dummies

	Dependent variable:	
	Log(followers)	
	Main model	Pooled OLS model with playlist ownership
Log(1 + Universal share)	-0.067	0.315***
	p = 0.604	p = 0.000
Log(1 + Sony share)	-0.283**	0.170***
	p = 0.038	p = 0.000
Log(1 + Warner share)	-0.141	0.591***
	p = 0.274	p = 0.000
Log(1 + Universal share other playlists)	2.347***	2.478***
	p = 0.00000	p = 0.000
Log(1 + Sony share other playlists)	-9.099***	-8.048***
	p = 0.000	p = 0.000
Log(1 + Warner share other playlists)	7.867***	7.907***
	p = 0.000	p = 0.000
Linear time trend	0.001***	0.001***
	p = 0.000	p = 0.000
Month April	-0.007***	-0.004
	p = 0.010	p = 0.225
Month August	0.012***	0.014***
	p = 0.0003	p = 0.0001
Month February	-0.003*	-0.003
	p = 0.098	p = 0.379
Month January	0.003**	0.002
	p = 0.046	p = 0.608
Month July	0.008**	0.011***
	p = 0.028	p = 0.003
Month June	0.008**	0.012***
	p = 0.020	p = 0.002
Month March	-0.010***	-0.010***
	p = 0.00002	p = 0.005
Month May	0.007**	0.010***
	p = 0.041	p = 0.007

Month November	0.006***	0.005
	p = 0.00001	p = 0.149
Month October	0.022***	0.022***
	p = 0.000	p = 0.000
Month September	0.012***	0.012***
	p = 0.0001	p = 0.001
Tuesday	-0.0003***	-0.0002
	p = 0.00000	p = 0.939
Thursday	0.00003	0.0001
	p = 0.712	p = 0.970
Monday	0.0001**	0.0001
	p = 0.035	p = 0.972
Friday	0.00000	0.0001
	p = 0.998	p = 0.963
Wednesday	-0.0004***	-0.0003
	p = 0.000	p = 0.907
Saturday	-0.0002***	-0.0002
	p = 0.009	p = 0.940
Number of tracks	0.00001	-0.00004***
	p = 0.371	p = 0.000
Energy	-0.274	0.105***
	p = 0.330	p = 0.000
Speechiness	-0.179	0.276***
	p = 0.626	p = 0.000
Acousticness	-0.116	-0.036***
	p = 0.520	p = 0.00004
Instrumentalness	-0.110	-0.009
	p = 0.514	p = 0.178
Liveness	-0.352	-0.989***
	p = 0.137	p = 0.000
Valence	-0.362**	-0.902***
	p = 0.050	p = 0.000
Tempo	-0.001	-0.005***
	p = 0.590	p = 0.000
Loudness	0.030***	0.042***
	p = 0.004	p = 0.000
Owner is Indie		-2.061***
		p = 0.000
Owner is other		-2.162***

		p = 0.000
Owner is Sony		-1.714***
		p = 0.000
Owner is Universal		-1.693***
		p = 0.000
Owner is Warner		-2.067***
		p = 0.000
Intercept		13.484***
		p = 0.000
Observations	2,194,819	2,194,819
R^2	0.345	0.343
Adjusted R^2	0.344	0.343
Residual Std. Error		1.062 (df = 2194780)
F Statistic	35,015.790*** (df = 33; 2192729)	30,210.280*** (df = 38; 2194780)

Note: Spotify is the base level of the playlist ownership dummies in the pooled OLS model.
*p<0.1;**p<0.05;***p<0.01