An American bias in a local music streaming market?

Marianne Lumeau[1], François Moreau[2], Patrik Wikström[3,4], Ola Haampland[4], Rune Johannessen[4]

[1]University of Rennes, CREM-CNRS, France; [2]University Sorbonne Paris Nord, CEPN-CNRS, France; [3]Queensland University of Technology, Australia, [4]INN University, Norway;

March 2023

Abstract. In cultural industries, the fear of large-country hegemon has been revived with the growth of subscription services (such as Netflix, Spotify or Apple Music) and the concentration of these markets. The aim of this paper is to provide evidence on this issue from music streaming consumption. Using a unique dataset that encompasses both streaming and download consumption in France, a dynamic local music market representative of the global evolution, our results are as follows. At the top end, we find no difference between streaming and download consumption based on the recording artist's geographical origin. Outside the top of the distribution, however, we observe that the US market share increases in the streaming model to the detriment of local (French) music. This advantage is even stronger for the US back catalogue (i.e., songs released at least ten years ago): the market share of a US song from the back catalogue is 26% higher when listened to on a streaming platform compared to when purchased as a download. These results are economically significant as non-chart and back catalogue songs cumulate a significant market share. From a second dataset of all the streams from a sample of French subscribers to a European music streaming platform, we show that the implementation of algorithmic recommendations by streaming platforms to guide users contributes to explain the dominance of US back catalogue. Altogether, these results document that streaming platforms introduce an American bias in music consumption.

1. Introduction

In entertainment industries, fears associated with a large-economy hegemon, especially an American dominance, are not new (Ferreira and Waldfogel, 2013) and motivate to consider cultural goods and services as exceptions in international treaties.[1] Many countries have adopted instruments to protect their local culture and domestic content, such as quotas in broadcasting (Richardson, 2006; Richardson and Wilkie, 2015). More recently, the growth of subscription services such as Netflix (for video) or Spotify and Apple Music (for music) revives the fears of an American hegemon among industry actors and policymakers (Aguiar and Waldfogel, 2018, 2021).[2]

In this context, this paper aims at examining the relationship between streaming consumption and the geographical origin of music. First, we document whether the fear of an over-representation of US content on the steaming platforms is entirely founded in smaller domestic markets. Second, we identify whether the American hegemon is a bias, i.e. whether American and local songs are treated differently by the streaming platforms. These questions are economically relevant as in the recorded music industry most revenue comes from subscriptions to audio streaming platforms (IFPI, 2022). Moreover, as in many other online markets, the audio streaming market is highly concentrated (Bourreau and Perrot, 2020; Lenard, 2019; Parker and Van Alstyne, 2005). Powerful network effects, economies of scale, the development of huge catalogues, and attractive functionalities mean that Spotify and Apple Music have come to (mostly) dominate this global market. Consequently, concerns have been raised about the market power of these platforms and their impact on the music industry (Aguiar and Waldfogel, 2021; Aguiar et al., 2021; Bourreau and Gaudin, 2022; Mariuzzo and Ormosi, 2022).

As far as we know, the study by Ferreira and Waldfogel (2013) is the first attempt to investigate how digitization impacts the market share of local and US content. They report that digitization has not reduced, but instead increased the market share of local content. With respect to streaming, George and Peukert (2016) report that YouTube consumption favors both local and US content. However, Aguiar et al. (2018) contradict these findings, and conclude that although local music benefits less from consumption on Spotify compared to physical sales or downloads, the US repertoire does not benefit in particular. It should be noted that all of these results were obtained by analyzing the very top end of the consumption distribution; the “charts” (the most successful

[1] See for instance the cultural exception clause in the GATT/WTO introduced in 1993, the cultural exemption clause in the NAFTA included in 1994, the UNESCO’s Convention on the protection and Promotion of the Diversity of Cultural Expressions adopted in 2005.

[2] See, for instance, the report published by the French and Quebec Ministries of Culture: https://www.culture.gouv.fr/en/Thematiques/Europe-et-international/Decouvrabilite-en-ligne-des-contenus-culturels-francophones

100 or 200 songs), which now represent a much smaller share of market revenue compared to the CD era (Anderson, 2006; Brynjolfsson et al., 2011). Since the market share of chart music is declining, neglecting consumption outside the chart could mean that any conclusions are unrepresentative of the full distribution.

Identifying a possible hegemon of US content requires to compare local and US market shares on streaming platforms to different consumption channels over the same period. We have collected a unique dataset from music streaming subscribers and from downloads purchasers on the main online platforms in France during 2017. The French market is a dynamic music market that is characterized by a significant local production and a high level of protection of cultural diversity.[3] France is the fifth-largest music market in the world (IFPI, 2022) and could be considered as representative of all digital markets (see sub-section 2.1). In 2017, music consumptions on streaming and download platforms were significant, even if premium streaming generates most revenue for the industry (see sub-section 2.1). The dominant platforms were Spotify and Deezer for streaming and iTunes for download and catalogs in all dominant platforms were huge and similar. The dataset is composed of almost all the consumption, including the charts (the top end of the distribution) and the non-charts (the bottom of the distribution). This allows us to improve the literature by analyzing the whole distribution of consumption.

Our results are the following. First, at the top end of the distribution, local (French) content aggregates a greater market share than US content, in both streaming and ownership models. This result is consistent with Ferreira and Waldfogel (2013), and George and Peukert (2016). Second, outside the top end of the distribution, we observe that the US catalogue, and especially songs from the back catalogue (i.e., released more than ten years ago), perform better in the streaming model than in the download model, to the detriment of the local content. We estimate that, compared to a French song, the market share of a US song from the back catalogue is 26% higher when listened to on a streaming platform compared to when purchasing the song as a download. It should be noticed that these results are economically significant: non-chart consumption cumulates 69% of the market shares and the back catalogue accounts for 20% of the consumption on premium streaming platforms. Altogether, these results highlight the need to study non-chart consumption in the streaming era. They also stress the importance of the back catalogue, which is in line with its growing economic value.[4]

The second aim of this paper is to study whether the better performance of the American content on streaming platforms is a bias, i.e. whether American and local content are treated differently.

[3] France is the fifth-largest music market in the world (IFPI, 2022) and could be considered as representative of all streaming markets. Sub-section 2.1 presents this market in detail.
[4] Recently, several back catalogues of famous artists and songwriters have been purchased for hundreds of millions of dollars. This point is explored in detail in sub-section 2.2.

Streaming platforms have developed functionalities to promote songs and guide subscribers to choose among huge local and international catalogues: playlists and algorithmic recommendations. If an American bias exists, American songs should be more pushed than local songs in playlists and algorithm recommendations. An alternative explanation of the hegemon of American repertoire could be the relative consumers' preference for this content: the US (back) catalogue experiences more repeat listening because it better fits users' preferences. Then, the better performance of the US songs could be explained by the business model of streaming platforms. While only the purchase of a song is counted in the ownership model (on vinyl, as a CD or a download), each song that is listened to generates revenue in the streaming model. If US songs are played more than local songs, the former will generate more market share in the streaming model than in the ownership model.

To the best of our knowledge, the literature has never disentangled the "preference pull" and the "curation push" effects. Aguiar and Waldfogel (2021) are the only authors who have examined the geographical origin of songs in platforms' mechanisms, specifically the composition of playlists available on Spotify. The latter study concluded that the major playlists provided by Spotify tend to favor US content. It should be noted that playlists are only a small part of the music consumption pushed on streaming platforms (4.4% of the consumption in our data – see sub-section 3.2) as compared to algorithmic recommendations (19.1%). A robust analysis of the role of the streaming platforms should include streams from playlists and algorithmic recommendations and discriminate their respective impact.

How to test the existence of a bias toward American content in playlists and algorithmic recommendations and disentangle the curation push and the preference pull effect? An explicit approach would be to know how these mechanisms are designed, but this information is not available. An alternative – implicit – method could be to compare, for each song, the percentage of streams that have been listened to after being push and the percentage of streams that have been listened to after an autonomous decision according to its geographical origin. This method allows us to control for a potential endogeneity issue: a song could be push by the platform because it fits users' preferences. As compared to a local song, if we observe that an American song is more often listened to autonomously than in a context of playlists or recommendations, this suggests that the hegemon of American songs from back catalogue is explained by a relative preference for this type of songs. Alternatively, if the difference between "push streams" and "autonomous choices" is higher for American than French songs, this means that playlists and algorithmic recommendations are bias toward American content.

To disentangle "preference pull" and "curation push" effects, we use another original dataset based on all the streams of 5,500 French subscribers to one of the main European music streaming platforms in 2019 (about 15 million streams). This dataset offers information at the user level on the context of streams, notably whether each stream was played following a platform algorithmic recommendation or as part of a playlist generated by the platform or recorded music labels (push), or following an autonomous search or within a user-generated playlist (pull). We find that compared to French songs, the US repertoire (and especially the US back catalogue) is streamed more when it is pushed compared to when it is pulled. More precisely, this result holds only for algorithmic recommendations and not for playlists. While we do not presume that this is the intention of streaming platforms[5], our result suggests that these algorithmic tools create a bias that favors large-country content to the detriment of local content. It should be noted that this bias is particularly harmful in a market dominated by few dominant platforms.

These results naturally contribute to the nascent literature on the impact of playlists (Aguiar and Waldfogel, 2021; Aguiar et al., 2021) by extending the analysis to another platform (beyond Spotify) and mechanisms that all streaming platforms develop to guide consumption (algorithmic recommendations). More generally, they also contribute to the literature on the market power of dominant platforms, and the possible bias introduced by algorithms, notably gender discrimination (Lambrecht and Tucker, 2019) or ideological polarization (Levy, 2021). Finally, these results rise questions about the competition between music suppliers (artists, labels) on streaming platforms as two suppliers with different nationalities receive a different treatment from platforms and between distribution channels as French radios are subject to local music quotas whereas streaming platforms are not.

Our paper is organized as follows. In the next section, we outline the background on the economics of streaming. We then describe our datasets, before presenting our results on the impacts of streaming on local content, and the role of recommendation systems and curated playlists. Finally, we discuss their implications and offer some conclusions.

2. The economics of streaming

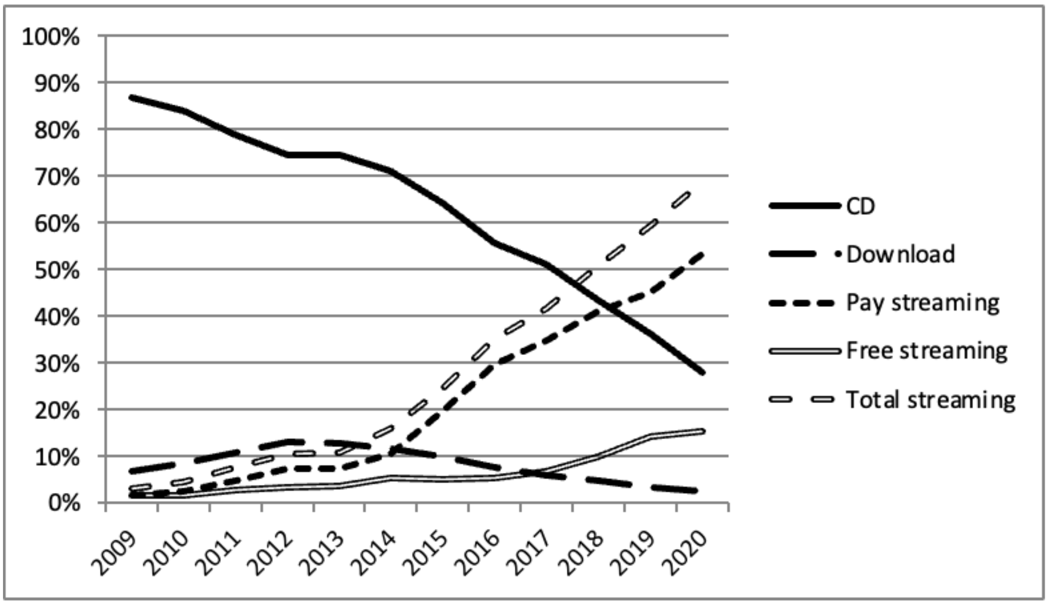
2.1 Streaming revenue in the French recorded music market

As in most countries, streaming has gradually become the dominant way to consume recorded music in France between 2009 and 2020. Over this period, the CD market share fell to 28% (down from 87%), while the streaming market share increased to about 70% (up from 3%). The streaming market is actually made up of three different models: (i) the subscription model

[5] For instance, in a global market that is dominated by a few international platforms, a song that comes from a large country, especially the US, is more likely to fit with the preferences of a large audience and thus to be selected by the algorithms.

(Spotify, Apple Music, Deezer, etc.) in which users pay a monthly flat rate for unlimited access to a vast catalogue of tens of millions of songs without advertising; (ii) the free model in which users have free access to the same catalogue, but with advertising and without some functionalities; and (iii) the video model (with YouTube as the dominant player) in which users have free access to millions of music videos with advertising. The latter two free models accounted for 22% of streaming revenue in 2020, while the lion's share (78%) went to the premium, paid-for model. A fourth component in the digital recorded music market is the download model, in which consumers purchase a digital copy of an album or a song. While dominant until 2013, the download model has now been superseded by the streaming model, as illustrated in Figure 1.

***Figure 1 – The increase in streaming revenue in the French recorded music market 2009–2020* (*Source: Snep*).**



2.2 Characteristics of the streaming model

Digitization of the legal distribution of recorded music began with iTunes, originally a download platform. This was followed by the streaming model, in the form of platforms such as Spotify and YouTube. While download and streaming models share some similarities, there are also some clear differences. The main similarity is the huge catalogue, which is comparable on all platforms, while the main difference is the business model. On the download market (and the physical market), consumers purchase an immaterial (material) product containing the recording, and must pay for each additional song (or album) they wish to play. Conversely, on the streaming market, consumers pay a monthly fee (or accept advertising), and have unlimited access to the entire catalogue, along with extra functionalities. In the context of an economic analysis, the change from an ownership model (CD/download) to an access model (streaming) has several important consequences: (1) the zero marginal cost of accessing music; (2) the way revenue is generated; (3) the role of the medium (songs released between two and ten years ago) and back catalogue (songs released more than ten years ago); and (4) the increasing role played by playlists and recommender systems in consumer choices. These four characteristics are discussed below.

Firstly, differences in the marginal cost of accessing music are a key factor in consumer choices. In the ownership model, the marginal cost is positive: a consumer is expected to purchase a song or an album only if the *ex ante* utility is high enough to cover the unit price. In the streaming model, users can access any song, including those with very low *ex ante* utility, which favors exploration-like behavior (Datta et al., 2018). In practice, it is easier to discover new music on a streaming platform, especially songs with low *ex ante* and high *ex post* utility, compared to a download platform.

A second important difference between ownership and access models relates to revenue generation. In the ownership model, only the purchase matters, and the number of times the consumer actually listens to the digital file (or the CD) has no impact on revenue. In other words, only the *ex ante* utility matters, and not the *ex post* utility—only the number of purchases has an impact on revenue generation. Conversely, in the streaming market, *ex post* utility does matter. The income that a song generates depends not only on the number of users who listen to it, but also the number of times each user listens to it. A song with low *ex ante* and high *ex post* utility will be listened to again and again, and will generate more revenue than in the ownership model (in which it would probably never, or only marginally, have been downloaded). Conversely, a song with high *ex ante*, but low *ex post* utility will probably be listened to just once in the access model, whereas it would have been purchased in the download model. Hence, in the ownership model a disappointed consumer generates the same income as a fully-satisfied consumer, whereas the latter generates more revenue in the access model.

Thirdly, this difference in how revenue is generated can have a significant impact on the medium and back catalogue. In the ownership model, a song purchased years ago and still listened to on a regular basis by an individual does not generate any revenue beyond the initial purchase. Conversely, the same song continues to generate revenue in the access model. Hence, successful songs from the back catalogue are expected to play a more important role in the access model. For instance, Waldfogel (2012) shows that the quality of western music peaked in the 1960s and 1970s. The importance of the back catalogue in the access model is illustrated by recent acquisitions of discographies by private equity firms (e.g., KKR & Co), specialized start-ups (e.g., Hipgnosis Songs Fund Ltd) or even recorded music companies. By early 2021, Hipgnosis had spent £1.2bn buying the back catalogues of famous artists and songwriters—from the Colombian pop star Shakira to the US folk-rock star Neil Young.[6] KKR & Co paid $200m for a majority stake in the catalogue of Ryan Tedder, a songwriter for popular acts such as Beyonce, Lady Gaga, Adele, Paul McCartney, Stevie Wonder and U2.[7] Likewise, in December 2020, the iconic folk-rock star Bob Dylan sold the rights to his entire songwriting catalogue (about 600 songs) to Universal Music Group.[8] In all of these cases the purchasers bet that the value of the best-selling tracks would generate reliable returns for investors, in a context of the growing popularity of streaming services. According to the funder of Hipgnosis, "great, proven songs have predictable, reliable income. It is better than gold or oil".[9]

Fourthly, the streaming model is associated with the growing importance of playlists and recommender systems. As Aguiar and Waldfogel (2021) underline, playlists have two broad functions: they are potentially informative lists of songs, and they provide utilities for playing the songs on these lists. They may be created by individuals, major record labels (through brands such as Digster, Topsify, and Filtr), or streaming platforms such as Spotify or Apple Music. Streaming platforms offer both curated and general chart-based algorithmic playlists, along with lists that are personalized to the individual user. The most influential ones are human-curated, and created by the streaming platform (Aguiar and Waldfogel, 2021). Recommender systems take several forms: *collaborative-filtering systems* (algorithmic recommendations are based on the preferences of similar users) or *content-based filtering systems* (the consumer is recommended products that are similar to those they already like).

Listening to a playlist or following an algorithmic recommendation from the platform is very different to autonomous or 'organic' listening, as the user does not actively choose the songs, only the genre, mood, situation, etc. Hence, a curated song frequently turns out to have low *ex post* utility, which is acceptable since its marginal cost is close to zero. Playlists and recommender systems now constitute a significant share of total listening time on streaming platforms. According to Spotify's CEO, in 2018, over 30% of consumption on the platform was a direct result of recommendations made by its own algorithms and curation teams.[10] Aguiar and Waldfogel (2021) show that being included in a playlist significantly impacts the number of times a song is streamed.

---

[6] https://www.bloomberg.com/news/articles/2021-01-13/shakira-s-145-song-catalog-latest-deal-for-hipgnosis-fund
[7] https://www.reuters.com/article/us-kkr-ryan-tedder/kkr-bets-200-million-on-onerepublic-frontman-ryan-tedders-catalog-idUSKBN29G1FI
[8] https://www.nytimes.com/2020/12/07/arts/music/bob-dylan-universal-music.html
[9] https://www.ft.com/content/71c2be62-b823-47d9-9f43-ab322883aa8c
[10] https://www.musicbusinessworldwide.com/is-the-power-of-the-streaming-playlist-on-the-wane/

2.3 Relationship to the existing literature

This paper is relevant to various streams of the literature. Firstly, it connects with the literature that explores the impact of digitization on international trade in cultural goods; this research is illustrated by Aguiar and Waldfogel (2018), who examined whether the expansion of Netflix favors US hegemony. The latter study showed that Netflix catalogues in small markets are dominated by US content, but noted that this advantage remains small compared to theatrical distribution.

Other relevant findings are reported in Ferreira and Waldfogel (2013), who studied consumption of recorded music in the pre-streaming period. The latter authors analyzed weekly charts in 22 countries over the period 1960–2007, and the study provided evidence that the share of local content had increased, rather than decreased, in the digital era. In 2007, the average market share of local content in weekly charts had increased to 70%, compared to 50% in the 1980s. The authors argued that this change was mainly due to digitization, the rise of TV music channels tailored to each country and, to a lesser extent, local music quotas on radio.

There are only a few studies of the streaming model, and results are conflicting. In Germany, George and Peukert (2016) were able to run a natural experiment following the removal of musical content from YouTube for several months due to a conflict with rights' holders. As musical tastes are similar in Germany and Austria, the authors were able to compare sales in the two countries during the ban; the study measured the impact of video streaming and, in particular, the geographic origin of music. The results showed that video streaming favored both US and local content, and the authors concluded that streaming did not threaten the local repertoire. However, in a later study, Aguiar et al. (2018) focused on audio rather than video streaming, and reached a different conclusion. The authors examined Spotify's weekly charts in 17 countries over the period 2014–2015, along with data on the best-selling recorded music (all delivery channels) over the same period. They concluded that streaming was less favorable to local content than other channels (physical sales, downloads). However, the beneficiary was not US content, but content from small countries.

It should be noted that these papers were based on an analysis of the top of the sales distribution, while significant revenue is now generated from other consumption. For example, Way et al. (2020), a team of researchers working for Spotify, analyzed all of the platform's consumption data in the 79 countries in which it was operating over the period 2014–2019, and concluded that the choice of local content increased over this period. However, the latter analysis did not investigate individual countries, and did not compare streaming with another consumption model. Our paper addresses this gap, and makes a major contribution to the literature by analyzing the overall distribution of consumption in both download and streaming models, which now generate most revenue.

Secondly, our paper is related to the literature that investigates the impact of recommendation tools on consumer behavior and the distribution of sales. Beginning with Bakos (1997), a growing body of literature has examined the impact of recommendation systems on consumer behavior. Belleflamme and Peitz (2018) provide a useful overview of theoretical and empirical works dedicated to recommendation systems and, more broadly, reviews and ratings. For instance, a positive impact of best-seller lists on sales has been identified (e.g., Salganik et al., 2006). Similarly, Fleder and Hosanagar (2009) showed that recommender systems tend to promote the most popular products, which reduces sales diversity at an aggregate level, although the latter authors argued that individual diversity could increase. In line with these findings, other empirical studies have provided some evidence of the impact of streaming platforms on the discovery of new products, but do not explore the specific role of curation (e.g., Datta et al., 2018).

More recently, some papers have examined a potential bias in the recommendations made by platforms (intentional or not). While several theoretical works have explored whether streaming platforms could have an incentive to bias their recommendations (e.g., Bourreau and Gaudin, 2022; Hagiu and Jullien, 2011), there are few empirical studies. One exception is Edelman (2011), who studied whether Google biased its search results in favor of its own interests. In the music industry, Aguiar et al. (2021) found evidence of a platform bias that favored independent labels, while Mariuzzo and Ormosi (2022) argued that major labels have disproportionate access to platform-generated playlists. Much closer to our approach, the empirical study by Aguiar and Waldfogel (2021) showed that inclusion on a playlist had a positive impact on the success of a song, and that leading playlists seemed to be biased toward US content. However, the authors did not disentangle a possible preference for US content from a real bias. Furthermore, their analysis only focused on playlists, and ignored all other (algorithmic) recommendations that a consumer could potentially follow. Our paper overcomes these issues by disentangle the role of the consumers' preferences and the role of all mechanisms implemented by streaming platforms (playlists and algorithmic recommendations).

In the same vein, our paper is related to the literature that studies bias introduced by algorithms. Earlier work has documented that algorithms can lead to gender (e.g., Datta et al., 2015; Lambrecht and Tucker, 2019) or race (e.g., Obermeyer et al., 2019) discrimination. Other authors have studied the role of social network algorithms in media consumption, and the findings indicate that algorithms tend to limit user exposure to news from a different ideology, and therefore favor ideological polarization (e.g., Levy, 2021). Our study naturally complements these papers by documenting a new algorithmic bias: as compared to local content, algorithmic recommendations favor US (old) songs on streaming platforms.

Finally, our paper is linked, albeit marginally, to the literature that explores the role of service quality as a dominant driver for repeat purchase behavior (Zeithaml et al., 1996; Paul et al., 2009).

3. Data

We investigate our research questions using two datasets. The first allows us to study the existence of an American hegemon on streaming platforms by comparing consumption via the ownership model (download) with consumption via the access model (streaming) on a weekly basis at the song level. More specifically, this dataset enables us to study the correlation between the success of a song and its geographical origin in both models in the same country during the same period (sub-section 3.1). After checking for the robustness of our analysis, by showing that differences between the population of streamers and downloaders do not explain our main result, we draw upon a second dataset to study if observed differences between the two models could be due to bias introduced in the recommender systems and playlists provided by streaming platforms. This second dataset includes at a user level all of the streams listened to by a sample of several thousands of French subscribers to one of the main European streaming platforms. In particular, it provides the context for each stream, and allows us to distinguish between streams that have been curated and those that have been autonomously selected by the user (sub-section 3.2). This allows us to control for potential endogeneity issue as a song could be pushed because it fits users' preferences.

3.1. Consumption data for access and ownership models

The first dataset consists of digital sales for 36 weeks in 2017 in France. Almost comprehensive weekly sales data, at the song or album level, was provided by the GfK Group. The company tracks music sales from all distribution models in France, and information is collected from the main digital providers (Spotify, iTunes, Apple Music, Deezer, Napster, Wimp, Microsoft, YouTube, etc.). For confidentiality purposes, data were aggregated by distribution model. We selected consumption data for two models: download and premium audio streaming (subscription). In both models, major platforms are characterized by a similar huge catalogue (approximatively 30 millions of songs in 2017).

For the download model, the dataset consists of weekly, detailed sales information at the song or album level. For the streaming model, only songs that were streamed at least 100 times in a single week were included. It should be noted that in the streaming model all streams refer to a song, while in the download model, a download may refer to a song or an album. In the latter case, we transformed sales data for the downloaded album into data at the song level (Last.fm[11] was used to find out which songs were included on an album).

The GfK database provides several variables for each triplet (song, week, distribution model): the name of the artist, the name of the song, the provider (Universal Music Group, Sony Music Group, Warner Music Group, Believe, etc.), and the musical genre (the main genres were Pop, Rock, Urban, French variety, and ElectroDance). We were also able to build a variable that measured the volume of an artist's sales via downloads over the period 2006–2014 (an index of their past success). Finally, we also collected the date of release for each song, based on the year indicated in its ISRC.[12] Three categories were developed to represent the vintage of a release: less than two years (known as the frontline catalogue in the music industry), from two to ten years (the medium catalogue), and more than ten years (the back catalogue). Much time was dedicated to the laborious task of checking the spelling of the name of each artist and song, and we developed automatic algorithms to clean the data and identify artists who appeared several times under different names (e.g., BOB MARLEY, MARLEY BOB, B. MARLEY).

After cleaning, the database contained 23,539 distinct artists, 82,362 distinct songs, and 260 providers. However, it did not provide one crucial piece of information: the country of origin of the artist that released the song. We thus used Musicbrainz, an online database that collects music metadata[13], to document the nationality of each artist (individual or group). This database was consulted in February 2020 and, at that time, it contained the nationality of 754,903 artists worldwide. Nationality was determined from three alternative pieces of information: the area (available for 735,659 artists), nationality at birth (available for 232,154 artists), and nationality at death (38,122 artists). Area indicates the area an artist is primarily associated with, and it is often, but not always, their country of birth or childhood.[14] Information from these three sources was highly consistent. For instance, among the 213,683 artists for whom we had both the area and nationality at birth, nationality was the same for 92% of them. We thus gave priority to the area

[11] Last.fm (https://www.last.fm) is both a web radio and a website that provides statistics and a recommendation system.

[12] The International Standard Recording Code (ISRC) is an international code that is used to uniquely identify sound recordings. It is associated with a recording, and not the work (lyrics and music). Hence different recordings, edits, and remixes of the same work should each have their own ISRC. However, recordings that are remastered without significant audio changes usually retain their ISRC (source: Wikipedia). As a further check, we also noted the date of the first week the song appeared in the GfK database. The comparison found no significant difference.

[13] See https://musicbrainz.org

[14] See https://musicbrainz.org/doc/Artist#Area

(which was available for 97.4% of artists). When this was not available, we considered nationality at birth and, as a last resort, the country of death.

Finally, the algorithm we had developed to clean the GfK database was run to ensure that the spelling of names in the Musicbrainz database was the same. We then matched the two databases, and the process ended with a manual identification of the most successful artists in each model for whom the country of origin remained unknown. Wikipedia, along with a few other sources, was used to identify their nationality, which allowed us to add another 525 artists to our database. Following this processing step, 83.6% of the songs in our initial database could be associated with a country, which represented 89% of entries (some songs appeared several times in different channels or different weeks) and 98.3% of sales volume. In the remainder of this paper, we consider three countries of origin (France, the US and the United Kingdom) that together accounted for 77.1% of all observations. The remaining countries were labelled “Other countries”.

Our final database contained 2,440,965 observations pertaining to 14,466 artists, 68,834 songs, and 240 providers. Table 1 indicates the average number of songs for the two models (download, streaming) along with the mean, standard deviation, and minimum and maximum weekly consumption by song (a sale for the download model, a stream for the streaming model). Table A1 (see the Appendix) provides descriptive statistics for the main variables used in the quantitative analysis (see section 4).

Table 1 – Database 1.

	Observations	**Average number of distinct songs per week**	**Weekly volume per song**			
			Mean	**SD**	**Min.**	**Max.**
Download	983,755	27,327	14.95	109.41	1	18,165
Streaming	1,457,210	40,478	10,627.04	46,812.89	100	4,591,098

3.2 Data on the streaming context

The second dataset contains information regarding the consumption of French users of one of the main European music streaming platforms in 2019. A total of 5,500 subscribers were randomly selected, and the content of all of their streams over a period of six months in 2019 was collected at the user level. This database contains more than 15 million streams that pertain to an average of 300,000 songs per month. To be consistent with the first dataset, we aggregated all monthly streams at the song level, and focused on songs for which the artist’s nationality was known.[15] Table 2 indicates that there were around 150,000 songs per month in the final dataset, which

[15] As nationality information was not provided by the streaming platform, we repeated the process used with the previous dataset to identify the nationality of each artist.

accounted for about 80% of the total streaming volume (on average, songs for which the artist's nationality was unknown were less popular).

Table 2 – Database 2.

	Observations	Average number of distinct songs per month	Monthly volume per song			
			Mean	**SD**	**Min.**	**Max.**
Streaming	944,297	157,382.8	13.64	80.53	1	15,347

In addition to the nationality of the artist (US, UK, France or Other countries), we identified the context for each stream, either curated or on-demand. *Curated* refers to streams that follow an algorithmic recommendation from the platform (19.2% of all streams) or that were included on a playlist that was built either by an audio streaming platform or by a record company (4.4%). *On-demand* refers to streams that followed an autonomous search on the platform (48.1%) or that were included on a playlist created by the user (28.3%). Hence, for each song, we were able to calculate the share of streams that were curated (*push*), and the share of true music on-demand (*pull*) content.

Furthermore, for each song, we collected information about the musical genre (classical, jazz, R&B, pop, rock, urban, electro, other, unknown), the popularity of the artist measured as their monthly rank (in the top 100, from 101 to 1,000, from 1,001 to 10,000, above 10,000), the type of provider (major label, independent label, digital aggregator), and the total monthly volume of streams. Table A2 (see the Appendix) provides descriptive statistics for the main variables that were used in the quantitative analysis of this second dataset.

4. The geographical origin of songs in the streaming and download models

4.1 Descriptive analysis

While most studies of the impact of digitization on international trade in music examine a dataset that is limited to the charts (e.g., Ferreira and Waldfogel, 2013; Aguiar et al., 2018), Table 3 shows that these charts only account for a small share of music sales. Moreover, the share is even smaller for streaming compared to download sales. More specifically, Top 200 sales in premium streaming account for less than a quarter of all music sales in this model. As a comparison, Feirrera and Walfogel (2013) report that the Top 40 corresponded to over 98% of the recorded music market in 2003. Analyzing what happens beyond the charts is thus more crucial than ever in the streaming age.

Given that the number of songs making up the dataset differ for each of the two models, we split the distribution into two: chart songs (the most popular 1%); non-chart songs (the remaining). As shown on Table 3, on both models the charts represent only about one third of the total number of streams.

Table 3 – How important are the charts in music consumption

#% of total music consumption	Download	Streaming
Top 200	33.8	22.6
Chart songs (Top 1%)	37.4	30.7
Non-chart songs	62.6	69.3

Our analysis shows that consumption as a function of the country-of-origin changes significantly from the top to the bottom of the distribution. In the charts, at the top of the distribution, Table 4 confirms the strong home bias already highlighted by the literature. In both download and streaming models, French songs account for more than half of consumption, while the market share of American songs barely exceeds 15%. However, when moving down the distribution (outside the charts), Table 4 illustrates that in both models, the market share of French content decreases significantly, mainly to the benefit of US content. The change in the market share of UK and Other countries is smaller.

Table 4 - Market share by country of origin.

		France	US	UK	Other	Total
Download	**Chart songs**	50.2	11.8	21.7	16.3	100
	Non-chart songs	41.3	24.5	15.0	19.2	100
Streaming	**Chart songs**	55.0	15.5	10.0	19.5	100
	Non-chart songs	42.3	27.0	11.4	19.3	100

In addition to the significant change in the market share of French and US products, looking beyond the top of the distribution highlights another important difference: the growing importance of the medium and back catalogues. Hence, Table 5 indicates that songs that were released at least ten years ago represent around 30% of consumption outside the charts, compared to 5% in the top 1%. This pattern is observed for both download and streaming models.

Table 5 - Market share by vintage

%		Frontline (< 2 years)	Medium (2–10 years)	Back (>10 years)	Total
Download	**Chart songs**	85.3	9.5	5.2	100
	Non-chart songs	36.9	33.1	30 .0	100
Streaming	**Chart songs**	80.9	13.9	5.2	100
	Non-chart songs	36.9	36.9	26.2	100

Another pattern common to both models is a strong correlation between the vintage of the catalogue and the country of origin. Table 6 highlights, for both models, the geographic composition of the back catalogue (more than ten years) and the frontline catalogue (less than two years). Here, the share of French content is significantly lower (by almost 20 percentage points) in the back catalogue compared to the frontline catalogue. Conversely, the US market share is at least twice as large in the former than in the latter.

Table 6 – Market share by country of origin for different vintages (more than ten years vs. less than two years)

%	Download		Streaming	
	Frontline	**Back**	**Frontline**	**Back**
France	52.1	30.8	56.1	27.6
US	15.1	29.7	16.9	38.7
UK	15.2	20.5	7.5	15.8
Other	17.6	19.0	19.5	17.9
Total	100.0	100.0	100.0	100.0

While both models illustrate the growing importance of the back catalogue when looking beyond the charts, together with the specific role of French and US content in back and more recent catalogues, there is an intriguing difference between download and streaming models that deserves further attention. In the download model, the French content market share remains (slightly) larger than the US market share in the back catalogue. This is not the case for the streaming model: here, US songs account for 38.7% of the market share of the back catalogue, compared to 27.6% for French songs. It seems that while the streaming model does not favor the back catalogue *per se*, it does favor the US back catalogue. Altogether these results suggest that music consumption on streaming platforms favors American "old" songs to the detriment of local music. This effect is economically significant as medium and back catalogues represent a vast majority of consumption in streaming platforms.

The higher market share of American old songs could be explained by differences in the characteristics and the number of US songs consumed in both type of platforms. We test for that in the next subsection (4.2). Of course, the difference in the composition of the populations of streaming subscribers and download purchasers (age, music tastes, etc.) could also contribute to explain these results. Although we are not able to properly control for these differences, we deal with this limitation in a robustness sub-section (4.4).

4.2. Econometric results

The descriptive analysis indicated that consumption on streaming platforms seems to favor US back catalogue songs. However, this result could be explained by differences in the characteristics and the number of US songs consumed. To control for this, and clarify the effect of streaming, we restricted our sample and used a subsample made up of songs that were simultaneously available on both channels, and compared their market share for a given week. For each of the 36 weeks, we considered 7,200 songs available via the two models. As we were able to document the country of origin of 94.7% of these songs, the final sample contained an average of 6,814 songs per week.

We then ran regressions with the song as a fixed effect. Our dependent variable was the weekly market share of each song in each of the two models (download, streaming) and for two parts of the distribution: in the charts (the top 1%) and outside the charts. We also distinguished songs by their vintage (Frontline, Medium or Back). The main dependent variables included a dummy variable for the geographical origin of the song (US, UK, Other, with France as the reference), and a dummy variable for streaming (with download as the reference). We also introduced interaction variables for models and geographical origin. Control variables included the artist's past success (volume of download sales between 2006 and 2014), the provider (the three major independent distributors, Universal Music Group, Sony Music Group and Warner Music Group, and Believe Digital, the main digital aggregator) and the musical genre (Pop, Rock, Urban, ElectroDance, Popular or Other). Table A1 (see the Appendix) presents the main descriptive statistics.

Table 7 presents the results of the analysis. Column 1 shows results for chart songs, and Columns 2 to 4 show results for songs that did not chart. For the latter, we also distinguish songs according to their vintage: Column 2 shows results for the frontline catalogue (two years old or less), Column 3 for the medium catalogue (from two to ten years old), and Column 4 the back catalogue (more than ten years). It should be noted that we do not make the same distinction for the Top 1%, as only a very small number of songs from the back catalogue enter the Top 1%; furthermore, we were unable to implement a song fixed effect due to the small number of observations.

Table 7 shows that, for chart songs, market share is lower on a streaming platform than a download platform (Column 1). The reverse is true for non-chart songs: here, the market share of a song is higher when consumed on a streaming platform compared to a download platform, except for the back catalogue but with a low magnitude (Columns 2, 3, and 4).

Regarding geographical origin, we also observe significant differences between chart and non-chart songs for US songs. For chart songs (Column 1), the coefficient of the variable *USA × Streaming* is not significant. This indicates US content does not benefit from the rise of streaming in the top 1%. Beyond the top 1%, however, things are very different (Columns 2, 3, and 4). The coefficients of the variable *USA × Streaming* are positive and significantly different from zero in columns 3 and 4. This suggests that the marginal gain for US content in the streaming model is significantly higher than for French repertories outside the chart, except for most recent songs (the frontline catalogue, column 2). It should be noted that the better relative performance of US content compared to French content in the streaming context compared to the download model seems to be specific to US content. Songs from the UK and other countries are consistently less consumed (compared to French ones) in the streaming model compared to the download model.

These results confirm that, as compared to download, music consumption on streaming platforms favors of American repertoire to the detriment of the local one. The effect holds for medium and back catalogues outside the top 1% of the distribution.

Table 7 – Determinants of market share for chart songs (top 1%) and non-chart songs

Dependent variable: Market Share	**Chart songs**	**Non-chart songs**		
	(1) All	(2) Frontline	(3) Medium	(4) Back
France	Ref.	Ref.	Ref.	Ref.
USA	−0.0222	-0.0127***	-0.00158***	-0.00345***
	(0.0229)	(0.00107)	(0.000322)	(0.000835)
UK	0.0962***	0.00117	0.0134***	0.000653
	(0.0221)	(0.00155)	(0.000389)	(0.000874)
Other	0.0615***	-0.00597***	0.00632***	0.00107
	(0.0224)	(0.00103)	(0.000344)	(0.000780)
Download	Ref.	Ref.	Ref.	Ref.
Streaming	−0.197***	0.00585***	0.00275***	-0.000215***
	(0.0168)	(0.000178)	(0.0000572)	(0.0000625)
France × Streaming	Ref.	Ref.	Ref.	Ref.
USA × Streaming	0.00249	-0.00240***	0.000341***	0.00232***
	(0.0283)	(0.000294)	(0.0000926)	(0.0000869)
UK × Streaming	−0.0143	-0.0136***	-0.00292***	-0.00284***
	(0.0299)	(0.000478)	(0.000135)	(0.000118)
Other × Streaming	−0.0495*	-0.00421***	-0.00109***	-0.000584***
	(0.0278)	(0.000282)	(0.000101)	(0.000100)
Control for past success	Yes	Yes	Yes	Yes
Control for musical genre	Yes	Yes	Yes	Yes
Control for providers	Yes	Yes	Yes	Yes
Song fixed effect	No	Yes	Yes	Yes
Constant	Yes	Yes	Yes	Yes
N	4,652	121,379	216,457	148,116

Robust standard errors in parentheses; * p<0.1, ** p<0.05, *** p<0.01

4.3. Magnitude of the effect

Table 7 indicates the statistical significance of the difference between US and French content regarding the rise of streaming at the expense of the download model. The magnitude of these differences could be measured by comparing the average gain in market share of US songs of a specific vintage from the switch from download to streaming (compared to a French song) to the average market share in the streaming model of a US song from the same vintage. Table 8 reports these magnitudes.

Table 8 – Magnitude of the effect of streaming consumption vs. download consumption for US content compared to French content

	Chart songs	Non-chart songs		
		Frontline catalogue	**Medium catalogue**	**Back catalogue**
Marginal gain (%)	ns	-0.8	+3.5	+26.0

Note: the marginal gain for a given vintage is calculated as the following ratio: [gain in absolute market share of a switch from download to streaming for a given vintage (the coefficient of *USA* x *Streaming* in Table 7) / Average market share via streaming of a US song from the same vintage]; ns: not significant

Table 8 confirms a slightly negative difference between US and French content for the frontline catalogue outside the charts and no significant difference in the top 1% where frontline songs are largely dominant. One possible explanation is that new songs have a higher probability of being promoted through radio or TV broadcasts than older ones, which could impact simultaneously download and streaming consumption. Table 8 also confirms the observation of a relative gain for US content compared to domestic content for the older songs, linked to the growth of streaming of non-chart songs: outside the top 1%, the market share of an American song from the back (medium) catalogue is 26% (3.5%) higher when listened to a streaming platform than on a download platform. This effect could be explained by a higher match between consumer preferences and this type of songs (or relatively higher quality of US medium and back catalogues), which generate revenue each time they are listened to in the streaming model, but only after a sale in the download model. Another possible explanation could be that curation by platforms is biased, intentionally or not, toward the US back catalogue. We test these two possible explanations in the next section (section 5).

4.4. Robustness

The size and completeness of our dataset (weekly consumption data for the same sample of around 7,200 songs on the two platforms) support the robustness of our results. It is possible, however, that there is a composition effect. We indirectly control for such an effect at the supply side as catalogues of the major platforms were huge and similar (around 30 million of songs on Spotify or iTunes), but we cannot control at the demand side: Consumers on download and streaming platforms could be different (according to age, music tastes, etc.). Specifically, a conjecture that could qualify our results is that the download market may be made up of a group of consumers with very specific preferences; people who have a preference for US music, especially medium and back catalogues, may have switched from download platforms to streaming platforms. Wlömert and Papies (2016) indeed show that consumers who adopt a paid streaming service are likely to quit pay-download platforms.[16] However, if this hypothesis holds, we should observe that in 2017, downloaders demonstrated a greater preference for French content over US content than previously. Hence, we should observe a decrease in the market share of the US back catalogue in the download market over the latter period and, conversely, an increase in the market share of French content. To explore this issue, we collected GfK data for the 20,000 most popular songs for each of the 52 weeks of 2007 (three years after the introduction of iTunes in France) and 2012 (the peak year for download sales in France). Using Musicbrainz, we obtained the artist's nationality for 93.4% of songs in 2007, and 93.3% in 2012. We then computed the market share of the back catalogue for the three years, along with the respective market share of French and US content in back catalogue sales.

Table 9 – Size and composition of back catalogue sales in the download model.

%	**2007**	**2012**	**2017**
Share of the back catalogue (≥10 years) in total download sales	12.5	19.2	20.7
Share of French content in the back catalogue	43.6	45.4	30.8
Share of US content in the back catalogue	23.4	28.9	29.7

Note: for each year, calculations are based on the 20,000 best-selling songs for each week.

Table 9 presents the results. It indicates that the share of the back catalogue increased over the years in the download model, and that the share of French content was higher than the share of US content in the back catalogue. Moreover, over the period 2007–2017, the share of French content did not increase, nor did the US market share decrease. These results do not support the hypothesis of a composition effect: download users' preferences in 2017 do not explain the relatively higher consumption of US content on streaming platforms compared to download ones.

To sum up, two results deserve to be highlighted. Firstly, we document a relative benefit for US content compared to French content for non-chart music in the streaming model. This effect is significantly larger for medium and back catalogues compared to frontline songs. Secondly, this advantage seems to be specific to US content. A similar relative advantage compared to local content is not observed for content from the UK and other countries.

[16] "You're seeing the most valuable music customers, who were happily buying bucketloads of downloads from iTunes, now finding they can get even more value by spending £9.99 a month".
https://www.theguardian.com/technology/2014/aug/29/music-streams-downloads-mark-mulligan

5. Curation push vs. preference pull

5.1 Econometric results

As the hypothesis of a composition effect does not seem to explain our results, we were left with two possible explanations, depending on the characteristics of the streaming model. Firstly, the effect could reflect the business model of streaming platforms and translate a relative consumer preference could favor US medium and back catalogues. Secondly, the effect could be driven by streaming platforms, which have developed recommendation systems and curated playlists. Aguiar and Waldfogel (2021) show that the major platform-operated playlists have large and significant causal impacts on streaming. However, the latter authors also stress that US-origin songs benefit most from global playlists, while other lists (especially Spotify's New Music Friday) contain more domestic music. In our approach, the issue with streams generated by algorithmic recommendations or played within platform playlists is that they have not been chosen by the user and could be listened passively and uncarefully. Yet, when it comes to calculate each song's market share to pay royalties to right holders, those streams are equivalent to streams actively and carefully chosen by the user. This is the source of the *curation push* bias.

To disentangle these two possible explanations – a consumer preference for US back catalogue content (*preference pull*), or a US bias in music curated by platforms (*curation push*) – we use our second dataset and ran an econometric analysis. Table A2 in Appendix indicates that 76.5% of streams are on-demand music (or *preference pull*), from an autonomous search by the users (48.1% of all streams) or from a playlist generated autonomously by the users (28.3%). In contrast, 23.5% of streams have been *push* in an algorithmic recommendation (19.1%) or a playlist generated by the streaming platform or the recorded company (4.4%).

We built our dependent variable *DiffPushPull* as the difference, for a given song in a given month, between the percentage of curated streams (*push*), either following a recommendation or included on a curated playlist, and the percentage of streams that followed an autonomous search by the user or that came from a user-generated playlist (*pull*). We considered that curation push tended to increase the value of *DiffPushPull* and, conversely, preference pull tended to decrease its value. Relying on percentage rather than on absolute values allows us to control for a potential endogeneity issue since a song could be more push because it better fits users' preferences. When we regress the dependent variable *DiffPushPull*, there is a priori no reason to believe that the independent variables of geographical origins are correlated with the error term.

Given our research question, we focused on the role of the artist's nationality and the vintage of the catalogue, to assess whether US songs did indeed benefit more from curation than domestic songs, especially for the back catalogue. As previously, vintage was modeled as three dummies:

Frontline (released less than two years ago), *Medium* (released between two and ten years ago), and *Back* (released more than ten years ago).[17] Table A2 in appendix shows that around 10% of streams in the sample are from the frontline catalogue, whereas over 43% are ten years old or more. Geographic origin was the second variable that was expected to impact the dependent variable. In our second sample, US streams accounted for 33.7% of observations, UK songs for 12.7%, and domestic (French) songs for 36.8%. Finally, we added the following control variables: provider (major, independent, digital aggregator), genre (Pop, Rock, Urban, ElectroDance, Popular and Other), song popularity, and artist popularity.[18]

Column 1 in Table 10 presents results for the overall sample without interactions. Column 2 present results for chart songs (as we noted in section 4, the competitive advantage of streaming for US songs was not observed at the top of the distribution) and Column 3 for non-chart songs.

Results presented in Column 1 suggest that, all other things being equal, and controlling for the provider, the musical genre, the artist's popularity and the song's popularity, a song from either the medium or the back catalogue is, on average, slightly more pushed than a frontline song. For instance, all other things being equal, the difference between the share of streams that are pushed and the share that are on-demand for a song from the back catalogue is 2.1 percentage points higher than for a frontline song. Likewise, a UK song is more pushed than any other nationality: *DiffPushPull* increases by 11.2 percentage points for a UK song compared to a local one. US songs are also more pushed than French ones (+9.8 percentage points).

Column 3 disentangled for non-chart songs the curation push effect and the preference pull effect for the different vintages and for each geographical origin. The positive coefficients US, UK and Other indicate that foreign songs in the Frontline catalogue are more pushed than chosen autonomously as compared to French ones. As compared to the Frontline catalogue, this advantage is always less important for foreign songs from the medium catalogue as compared to French songs of the same vintage (all interaction terms *Country* × *Medium* are negative). Turning to the back catalogue, US songs turn out to be the only ones to be more pushed in the back catalogue than in the frontline one: the coefficient of the variable *USA* × *Back* is positive and significant. For instance, while a back catalogue French song is slightly more pushed than a frontline French song (*DiffPushPull* increases by 4.1 percentage points), the difference is larger (+5.8 percentage points) for a US song from the back catalogue as compared to a US song from the

[17] We alternatively also use a discrete variable (Vintage) that covers all the possible categories: for Frontline songs, Vintage = 0; for 2 to 10 years old songs, Vintage = 1; for more than ten years old songs, Vintage = 2. The results remain unchanged.

[18] The variable *song popularity* represents the volume of streams for a given song i over all six months, and *artist popularity* specifies how the artist who performs the song i is ranked in our sample (in the Top 100, from rank 101 to 1,000, from rank 1,001 to 10,000, beyond rank 10,000).

frontline catalogue. Likewise, we note that the difference between *push* and *pull* for back/medium and frontline catalogues is almost always lower for songs from the UK and other countries compared to songs from the US. These results support our claim that US songs from the back catalogue, which account for 26.4% of the total volume of streams in our sample, are especially favored by curation. Moreover, we ran separate regressions for the two categories of curated streams (algorithmic recommendations and human-created playlists). The coefficient US × Back in Table A3 in Appendix shows that as compared to the US frontline catalogue, the US back catalogue benefits only from algorithmic recommendations and not from platform playlists.

Focusing on the best-selling songs in our sample (the monthly top 1%), Column 2 indicates that medium or back US catalogues are less pushed than French songs from the same vintage, and also less pushed than the frontline US catalogue. This result is consistent with our claim that the benefits of platform curation for US content are focused on the back catalogue, which is almost absent from the list of the most-streamed songs (see section 4). Altogether these results suggest that streaming platforms bias music consumption in favor of American back catalogue songs, especially through their algorithmic recommendations.

5.2 Robustness

Previous results indicate US songs from the back catalogue benefit from algorithmic recommendation. One may argue that this result comes from a composition effect: the streaming users that rely the most on curation may also be those who have a relative preference for this type of songs, i.e. who are more prone to listen to the US back catalogue. If this effect is true, we should observe a positive correlation between the share of subscribers who use recommendations and the share of autonomous consumption of US back catalogue (without the use of recommendations or playlists curated by platforms or labels). To test this potential effect, we calculate this correlation on the second dataset at the subscriber level. The correlation coefficient is negative (-0.197, $p<0.0000$): the more users use the recommendations, the less US back catalogue songs they choose autonomously. This suggest that American bias is not explained by a composition effect of our sample.

This result is in line with another study on several hundred of French subscribers to music streaming platforms (Hadopi, 2020). Results show that the youngest subscribers – presumably less interested in the US back catalogue – are over-represented among the users that rely on recommendations from the platform and playlists to discover new music. Moreover, fans of

French music are neither over-represented nor under-represented among the users of recommendations and playlists.[19]

Table 10 – The determinants of curated vs. on-demand streams.

Dep. variable: DiffPushPull = share of push – share of pull	(1)	(2)	(3)
	All	Chart songs	Non-chart songs
Frontline	Ref.	Ref.	Ref.
Medium	0.0188***	0.0516***	0.0568***
	(0.00245)	(0.00747)	(0.00336)
Back	0.0210***	0.196***	0.0412***
	(0.00256)	(0.0116)	(0.00356)
France	Ref.	Ref.	Ref.
USA	0.0976***	0.215***	0.110***
	(0.00188)	(0.0170)	(0.00513)
UK	0.112***	0.461***	0.213***
	(0.00259)	(0.0362)	(0.00888)
Other	0.0683***	0.214***	0.108***
	(0.00225)	(0.0172)	(0.00598)
USA × Medium		-0.172***	-0.0416***
		(0.0198)	(0.00567)
UK × Medium		-0.305***	-0.113***
		(0.0392)	(0.00954)
Other × Medium		-0.112***	-0.0372***
		(0.0202)	(0.00668)
USA × Back		-0.0820***	0.0164***
		(0.0215)	(0.00578)
UK × Back		-0.356***	-0.105***
		(0.0398)	(0.00958)
Other × Back		-0.193***	-0.0556***
		(0.0244)	(0.00694)
Constant	-0.767***	-0.778***	-0.814***
	(0.00606)	(0.0181)	(0.00679)
Control for:			
Provider	Yes	Yes	Yes
Musical genre	Yes	Yes	Yes
Artist's popularity	Yes	Yes	Yes
Song's popularity	Yes	Yes	Yes
Month dummy	Yes	Yes	Yes
N	944,297	18,167	926,130
R^2	0.026	0.312	0.026

Robust standard errors are shown in parentheses; * p<0.1, ** p<0.05, *** p<0.01.

6. Discussion and conclusion

In a music market that is dominated by a few international streaming platforms, concerns have been raised about a potential bias toward large-country repertories, notably the dominance of the US repertoire. In this paper, we studied if, and why, streaming consumption favors US content to the detriment of the local repertoire in a market that is characterized by significant domestic production.

Unlike most earlier works, which has focused on the music charts, we first compared weekly consumption of several thousands of songs in two distribution models in France: premium streaming, which is the most lucrative model for the music industry, and download. Focusing on geographic origin, our results documented that the average market share of US songs is greater in the streaming model than in the download model, to the detriment of French songs. This better performance of the US content is strong for non-chart music and for the back catalogue (i.e., songs released more than ten years ago). A similar relative advantage is not observed for content from the UK and other countries compared to local (French) songs.

We document that non-chart music and the back catalogue are economically significant as they cumulate 69% and 20% of the total consumption in France, respectively. As a consequence, the higher performance of US content and the corollary lower performance of local repertoire are far from being negligeable and are particularly harmful in a market dominated by few platforms.

We then tested two possible explanations for this American hegemon: a "curation push" effect or a "preference pull" effect. The former refers to the role of mechanisms implemented by streaming platforms to guide consumers who have access to a huge catalogue of tens of millions of songs: their playlists and recommendations could be biased toward the US repertoire. The second explanation finds its roots in the way revenue is generated in the streaming model. Unlike the ownership model (download or CD) where each sale only generates revenue once, in the streaming model, revenue is generated each time a song is listened to (section 2.2). This could favor US (back catalogue) songs if consumers have a relative preference for this type of content and, on average, listen more times to songs in this catalogue than songs in other catalogues.

Drawing upon consumption data for a sample of several thousands of subscribers to a main audio streaming platform, our results support the idea that streaming platforms have introduced a bias toward American songs from the back catalogue and to the detriment of local one: compared to French songs, the US back catalogue is pushed more than chosen autonomously. This result holds only for algorithmic recommendations and not for the playlists generated by the streaming

[19] While 82.1% of subscribers rely on curation to discover new music, this figure increases to 86.6% for users between 15 and 34 years old and equals 82.2% for fans of French music (Hadopi, 2020).

platform and the recorded music labels. It differs from Aguiar and Waldfogel (2021) who show that main playlists tend to favor US songs on Spotify (except in the New Music Friday playlists). The difference could be explained by the fact they studied the composition of playlists whereas we examined the consumption of playlists.

These results contribute to the literature on the international trade in cultural goods. In particular, they show how music consumption via the streaming model differs from consumption via the download model. We suggest that this finding is linked to the characteristics of the streaming model, and is not a consequence of digitization. We observe a bias toward the US repertoire at the expense of the local, French repertoire. This trend is not found at the top end of the consumption distribution, but rather in the remaining of the distribution. From a methodological point of view, this result highlights the need to study the whole consumption distribution in the streaming age, rather than only the charts, as has been the norm in most previous studies. From an economic perspective, the observed effect is non-negligible and should be taken into consideration by the industry. Our finding that the bias is especially strong in the back catalogue is also consistent with the recent appetite of investors for the back catalogue of US stars.

Our results also contribute to the growing literature on the impact of playlists and the (algorithmic) bias that digital platforms may introduce (Aguiar and Waldfogel, 2021; Aguiar et al., 2021). In addition to gender discrimination (Lambrecht and Tucker, 2019) or ideological polarization (Levy, 2021), platform recommendations can also influence music consumption (and revenues) and tend to favor large-country content to the detriment of local content.

Overall, our results raise concerns about cultural diversity and how to protect it. Since the 1990s, cultural goods have been exempted from free trade treaties, and many countries, including France, have adopted domestic content protection measures. In the music market, this has mainly taken the form of quotas on radio stations. Previously, local content has been promoted by radio stations, enabling consumers to discover new domestic music. Today, streaming platforms play a significant role in the promotion and discovery of recorded music. While we still cannot be sure whether domestic quotas are efficient (Ferreira and Walfogel, 2013; Richardson and Wilkie, 2015), it should be noted that they do not apply to streaming platforms.

Our work is not without limitations. First, we only measured the impact of algorithmic recommendations and playlists on songs that were listened to. Our dataset did not make it possible to examine all of the recommendations made by a platform, and all of the songs included on a playlist, according to their geographical origin. This is challenging from a methodological point of view: how can we collect data on algorithm recommendations that are mostly personalized, and isolate the impact on geographic origin for consumers with heterogeneous preferences?

Another limitation is that our results are based on a single market. Testing the bias toward US content in countries other than France, such as Canada (a neighboring North American country), the UK (another Anglophone country with a dynamic music market) or Germany (another European non-Anglophone country), are other interesting opportunities for future research. It should however be noted that future research could conduct to different results because curation tools permanently evolved.

This paper also ignores the consequences of the push effect. How do consumers react to the promotion of US content in algorithmic recommendations? Do they listen repeatedly to these songs? Do they modify their preferences and future choices in favor of the US repertoire? We leave these questions to future studies.

Finally, our results raise questions about the impact of the observed bias on the supply side of the local music market. In the case of newspapers, George and Waldfogel (2006) found that the availability of a national newspaper (*The New York Times*) on local markets modified the position of local newspapers. How does a local music industry react to a bias toward the US (back) catalogue? This is clearly another interesting subject for future research.

7. References

Aguiar, L., Gomez-Herrera, E., and J. Waldfogel (2018). Does digitization threaten local culture? Music in the transition from iTunes to Spotify, mimeo.

Aguiar, L., and J. Waldfogel (2018). Netflix: global hegemon or facilitator of frictionless digital trade? *Journal of Cultural Economics*, 42(3), 419-445.

Aguiar, L., and J. Waldfogel (2021). Platforms, power, and promotion: Evidence from Spotify playlists. *The Journal of Industrial Economics*, 69(3), 653-691.

Aguiar, L., Waldfogel, J., and S. Waldfogel (2021). Playlisting favorites: Measuring platform bias in the music industry. *International Journal of Industrial Organization*, 78, 102765.

Anderson C. (2006), *The long tail*, London: Random House.

Bakos J.Y (1997). Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science*, 43(12), 1609–1755.

Belleflamme, P., and M. Peitz (2018). Inside the engine room of digital platforms: Reviews, ratings and recommendations. In J.-J. Ganuza & G. Llobet (Eds.), *Economic Analysis of the Digital Revolution* (Chapter 3, pp. 75–114). Funcas.

Bourreau, M., and G. Gaudin (2022). Streaming platform and strategic recommendation bias. *Journal of Economics & Management Strategy*, 31(1), 25-47.

Bourreau, M., and A. Perrot (2020). Digital platforms: Regulate before it's too late. *Notes du conseil d'analyse économique*, 60, 1-12.

Brynjolfsson, E., Hu,Y., and D. Simester (2011). Goodbye Pareto Principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, 57(8), 1373– 1386.

Datta, A., Tschantz, M.C., and A. Datta (2015). Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1), 92-112.

Datta, H., Knox G., and B.J. Bronnenberg (2018). Changing Their Tune: How Consumers' Adoption of Online Streaming Affects Music Consumption and Discovery. *Marketing Science*, 37(1), 5–21.

Edelman, B. (2011). Bias in search results: Diagnosis and response. *Indian Journal of Law & Tech.*, 7, 16.

Eriksson, M. (2020). The editorial playlist as container technology: on Spotify and the logistical role of digital music packages. *Journal of Cultural Economy*, 13:4, 415-427.

Ferreira, F., and J. Waldfogel (2013). Pop Internationalism: Has A Half Century of World Music Trade Displaced Local Culture? *Economic Journal*, 123(569), 634–664.

Fleder, D., and K. Hosanagar (2009). Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science*, 55(5), 697-712.

George, L.M., and C. Peukert (2016). Youtube Decade: Cultural Convergence in Recorded Music. *NET Institute Working Paper No. 14-11.* http://dx.doi.org/10.2139/ssrn.2506357

George, L.M., and J. Waldfogel (2006). The New York Times and the Market for Local Newspapers. *American Economic Review*, 96(1): 435-447.

Hadopi (2020). *Les pratiques d'écoute de musique en ligne – phase quantitative*, septembre.

Hagiu, A. and B., Jullien (2011). Why do intermediaries divert search? *The RAND Journal of Economics*, 42: 337-362.

IFPI (2022). Global music report 2022. Available at https://www.ifpi.org/resources/

Lambrecht, A., and C. Tucker (2019) Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science,* 65(7): 2966-2981.

Lenard, T.M. (2019). Introduction to the RIO Special Issue on Antitrust and the Platform Economy. *Review of Industrial Organization*, 54: 617-626.

Levy, R. (2021). Social Media, News Consumption, and Polarization: Evidence from a Field Experiment. *American Economic Review,* 111: 831–870.

Mariuzzo, F., and P.L. Ormosi (2022). Independent v major record labels: Do they have the same streaming power (law)? *Available at SSRN 3729966*.

Obermeyer, Z., Powers, B., Vogeli, C., and S. Mullainathan (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464): 447-453.

Parker, G.G., and M.W. Van Alstyne (2005). Two-sided network effects: A theory of information product design. *Management science*, 51(10): 1494-1504.

Paul, M., Hennig-Thurau, T., Gremler, D. D., Gwinner, K. P., and C. Wiertz (2009). Toward a theory of repeat purchase drivers for consumer services. *Journal of the Academy of Marketing Science*, *37*(2): 215-237.

Richardson, M., and S. Wilkie (2015). Faddists, enthusiasts and Canadian divas: broadcasting quotas and the supply response. *Review of International Economics*, 23(2): 404–424

Salganik, M.J., Dodds P.S., and D.J. Watts (2006). Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, 311: 854–856.

Waldfogel, J. (2012). Copyright protection, technological change, and the quality of new products: evidence from recorded music since Napster. *Journal of Law and Economics*, 55(4): 715–740.

Way, S.F., Garcia-Gathright, J., and H. Cramer (2020). Local Trends in Global Music Streaming. *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media* (ICWSM 2020).

Zeithaml, V. A., Berry, L. L., and A. Parasuraman (1996). The behavioral consequences of service quality. *Journal of Marketing*, 60(2), 31-46.

APPENDIX

Table A1 – Descriptive statistics for dataset 1.

Variable	Obs.	Mean	Std. Dev.	Min	Max
MarketShare (DV)	490,604	.0146758	.0556926	0	10.04546
Geographic Origin					
France	490,604	.3953372	.4889235	0	1
USA	490,604	.2832835	.4505933	0	1
UK	490,604	.0980873	.2974329	0	1
OtherCountries	490,604	.2232921	.416453	0	1
Musical genre					
Pop	490,604	.1907608	.3929013	0	1
Rock	490,604	.0899667	.2861343	0	1
Urban	490,604	.2612046	.4392916	0	1
French variety	490,604	.0946344	.2927095	0	1
ElectroDance	490,604	.076897	.2664283	0	1
OtherGenre	490,604	.2865366	.4521436	0	1
Provider					
Believe	490,604	.10781	.3101406	0	1
Universal	490,604	.3090395	.4620979	0	1
Sony	490,604	.246382	.430904	0	1
Warner	490,604	.2228844	.4161818	0	1
OtherProvider	490,604	.1138841	.3176708	0	1
Vintage					
Frontline (< 2 years)	490,604	.2553505	.4360585	0	1
Medium (2–5 years)	490,604	.3092148	.4621703	0	1
Back (≥ 5 years)	490,604	.4354347	.4958143	0	1
PastSuccess	490,604	580,567.3	969,715.2	0	6,764,008

Note: The difference in the number of observations between Tables 1 and A1 pertains to the fact that in Table A1 only songs that have been "sold" in both streaming and download formats in the same month are considered. Furthermore, only songs with a known nationality are included.

Table A2 – Descriptive statistics for dataset 2.

Variable	Obs.	Mean	Std. Dev.	Min	Max
Context					
Music on-demand (pull)	944,297	.7648	.3583	0	1
Autonomous search	944,297	.4814	.4281	0	1
User-Generated Playlists	944,297	.2834	.3732	0	1
Curated streams (push)	944,297	.2352	.3583	0	1
Algorithmic recommendations	944,297	.1909	.3318	0	1
Playlists	944,297	.0443	.1672	0	1
Geographic Origin					
France	944,297	.3684	.4824	0	1
USA	944,297	.3367	.4726	0	1
UK	944,297	.1265	.3324	0	1
OtherCountries	944,297	.1684	.3743	0	1
Musical genre					
Classical	944,297	.0154951	.1235113	0	1
Electro	944,297	.0613726	.2400127	0	1
Urban	944,297	.1177013	.3222543	0	1
Pop	944,297	.2902074	.4538582	0	1
Rock	944,297	.1760389	.3808535	0	1
Jazz	944,297	.0217463	.1458542	0	1
R&B	944,297	.0412963	.1989748	0	1
Other	944,297	.0709904	.2568089	0	1
Unknown	944,297	.1945987	.3958917	0	1
Provider					
Major	944,297	.7098519	.4538308	0	1
Independent	944,297	.1679546	.3738261	0	1
Digital aggregator	944,297	.1221935	.3275094	0	1
Vintage					
Frontline (< 2 years)	944,297	.1037	.3049281	0	1
Medium (2–10 years)	944,297	.4682	.4989908	0	1
Back (> 10 years)	944,297	.4280	.4947909	0	1

Table A3 – The determinants of curated vs. on-demand streams disentangling the playlists and recommendations impact

	(1)	(2)
Dep. variable:	share of reco – share of pull	share of playlist – share of pull
Frontline	Ref.	Ref.
Medium	0.0704***	0.0149***
	(0.00395)	(0.00264)
Back	0.0526***	0.00928***
	(0.00407)	(0.00271)
France	Ref.	Ref.
US	0.0734***	0.0912***
	(0.00538)	(0.00359)
UK	0.144***	0.176***
	(0.00803)	(0.00536)
Other	0.0705***	0.0915***
	(0.00607)	(0.00405)
US × Medium	-0.0170***	-0.0453***
	(0.00586)	(0.00391)
UK × Medium	-0.0633***	-0.106***
	(0.00864)	(0.00577)
Other × Medium	-0.0177***	-0.0381***
	(0.00667)	(0.00445)
US × Back	0.0355***	-0.0109***
	(0.00591)	(0.00394)
UK × Back	-0.0459***	-0.112***
	(0.00866)	(0.00578)
Other × Back	-0.0283***	-0.0551***
	(0.00688)	(0.00459)
Constant	-0.800***	-0.883***
	(0.00837)	(0.00559)
Control for:		
Provider	Yes	Yes
Musical genre	Yes	Yes
Artist's popularity	Yes	Yes
Song's popularity	Yes	Yes
Month dummy	Yes	Yes
N	926,130	926,130
R^2	0.029	0.024