An Empirical Analysis of Music Streaming Revenues and Their Distribution

Daniel Antal, CFA

2021-09-23

Contents

List of Figures

Executive Summary

The research questions asked in this report are related to the Music Creator Earnings' Project (MCE), exploring issues concerning equitable remuneration and earnings distributions. We were tasked with providing a longitudinal analysis of earnings development and relating our findings to equitable remuneration. The starting point of our work was centred around a very broadly defined problem: how much money music creators (rightsholders) earn from streaming, how these earnings are distributed, and how the earnings and their distribution have developed during the last decade.

The highly globalized music industry generates two important international reports, as well as several national reports, but these are not suitable for the analysis of the typical or average rightsholder, nor for small labels and publishers who do not represent a large and internationally diversified portfolio of music works or recordings. Copyright and neighboring right revenues are collected in national jurisdictions. Because British artists are almost never constrained by their use of language, and the UK Music Industry is highly competitive in the global music markets, even relatively less known rightsholders earn revenues from dozens of national markets. The lack of market information on music sales volumes, prices for each jurisdiction, and the unaccounted for national, domestic, and foreign revenues makes the analysis of the rightholder's earnings, or the economics of a certain distribution channel like music streaming or media platforms, impossible.

While total earnings are reported by international and national organizations, they hide five important economic variables: changes in sales volumes, changes in prices, market share on various national jurisdictions (which have their own volume and price movements), the exchange rates applied, and the share of the repertoire exploited. Even worse, the global music industry has no comprehensive database of rightsholders, music works, and recordings. Many rights are represented by heirs or passive investors. And because of the enormous number of works and recordings, in any given royalty payment period, most works/recordings are not used and not compensated. The lack of a known population and distribution makes usual indicators as average or median earnings arithmetically impossible to compute, and the large number of disused works and recordings makes even the estimation of the typical (median) value useless for economic analysis. The estimation of the arithmetic mean is equally problematic, because it is distorted by the earnings of very few global stars. To understand the streaming economy from the perspective of a typical or average rightsholder, or from the perspective of a small independent label or music publisher, requires very challenging sampling techniques either in surveying or in empirically observing and aggregating data from royalty accounts. National and international music organizations are not equipped with the data processing and statistical capacity to do so.

This is the problem that the Digital Music Observatory, a working demo of the European Music Observatory is solving. It grew out of the Central & Eastern European Music Industry Databases (CEEMID) initiative in 2014, in which rightsolders from three countries attempted to solve this problem. By 2019, CEEMID had collected information on 20 European markets, including the United Kingdom, and processed data on far more markets. The state51 music group, through its distribution arm, has been supporting the creation of the largest ever European market report, the Central European Market Report, and supported the creation of the CEEMID-CI indexes, which, for the first time provided a stock-index type of view from an individual rightsholder's perspective on volume and price movements in the UK and in other countries. The state51 music group drew attention to the observatory approach and this work

in the Digital, Culture, Media and Sport Committee (DCMS) Select Committee of the British House of Commons. The MCE project first individually contacted the Digital Music Observatory (successor of CEEMID) and state51, and eventually with the permission of state51, the project commissioned this report, which re-uses the CEEMID-CI indexes. The MCE project also committed to share data in the Digital Music Observatory.

The first chapter is divided into three subchapters: 1.1 the theoretical problems imposed by copyright law and intellectual property valuation methods; 1.2 policy issues related to the copyright management problems that lead to large amounts of unpaid royalties and value transfer, as well as potential problems with AI distribution systems and royalty distribution rules; and 1.3 identifying and offering solutions for empirical problems in estimating the relevant earnings indicators for individual rightsholders.

The second chapter presents the CEEMID-CI indexes, and gives an explanation to the seemingly paradoxical situation of total market growth, and the feeling of individual rightsolders that their earnings remain flat or decreasing. We show in 2.1 that revenues that are flat or slightly growing are caused by growing volumes and falling prices (See 2.2; 2.3.)

The third chapter, using computer simulations, gives a fuller picture. It shows that in the UK and many other markets, in the period of 2016-2019, the growth of streaming use hardly offset the diminishing price (value) of the streams. In many markets, rightsholders, expressed in British pounds, experienced flat earnings, which was by a large extent due to the benevolent USD/GBP, EUR/GBP, and CHF/GBP exchange rates. It is likely that the majority of individual rightsholders experienced flat or diminishing revenues, and later releases brought less and less revenue per a thousand streams. A less benevolent exchange rate environment could have brought very large losses of revenue (See 3.4 for more details.)

Our computer simulations show that British rightsholders are experiencing a rather varied picture of music streaming. Because of the generally falling value of streams, short-lived hits are more impacted by value decline than perennial hits (See subchapters 3.1 and 3.2.) Rightsholders with a diversified international audience are partly shielded by these effects (See 3.3.) Younger stars and new genres generally experience a windfall from the pro-rata distribution system and older artists with more conventional genres such as rock or blues are experiencing headwinds. The publicly available market totals are hiding completely different market experiences.

The MCE project and its stakeholders did not provide us with any data. Our detailed data from other jurisdictions, however, shows that music streaming has been successfully competing with other sales/distribution formats of music partly because it was much cheaper, and generally devalued music-related rights. This is particularly the case with media platforms like YouTube. For the UK, which is not planning to adopt new EU regulations, closing the value gap is an important policy task. We also show on the example of other countries that current policy debate in the UK around the re-definition of equitable remuneration rigths, or on the pro-rata or user-centrique distribution schemes, are gaining excessive focus, and changes in these aspects of the streaming ecosystem will solve the problems of few, if any, rightsholders. We believe that the music industry must look beyond the current practices of streaming pricing and distribution, for example, to telecommunication services, and investigate the adoption of a more just system.

Our report highlights some important lessons. First, we show that in the era of global music sales platforms it is impossible to understand the economics of music streaming without international data harmonization and advanced surveying and sampling. Paradoxically, without careful adjustments for accruals, market shares in jurisdictions, and disaggregation of price and volume changes, the British industry cannot analyze its own economics because of its high level of integration to the global music economy. Furthermore, the replacement of former public performances, mechanical licensing, and private copying remunerations (which has been available for British rightsholders in their European markets for decades) with less valuable streaming licenses has left many rightsholders poorer. Making adjustments on the distribution system without modifying the definition of equitable remuneration rights or the pro-rata distribution scheme of streaming platforms opens up many conflicts while solving not enough fundamental problems. Therefore, we suggest participation in international data harmonization and

policy coordination to help regain the historical value of music.

Chapter 1

Theoretical, Policy and Empirical Measurement Challenges

We were tasked with providing a longitudinal analysis of earnings development as well as relating our findings around equitable remuneration. The starting point of our work was centred around these questions:

- *How are earnings distributed along the curve from top earners to the 'long tail' of musical creators who make only very small amounts of money from music?*
- *How much money do music creators (rightsholders) currently receive from streaming?*
- *How much revenue is collected by streaming services through subscriptions, advertising, and any other sources?*
- *How is this revenue divided between streaming services, music publishing rights holders and sound recording rights holders?*

In the subsequent analysis, we point out that any meaningful answer to these questions requires a more precise definition of `earners`, the `long tail`, and the timeframe of the analysis. In any given royalty payment analysis, most works or their recordings do not make earnings—this is particularly true in the case of music streaming. This alone makes the use of the median earning a useless indicator. As we show in subsequent subchapters 1.3.1 Sampling the unknown rightsholder population and 1.3.2 Sampling the unknown universe of works and recordings, it is notoriously difficult to map the distribution of earnings among rightsholders, even when employing advanced surveying techniques or attemting to empirically observe royalty statements.

The question of `how much` will lead to significantly different answers in different years in different geographical markets: we must carefully chose the relevant geographical market in question and the relevant timeframe for the analysis. The publicly available data on the British or the global music industry contains only highly aggregated revenue data, which hide the different movements in volume of use, price of use, currency rates, number of works/recordings uses, and number of users—often even the number of jurisdictions which collected revenues for the British rightsholders. Because there is no comprehensive database of rightsholders, protected works, and protected recordings, along with total revenue data, it is arithmetically impossible to characterize the earning prospects and changes of the typical (median) or average (mean) rightsholder.

Weather we try to analyse rightsholders' earnings in a purely empirical, positive way (How much do they earn? How are earnings distributed) or in a normative way (Are earnings fair? Are they equitable?), we run into the same theoretical and empirical questions. We start with the more normative questions that require a knowledge of music economics, royalty accounting, and copyright law; we then turn our attention to empirical problems with the help of statistics and data science.

1.1 Theoretical Challenges

It is impossible to understand music streaming earnings outside the context of the music market or the music business ecosystem. Music streaming royalty rates and distribution methods largely follow long-established practices and regulations concerning valuing intellectual property, accounting for royalties, and regulating the administration of copyrights and neighbouring rights. Music streaming competes with other uses, such as public performances in television and radio broadcasting, or digital downloads. Music streaming cannot be analysed in isolation from the volume and price trends of public performance, mechanical licensing, and home copying.

1.1.1 Equitable remuneration

In the UK policy debate, a potential redefinition of equitable remuneration rights received noteworthy attention, and it is a good starting point to understand music earnings. Equitable remuneration is originally connected to a compulsory license that must be paid when recorded music is played on a publicly accessible location to the performing artists and music producers.

Compulsory licensing is subject to regulation, because it disallows individual negotiations between business-to-business rightsholders and users about the royalty rates or other contractual terms. (Compulsory licensing never applies to private end-users.) Instead of market forces, it is up to regulation to guarantee terms that are called *equitable* (and in some jurisdictions, *fair*) and similar to conditions that two business partners at arm's length–without buyer and seller power, or cross-ownership–would have agreed upon.

The equitable remuneration right is a legal concept, and not a guidance on how much music creators are earning, nor how much they should earn. The equitable remuneration right calls for fair valuation techniques (see in the next subchapter) and competition law guarantees ensuring that rightsholders are eventually paid fairly.

The legal concept of *equitable remuneration* has an economic aspect. In international law, it was first enshrined as Convention C100 of the International Labour Organisation (ILO), stipulating that men and women should receive equal pay for equal work (ILO, 1951). Within the context of international copyright law, it was introduced as a modification of the Berne Convention by the Rome Convention for the remuneration of the broadcasting of recorded fixation of music works (recordings) since 1971. This right is further elaborated by the WIPO Performances and Phonograms Treaty (WPPT).These copyright conventions are administered by the WIPO (WIPO, 1996a,b). The equitable remuneration rights ensure that users of music and rightholders should settle along terms *as if* they would have negotiated in an open market transaction in the absence of barriers to negotiations and asymmetric power.

These international treaties do not set an international standard on how to calculate equitable remuneration, and do not apply to all uses of music works and their recordings, but rather try to avoid certain market outcomes. For example, compensation must be proportional to use: rightsholders whose works were not used, or used only very few times, should not receive similar remuneration like those whose works were used frequently. Equitable remuneration is itself an important, but justified source of unequal royalty distribution, because the use of music favors novelty and popular performers or composers. The fact that the *equitable remuneration right* itself is the cause of inequality among rightsholders may seem to contradict the everyday use of the word 'equitable'.

The idea of equitable remuneration, and the right for equitable remuneration is not an absolute and well-defined right. It has gradually evolved over the last 60 years, and it has been defined and re-defined many times, like other aspects of copyright law. Because many UK rightsholders have a significant international market presence, they are not remunerated according to a single equitable remuneration right. They receive income from many jurisdictions, which all define differently the equitable remuneration right. There are very significant differences on the distribution of rights between producers, performers, and among performers, featured artists, and session musicians. A British recorded fixation may be subject to differently set rights in in the UK, Ireland, Germany and the United States. In an international context,

a study of Europe Economics and IVIR has shown that there are notable differences in how *equitable remuneration* is understood – and it is often used as synonymous to *fair* remuneration (Europe Economics & IVIR, 2015).

In our understanding, the treaties that define *equitable remuneration* are silent on how to calculate the remuneration, because this is enshrined in *intellectual property law* and *financial law*. The fact that some jurisdictions use the word *equitable* and *fair* interchangeably underlines the fact that equitable remuneration should be based on the principles of 1.1.2fair value, as defined by economics and international accounting standards.

The current UK debate is perhaps too much focused on the UK definition of equitable remuneration rights. The changes promoted by the Broken Record campaign were successfully made in several European jurisdictions, but they did not significantly change the earnings of most creators. By focusing on the re-definition of the equitable remuneration right, the debate is only focused on the re-distribution of a very small portion of the music ecoystems's income. This blurs the fact that the general level of earnings— which is not governed by the equitable remuneration right, but by fair value, arm's length standard, and competition law—is declining and seems to be inadequate compared to historical levels.

1.1.2 Fair value

When rightsholders are not in the position to negotiate rates freely, various legal, accounting, and economic norms, and institutional guarantees must be in place so that they receive fair and equitable remuneration. Because copyrights and neighbouring rights (in European parlance, or recording copyrights in U.S. terminology) are valuable intellectual property rights, they entail fair compensation for the use of rightsholders' works or recorded fixations.

When intellectual property is sold (against a lump-sum payment), or licensed (for periodical payment), or used as a pledge against a loan, the transaction must comply with the *Fair Value* standard. This standard is, similarly to copyright law, set by international law, which generally can only in detail be modified by UK law. The Fair Value standard of the *International Financial Reporting Board* has been incorporated into UK law via EU law, and remained UK law after Brexit.

In 2008, the global recording industry body, IFPI, published *Valuing the use of recorded music*, created by Price Waterhouse Cooper (PwC, 2008). This excellent methodological guide applies the WIPO and IFRS standards (IFRS, 2011; Flignor and Orozco, 2006; Puca and Zyla, 2019) on valuing copyrights in more practical terms for the music industry. It is important to recognize that IFRS, like copyright law, is international law, and it can be enforced in the UK courts. The valuation principles are enshrined in the fair valuation principles of the WIPO, and the fair value principles of the International Financial Standards Board. While the *equitable remuneration right* may only apply, from a legal perspective to a subset of rights, the economic techniques of valuing copyrights, or any type of intellectual rights are similar. Therefore, even when no law or jurisprudence stipulates that remuneration must be equitable, the same economic valuation techniques should be used that we use for setting the remuneration level for *equitable remuneration*.

The recognized fair valuation principles stipulate that the "most applicable method" must be used in valuations. In the music industry, the use of the "cost approach" is completely impractical, which leaves open the choice between the "income approach" and the "market approach".

The *income approach* compares the *expected future royalty cash flows* of a work or its recorded fixation using an appropriate "discount rate." When a user buys in a music store an mp3 file on 1 July 2015, it triggers a single royalty payment after the deductions of the cost of sale on the marketplace. In a streaming platform, the same user's royalty payments appear every month when she listens to the song, and in radio, usually every year. The discount rate provides a proper comparison between remuneration received in July 2015 and April 2021.

The *market approach* tries to identify a payment rate, regardless of whether it is made in lump sums, monthly or annually to established, among sufficiently similar uses. Many ideas were tried at an interna-

tional level to identify the sufficiently similar uses of music streaming; for example, relating ad-supported and automatically selected songs to radio streams, and relating cases where the user controls the selection of songs, and may even download them.

The application of *fair valuation principles* is particularly challenging in the case of private copying, where the transactions are not recorded (as they are not market transactions) and in streaming, which is a relatively new technology that is seen in licensing as a mixture of earlier mechanical copy-based and public performance-based licensing, and has so many transactions that most rightsholders (and even their national organizations) lack the data processing capacity to administer the rights or challenge incorrect payments. In our practice, we usually use the market approach, and we carefully compare after timeframe and other adjustments mechanical licensing revenues and public performance revenues with proceeds from music streaming platforms and media platforms, such as YouTube.

Comparing royalties is challenging because some royalties are paid upfront and for perpetuity (like in the case of the digital download model of mp3 music); music streams are remunerated monthly, based on actual use; and public performance streams, such as broadcasting, are usually remunerated annually, in equitable proportion to the use of the works and recordings. Both monthly and annual royalties are subject to accruals of small sums.

1.1.3   Relevant market for analysis

In competition policy, a relevant market is a market in which a particular product or service is sold—an intersection of a relevant product market and a relevant geographic market. This concept is particularly useful in our case, because music has no clear-cut product, service, or geographical market.

A relevant product market comprises all those products and/or services which are regarded as interchangeable or substitutable by the consumer by reason of the products' characteristics, their prices and their intended use—experience and research in other countries shows that music streaming can be substituted with other, differently licensed music, for example, listening to music via the media platform YouTube, or from mp3 files. Even the sale of physical music products, such as microcasettes and CDs, can be substituted with music services such as streaming subscriptions.

A relevant geographic market comprises the area in which the firms concerned are involved in the supply of products or services and in which the conditions of competition are sufficiently homogeneous. Because recorded music creates royalties from copyrights and neighbouring rights defined by national copyright law, and, with some exceptions (for example, the BIEM agreements), are set nationally, the relevant market for analysis is the national market. British music is licensed in almost 200 jurisdictions, which makes the analysis of music creators' earnings particularly difficult. Yet, there are very few national (or domestic) data available on music sales volumes and sales prices, particularly in streaming. And the empirical observation, either via surveys, self-reporting, or via the observation of various royalty statements is very difficult, because the population of rightsholders, works and recordings is unknown, and representative sampling is difficult. (See subchapter 1.3 Empirical Challenges.)

While observing their global income is relatively easy, it tells us nothing about the economics of those earnings, i.e. the change of demand, supply, volumes, and prices.

1.2   Policy Issues

The MCE project is taking place in a time when serious policy changes and discussions are taking place in the United Kingdom and Europe. As the UK is building a new regulatory system, its legislation must decide on which aspects of the copyright regulation to keep from retained EU law and which aspects to change.

Because of the important global position of the UK music industry, the international policy debates and expected or undergoing changes are relevant for understanding the earnings prospects of music creators. We believe that the most important factors at play are the consolidation of the data management of

a very fragmented music ecosystem, the creation of trustworthy AI in algorithmic music sales and the closure of the value gap towards media platforms. These problems are directly related to the low level of earnings from music streaming, and some unjustified differences among rightsholders.

1.2.1 Problems with copyright management

As we will show in the 3 Differences in Music Creators' Earngins: Simulation Results chapter, the fluctuation in streaming price and currency conversion to the function currency from the market of sale can cause much and justified earnings differences among British rightsholders. But there are differences which are not justified and related to some form of malfunction in the rights management system.

For producers and performers, the problems are partly different from the problems of composers, because streaming platforms license the copyrights and neighbouring rights (the publishing side and recording side) differently. But generally, both publishing and the recording side suffers from the problem of high documentation and data processing costs, high error rates, and large sums of unpaid revenues. The lack of data interoperability, frequent changes of standards, and the general fragmentation of the music ecosystem–which disallows the application of specialized documentation or data management workforce– causes delays or non-payment for rightsholders. In some markets, we observed that this could affect up to fifty percent of all rightsholders.

In the current market situation, with the current streaming prices and streaming volumes, the cost of proper (metadata) documentation of works and recordings often exceeds the value of expected future royalty cash flows. Because metadata documentation and data processing exhibit large economies of scale (i.e., the unit costs are much lower for a million songs than for a thousand songs, and the difference is even greater to a hundred songs), payment problems plague independent labels and publishers. In Europe, the publishing share in streaming revenues is unusually low compared to the recording side, therefore the proper documentation of composer royalties is particularly challenging.

Our Digital Music Observatory, and its related Listen Local Initiative was partly set up to allow the pooling of data processing, documentation automation for smaller organizations, like independent labels, or national organizations of smaller countries, who usually manage a much smaller national catalogue than major music businesses. We are participating in an international research consortium (which includes the UK Copyright and. Creative Economy Centre from the United Kingdom) to understand the problems in data and metadata management and propose copyright law changes in Europe and the UK.

We do not have estimates about the scope of the problem in the UK, but in the United States, after the introduction of the new Music Modernization Act, the Mechanical Licensing Collective received in January 2021 $424.38 million in unmatched royalties from Apple Music, Amazon Music, and Spotify among other streaming services. According to various reports, these streaming services paid accrued historical unmatched royalties, often called "black box money," to the MLC in order to get these royalties to copyright holders, songwriters, and publishers who have applied to become MLC members. Apple Music transferred the most money to the MLC with over $163 million, while Spotify wasn't far behindl submitting over $152 million. Amazon, meanwhile, transferred $42 million, while Google/YouTube sent $32 million. We experienced similar problems with a smaller absolute and a higher relative value on the smaller European markets.

1.2.2 Value transfer

During the creation of this report a legislative process is underway to implement a new European legislation that aims to close the value transfer to unlicensed media platforms. Media platforms, like YouTube, unlike licensed music streaming services like Spotify, Deezer, Apple Music or licensed audiovisual services like Netflix, did not get a license from rightsholders for the use of their music, audiovisual content, literary or artworks. Their objections to such licenses were based on the safe harbor of copyright law mandated by the U.S. Congress to help the evolution of internet businesses.

The representative associations of composers (GESAC, 2015b,a; CISAC, 2017), performers (AEPO-

ARTIS, 2016, 2017), and producers (Moore, 2016) in Europe were all strongly calling for a change in these practices, while technology companies on the other side were vehemently lobbying European governments and members of the European Parliament to stop or water down proposals. Eventually the European Union adopted the new Copyright Directive (EUR-Lex, 2019). The transposition of this European directive into national copyright law is happening in these months all over Europe (or it has already finished.) It is yet to be seen if the regulatory approach will close the value gap in Europe, but it is likely that revenues from the United Kingdom will fall under a different copyright regulation.

Value transfer is undermining mechanical licensing, public performance, mixed streaming licenses, and even compensation values for home copying. They offer a cheaper platform for viewers than licensed onces, which moves the basis of potentially royalty earning revenues towards the media platform. We have measured this effect in several European countries, and we have found this to be far more significant than the potential redistribution effect from pro-rata streaming distribution (see the next 1.2.3 subchapter) or the redefinition of the equitable remuneration right (subchapter 1.1.1).

1. Unlicensed uses create competition to the business-to-business users of music, audiovisual content, books and other copyright protected content. For example, television stations and radio stations lost audiences to YouTube, because YouTube could offer programming "for free", as it paid only about 1-10% of the similar royalties of broadcasters for content. This means that advertising revenue moved from TV and radio stations to YouTube. This revenue should have been shared with rightsholders by the broadcasters.

2. The availability of unlicensed platforms, such as YouTube, that claims not to be a streaming platform, and pirate websites reduced demand for licensed, legalized use of copyright protected content. The lower sales volumes and the direct loss of advertisement revenues reduced the market price of content.

3. Because private copying remuneration must be based on market values of music, filmed content, literary works, and other works, it is important that we enumerate the price loss due to illegal activities, so that the damage claim is truly based on arm's length rather than illegal market transactions.

We were not provided with data to estimate the size of the value transfer in the United Kingdom, but given the global nature of the music ecosystem, we do not believe that the British rightsholders are affected differently than European ones. Partly, because a large part of the British revenues are export revenues to the European Union, and partly because unlicensed media platform are competing precisely in the same manner in the UK with licensed public performers or streaming platforms like in other countries. We believe that the existence of the value transfer is one of the sources of the declining value of music shown in subchapter 2.3, but we have not attempted to estimate the size of this effect. However, our detailed data in Hungary, Croatia, and Slovakia suggest that the value transfer amounts to 10-20% of the total music market value.

1.2.3   Distribution rules

Recently we have seen a considerable discussion around the choice between the *pro-rata* versus *user-centric* model of royalty distribution from streaming platforms. Except for SoundCloud, and only from April 2021, music streaming services do not divide up the royalty fund (from subscription and advertising revenues) based on the number of streams per recording. We did not have a mandate to analyze this issue deeper, but we believe that in the long-term, the choice between these two models is a false dichotomy, and the music industry must look beyond its boundaries, for example, to telecom services, to find a better and more just distribution system. We summarize empirical studies made in Denmark, Norway and France in subchapter @ref(#earnings-distorsion).

1.2.4 Trustworthy AI

Recommendation systems of music and media streaming platforms are extremely complex systems, and they operate as a proprietary marketplaces where the rules of fair competition, fair play, and equitable remuneration are basically maintained by the owner of the platform. Forthcoming UK and EU regulations try to make these autonomous or semi-autonomous algorithmic systems accountable.

One important element of work we are doing with the founding partners of our Music Observatory–with the support of the Slovak Arts Council–is to find out why classical musicians are paid less, or why 15% of the Slovak, Estonian, Hungarian and Dutch artists never appear on anybody's personalized recommendations (Antal, 2020b).

When we analyze the effect of AI algorithms on royalty payments, we describe potentially bad outcomes of the use of automated sales or personalization, which undermines or contradicts national or EU-level cultural policy goals, national media regulation, national child protection rules, national or international fundamental rights protection against discrimination without basis. In a recommendation system many bad outcomes may happen that can eventually lead to lower or no payment for a rightsholder.

In these cases, we are not talking about *unjustified* differences of payment as simulated in 3 Differences in Earnings: Simulation Results, but a systematic breach of non-copyright rules that lead to an *unjustified* lower streaming *volume*, and therefore a lower royalty payment on lower volumes. A recommendation system might exhibit the following characteristics:

- It may recommend too few female or small country artists, or start recommending artists with hateful language.
- It may put certain label's music on less visible places.
- It may make the works of major labels easier to find than those of independent labels.
- It may place less works and recordings on personalized lists than local content guidelines (applied in about 90 territories of the world, see (Stein et al., 2012)) would require from competing local television or radio stations.
- Your personalized list is filled with misleading track/artists names that are hunting for accidental streams, much like click-hunting text content that plays search engines.

These *potential* undesireable outcomes are sometimes illegal, and, for example, they may hypothetically make Welsh artists earn significantly less than American or English artists.

In our work in Slovakia, we reverse engineered some of these undesirable outcomes and we are planning similar research in the United Kingdom. We do not have relevant empirical estimates for the United Kingdom, however, and our experience shows that one potential reason why UK rightsholders are facing such a different outlook on streaming revenues may be due to various bad outcomes of the largely autonomous system of playlist personalization and playlist curation. We want to stress that these problems are unlikely to be the results of foul play by streaming providers, but they are inherent to extremely complex systems that need constant and complicated monitoring.

1.3 Empirical Challenges

Many people are familiar with the work of *Sixto Rodriguez*, a Detroit-born songwriter, whose works were not licensed for more than 30 years after they were recorded between 1967 and 1971, until 1997, when it was famously discovered that he is a best-selling songwriter in South Africa. After the documentary *Searching for Sugarman* won the Academy Prize, his recordings, 40 years after their release, became golden and platinum records in many countries.

Many artists are active in their twenties, and in the 21st century they can expect to live 60-70 years longer, and their heirs enjoy copyright protection for further decades. Like Rodríguez's recordings, there may be many protected works and recordings that have not been observed in distribution schemes, media and streaming platforms, yet become successful in the future. In the absence of a global, compulsory

registry of copyrights, we do not even know how many works and recordings enjoy copyright protection in 2021.

Because there is no global database of copyright (and neighbouring right) protected music works and recordings, we do not know precisely how many rightsholders and how many works or recordings are protected by the law. From a strict mathematical perspective the average (mean) and typical (median) rightsholder earnings do not exist, because their calculation would involve division by, or ranking of an unknown number of rightsholders.

1.3.1   Sampling the unknown rightsholder population

Most rightsholders create works and recordings that have a limited lifespan: in a few years they lose their audience, and in the absence of their use, they no longer create income. But, as the case of Sixto Rodríguez shows, they may in the future, even decades, or a hundred years later. There are many rightsholders who create something, but never become full-time creators, or eventually they go back to part-time or occassional creators. Creating a representative survey of rightsholders is very challenging, becaues the rightsholder population has many inactive members.

Similarly to rightsholders, most of the works and recordings are not in use, or hardly in use at any given time period. Creating a representative sample of recordings or works for empirical observation is very difficult, because we simply do not know how many works or recordings exist.

Since 2014, we perfected sampling both rightsholder populations and recordings to create better and better representations of the creators and their creations. We created special surveys among music professionals to establish their typical earning levels and composition, and we created special samples of recordings to observe their typical revenues.

The way we sample or weigh musicians' self-reported earnings in the Digital Music Observatory's surveys is based on the known, but anonymised payouts of author's societies, and comparing it to the self-reported earnings from author's societies. The closer the self-reported median, quartile, or average payouts get to the know, but anonymous true payouts, the better the sample. In Hungary and Slovakia we were able to create highly representative music professional surveys for the active musician population. On active musician population we meant those living artists, who recieve some collectively managed royalties in a given year, i.e., their works are played on radio, television, directly identified public performance, collectively managed streaming services, on concerts, or other uses where author's societies record use and collect revenues. We believe that this is the best target population for investigating music creator's earnings. This target population is usually rather different from the payrolls of author's societies, not to mention neighboring rights soceities, becasue those contain heirs of many deceiced artists. This target population does *not* include new artist who do not have yet a registered work or commercially released sound recording, or whose works are not exploited in these channels. Generally, this is an appropriate target, though in some genres they may not be representative. (For example, old and classical musicians play music that has no author royalties, and some subgenres of hip-hop are not radio friendly and avoid any registration.)

1.3.2   Sampling the unknown universe of works and recordings

The approach we used for the Central European Music Industry Report, which is so far the most comprehensive European music industry report, was a systematic sampling of a very large portfolio managed by Consolidated Independent (in the state51 music group.) With the help of state51 music group's engineers, we pooled the royalty statements of several million recordings, and chose to examine further those which were used in one of the four services (Spotify, Deezer, Apple, YouTube) in 20 select markets, including the United Kingdom. We did not believe that the Consolidated Independent portfolio was representative of the entire European or each of its national markets. We used a sampling technique reminiscent of some bond and stock market indexes to select recordings that were good candidates to represent the earnigs of a typical recording in a given market and time period. Our 'index basket' chose the median use of more than 300,000 recordings for 20 distinct national copyright jurisdictions and months. The index basket

was 'rebalanced' every month: songs that were used more or less than the median value were replaced with the median used (played) songs of the month. We wil introduce these indice in the next chapter.

Our indexing was deliberately experimental, and connected to the creation of the report, not the Music Creators' Earning project. We believe that more sophisticated and even largely different sampling techniques could be used for understanding earnings in streaming channels. Our selection method tried to find the 'typical' song that was played in a given months, which is very different from the 'typical' recording. In any given month, in the CI portfolio, as in any large portfolio, the 'typical' recording was not played, and the median (typical) earning was zero. We avoided this problem with selecting every month the 'typical played' song, but if we would like to characterize the 'typical' rightsholder, we would have to give up on the use of both the arithmetic mean (which is cannot be calculated as we do not know the number of protected works/recording) and the median (because it non-descriptively zero.) We still believe that the CI-CEEMID index is the probably the best descriptive statistic that can characterize British and European individual rightholder earnings.

1.3.3 Market geography

Aggregated income indicators are usually made in two versions: national or domestic. The national music creator earning is the earning of rightsholders who live or are registered as legal persons in the United Kingdom. The domestic income of the music industry is the earning made by entities in the United Kingdom, including the earnings of non-resident rightsholders they represent. The earnings of a French artist residing in the United Kingdom, represented by a French label, is likely to be counted in the British national earnings, but not in the domestic earnings of the British music industry. However, a Dutch artist residing in the Netherlands but represented by a British label is likely reported in the domestic revenue of the British music industry (and add value to the British GDP.)

Defining the geographical scope of the analysis is relatively easy in the case of live music, and rather complicated in the case of recorded music. Because copyrights (and neighboring rights) are nationally granted on the basis of international treaties, rightsholders—via distributors or collective management socieites—collect their revenues from each copyright jurisdiction separately. However, this information is often lost in surveys and global reports. In some cases, tying the revenues to a country, or even a region of a country is possible—this maybe possible in the case of radio revenues or royalties connected to live performances.

The loss of information, as we will show in the next chapter, is critical, because volumes, prices and exchange rates move rather differently from the UK market in other countries. German revenues are driven by different volume growth rates, different local price fluctuations in the value of a stream, and by the EUR/GBR rate applied whenever a monthly or accrued royalty is paid. Because practically all stakeholders have non-UK revenues, without knowing how much they are affected by different international factors, they cannot know what steps in Britain can lead to a positive change in their visibility or income goals.

The British music industry is not constraint by language barriers, and most rightsholders receive income from dozens of territories or jurisdictions. This means that British rightsholders collect from many jurisdiction, subject to various definitions of equitable remuneration, public performance, mechanical royalty regulations, and in many currencies. Because of the significant differences in streaming prices across the world, and the presence of currency rate fluctuations, a very large part, if not most of the income differences of rightsholders are caused by the different geography of their audience distribution.

The Digital Music Observatory uses data from righthsholders directly, from collective management and distributors. It is not always possible to break down the figures to national territories—for example, in the case of musician surveys, it would not be possible to ask musicians to break up their royalty statements when they self-report revenues.

This is an important problem, because each artist, label, may have a different share of British, Irisih, American, German, and other income, and therefore their reported income may contain an unknown element of foreign income, and an unknown currency conversion effect. For example, in the 2016-2020

period the British pound generally lost value against the euro, so a label or artists with flat revenues partly arriving from the eurozone could have seen a rising pound revenue.

The CEEMID-CI indices are denominated in British pound, and each revenue is exactly tied to a jurisdiction. This cannot be said of self-reported surveys like that of the UK Music in Numbers, and we believe that there is no mechanism to prevent reports to IFPI to clearly delineate revenues per market. In the case of the CISAC Global Collection Report, we have a different problem: while we know that all reported income was generated in a particular national jurisdiciton, often a large share, even majority of those collections were made on behalf of foreign rightsholders.

1.3.4 Units of observation

Music is a global industry, and even less known artists or songs find occasionally international audience on streaming platforms. This means that practically every British rightsholder has *some* revenues that are *not* originally denominated in British pounds. In the post-Brexit period, the British pound lost value against the US dollar and the euro, which means that a British artist could have received the same 100 GBP revenue even if her euro or dollar quantities or prices were falling.

For any meaningful economic analysis, the volume, price, and exchange rate effect must be separated.

Revenues are multiples of use volumes over price expressed in a certain currency. They can be national or domestic market indicators if they contain revenues only domiciled rightsholder or neutralized rightsholders' revenues. Without the ability set up apart foreign versus national *or* domestic revenues, and break up the effect of changing volumes, prices and exchange rates, they are not suitable for economic analysis, and they do not indicate the earnings of the *average* or *typical* rightsholder. They still maybe useful for business target settings, but only for organizations with a large and diversified portfolio.

There are several problems with the existing national and international music industry reports:

- They usually do not separate domestic, national and foreign income.

- They do not consider the exchange rate, or use currency translations that may not be adequate for every type of analysis.

- They do not separate the effect of price and volume change.

- Sometimes they do not even control the number of territories (jurisdictions from where royalties are received.)

These problems give rise to the problem that growth in aggregates may lead to a diminishing revenue for smaller entities or individual artists. These indicators are not adequate for the use of small publishers, labels or managers of bands, artists, because they do not indicate growth or decline in volumes (and signal for changes in promotion), do not indicate growing or declining prices (and allow the focus of sales on more lucrative segments.) While currency risk can be hedged or otherwise insured, they do not even show which part of the earning change is attributed to exchange rate movements and which is related to the market performance of their catalogue.

Ideally, the data collection should record volumes (number of streams), prices (value of individual stream), and exchange rates (conversion rate applied for the particular stream) for all significant markets of the artist, company or national organization. Currency translations must be consistent and meaningful.

1.3.5 Timeframe of analysis

As we have experienced during the creation of the CI-CEEMID indices, most recordings' typical (median) earning over a single royalty payment cycle (a calendar month) is zero. But it is certainly not zero during the entire lifespan of the recording, and it is usually not zero for the first or second year. The choice of the timeframe for empirical observation is critical.

Various licensed uses of music have three royatly payment cycles. Mechanical royalties are paid once in a lifetime. Public performance royalties are collected on an annual basis, and small earnings (that would

be too costly to pay out during the year) are accumulated over two or more years. Streaming earnings are paid out monthly, but similarly to public performance royalties, small amounts are accumulated over two or more periods. Observing royalty statements alone cannot enable meaningful comparisons, because the same payouts on a given date–for example, 30 April 2020–refer to different earning periods.

Periodical earnings often do not reflect accrued but unpaid royalties, which may affect a very large number of rightsholders, given that the earnings in the long tail are smaller than the accounting and bank transfer cost. These accrued earnings are not always lost, and often carried over to the next period when the payment, after payment costs, is practical. It is very easy to significantly underestimate the payouts *in the long tail*, because their *payment frequency is lower than among successful artists.*

The different timeframes of various royalty collections has another impact on the analysis: different royalty cash flows are subject to different currency exchange rates. Two, seemingly equal royalty payments may veil a different quantity or price by an offsetting currency movement.

1.3.6   Survey design and harmonization

Because surveying musicians and rightsholders is very difficult, we use various techniques to improve our results. As mentioned earlier, we can use known payouts (such as various collectively managed income paid to living British rightsholders) to create post-stratification weights for responses. Another possibility that the Digital Music Observatory uses ex ante and ex post survey harmonization. Asking the same questions with the same methodology used in large, representative, national or Pan-European surveys allows us to understand the various demographic biases of musician surveys.

For example, in the Central European Music Industry Report, we used relative income and life satisfaction questions that have been asked for decades by Eurobarometer, and compared results. In some countries, we could create a representaitve sample of musicians; we could observe that everywhere musicians' earnigs were more precarious than that of the local population. A major advantage provided by survey harmonization is that we can compare musicians' earnings with various occupational groups, too.

Our suggested approach to the MCE project was the use of survey harmonization and post-stratification, but this was turned down by the Steering Group.

When we are surveying rightsholders, which I have done so far in 12 countries, we must keep in mind time and again that respondents are typically drawn from a very small minority of the rightsholders. Music careers are not always linear or exponentially growing, many rightsholders are inactive for years or decades when they often receive very little or no earnings, and do not participate in any surveys. Heirs almost always decline to participate in any music earnings survey. Statistically valid surveying is possible but challenging – in our surveys we use ex ante and ex post survey harmonization with national and international standardized surveys (which are available in the United Kingdom, too) and we compare surveyed, anonymously reported earnings with anonymised known earnings. For example, in the CEEMID surveys conducted in Hungary and Slovakia, we compared the anonymously reported income from Artisjus and SOZA with the anonymised, full payout of actual Artisjus and SOZA royalties to *living* rightsholders. Our successfully distributed surveys very accurately resembled the known, true payouts. Because the distribution of the earnings was very similar in 12 analysed countries, and in two we could very carefully compare the reported earning distribution with the true distribution, it is safe to assume that British payouts have a similar form.

In any empirical analysis, it is important to note that usually we can observe a minority of the works or recordings, and a minority of the rightsholders. If we observe actual streams or payouts, most of the works/recordings are not used in any given period, and the second largest group is so scarcely used that the payout is carried over to a next period. In any given period, most of the works/royalty payments are not observed; additionally, the end of the long tail is also not visible.

Chapter 2

Music Creator's Earnings

The global music industry has two well-known global reports: The Global Music Report of IFPI, which covers total revenues in almost all territories for producers and partly performers; and the CISAC Global Collections Report that contains total revenues from collectively managed author's revenues. Both organizations collect (rather different) international data which is only available to their members. There are also national music industry reports available with variable depth of content and analysis, like the UK Music in Numbers series. The more fragmented live music industry does not have any comprehensive global or European report. We also do not have a truly comprehensive report on global publisher revenues that are not collectively managed.

These reports, because they were designed to set business targets for larger organizations, and not to analyze the economics of music or to provide guidance to smaller entities and individual rightsholders, are not suitable for the analysis of music creator's earnings. They are also not suitable to govern business decisions of independent labels and artists, or to provide evidence for the claims when they do not receive a fair or equitable remuneration. Indicators for the average or typical rightsholder, work, or recordings cannot be inferred from these reports.

The Digital Music Observatory—following the blueprints of the planned European Music Observatory—was set up with many years of work to overcome these shortcomings. Originally called the CEEMID initiative (Artisjus et al., 2014), it started data collection and data harmonization among relatively data poor Central European countries, but as experience showed, Western, more advanced, and future markets like Armenia share vey similar problems.

This was the reason why Consolidated Independent (from the state51 music group) teamed up with CEEMID to make a showcase for a permanent, harmonized, international data collection program. Our Central European Music Industry Report was prepared with the help of 60 music organizations in 12 countries, including the United Kingdom, and analysed rightsholders earnings in various distribution channels, including, but not limited to streaming, in 20 markets.

The idea of this observatory was brought to the UK policy debate on music streaming by the observatory's only (former) British users, via the *Written evidence submitted by The state51 Music Group* to the *Economics of music streaming review* of the DCMS Committee (state51 Music Group, 2020; Antal, 2020a).

> "There are instructive initiatives in other industries in which there is perhaps a clearer and longer standing recognition of the role of economic analysis. This sometimes results in initiatives such as 'Observatories' like the European Market Observatory for Fisheries an Aquatorial Products or the European construction sector observatory [...] These tend to be collaborative endeavours, with a varying mix of government, industry, economists and in some cases funding bodies. [...] To date there have been few if any entities or initiatives for music similar to the above-mentioned observatories. We suggest this is something that policy makers can support

and encourage, but which ultimately needs to be driven by the industry itself. […] This is one reason we have worked with the economist Daniel Antal and his team, in particular on the Central European Music Industry Report 2020. Economists such as Daniel Antal produce data about the music industry that is consistent with international statistical standards and adhere to rigorous data ethics principles, seeking external validation through data and code repositories for underlying data and methodologies."

The data observatory concept is derived from Earth and natural sciences, where often many research stakeholders build large observation stations, such as the Hubble telescope in space, or CERN. Data observatories are managed by a triangular stakeholder base of business, scientific, and policy stakeholders. The music industry requires a permanent market monitoring facility to win fights in competition tribunals, because it is increasingly disputing revenues with the world's biggest data owners.

This was precisely the role of the former CEEMID program, that was initiated by a few collective management societies after a dropped GESAC project. Starting out from three relatively data-poor countries, where data pooling allowed rightsholders to increase revenues, the CEEMID data collection program was extended by 2019 to 12 countries. It was eventually transformed into the Demo Music Observatory in 2020 (Antal, 2021), which is now open for any national rightsholder, stakeholder organization or music research institute.

The average and median (typical) earnings cannot be calculated arithmetically—the average does not exist if we do not know the number of works, and the median cannot be observerd if we do not know how many works, recordings, or rightholders to rank. The mean values are also rather useless because of the large concentration of revenues in a greatest hits segment.

This problem is not specific for the music industry. Copyrights and neighbouring rights are claims for future cash flows, just like stocks and bonds. Many of the aspects mentioned above are present in the empirical analysis of stock and bond securities markets, where the global universe of investible securities is vast and changing daily, particularly in the bond market. Bonds often have only 30- or 90-days lifespan, they "expire" and new bonds are re-issued. Most copyrights and neighbouring rights only earn revenues in the first 1-2 years of their lifespan but remain in the copyright-protected universe potentially for 100-200 years.

- The fact that copyrights and neighboring rights are fixed in term (even though often they are fixed relative to an unknown event, the death of the composer), they are similar to fixed income (bond markets). Eventually all rights will enter the public domain and stop paying royalties.

- The fact that each music work and recording has a variable number of streams at a monthly varying price in varying territories makes streaming similar to dividend-paying stocks. Streaming payouts, like stock dividends or prices, are autocorrelated (March offers an insight into April earnings) but unpredictable in the long-run. Many songs enter a similar phase where they stop earning any royalties—much like the stocks of companies that cease to exist.

The reports of IPFI and CISAC, in a financial market analogy, are describing the annual growth of market capitalization, but not the individual or even typical performance of assets. Market capitalization can grow as the number of investible stocks, bonds, or streamable songs grows, even if representative sales volumes and prices fall. These are not useful for the characterization of the economic situation of investors or rightsholders. Instead, we pioneered the creation of indexes that represent the market view of a particular rightsholder.

The CEEMID-CI indexes slightly resemble the approach of the Standard&Poor, Dow Jones, or iBoxx indexes. Indexing is complicated—the formula and know-how to create these well-known indexes is protected intellectual property, and we only used some of their methods to create our proprietary streaming indexes. Our indexes were not created for the MCE report, and we would have used a different indexing for this purpose, but we believe that our "typical" indexes are very useful for the aims of the MCE project.

Our indexes paint a more rosy picture than that view from a 'typical' rightsholder. Our index shows what happens with the typical song of the month. Every month, we selected those songs that were performing

better than half of the songs in a particular country, and worse than the other half in the CI portfolio. This is the median value of the songs that were listened to by anybody in a particular country, and not a measure of all possible songs. The median (typical) value for all songs is zero, because there are so many more copyright and neighbouring right protected works and recordings available every given month than the amount which are played at least once.

In the period of 2015-2019, the typically exploited song in a large, independent portfolio exhibited the following characteristics:

1. The audience expressed in the volume (quantity) of the the monthly streams was generally increasing.
2. The monthly revenues were flat.
3. The prices of a single stream were declining.
4. The variability (risk) of the earnings from month to month was greater in emerging markets.
5. The greater risk in emerging markets was compensated with higher volume growth, and higher revenue growth, just like in financial markets.
6. The revenues of poor markets were high relative to the household cultural spending differences. Often we felt that the streaming providers made pricing mistakes, or inconsistent pricing in national markets. The likely cause is that neither the streaming providers nor the rightsholders made adequate market research into sales and price planning in the smaller markets, and naturally this increased the riskiness of international revenues.

2.1 CEEMID-CI Streaming Value Index

The revenues for the typical used song were roughly flat in almost all markets. This means a decline in streaming value in the case of flat volumes. Our finding is not contradicting the IFPI reports of rising total value. Streaming services are licensed to more and more markets, and more and more recordings (from the newer territories, from the back catalogue, and new releases) are becoming available to a growing global audience. Because of the internationally competitive British repertoire, this means growth for the UK industry, at least in the *domestic* perspective, though in the *national* perspective the growing international competition may lead to loss of market share for invidiual rightsholders who are UK nationals.
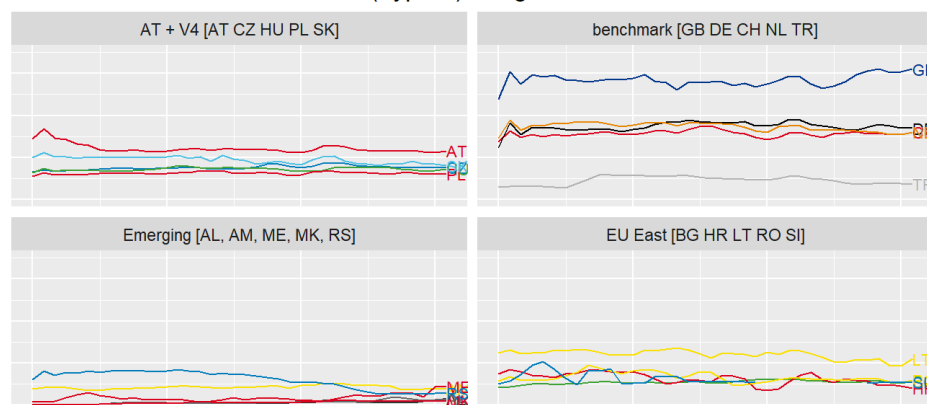
Figure 2.1: Total Monthly Streams of a Typical Song in the United Kingdom and 19 European Markets

2.2 CEEMID-CI Streaming Volume Index

The streaming volumes in the earlier saturating, advanced markets of the United Kingdom and Germany showed slightly increasing or flat volumes for the typical used song. This is likely to mean a decline for the typical song, but the typical (median) song is not played in any given period, i.e. month. Emerging markets, where the service is later introduced, show initial growth.
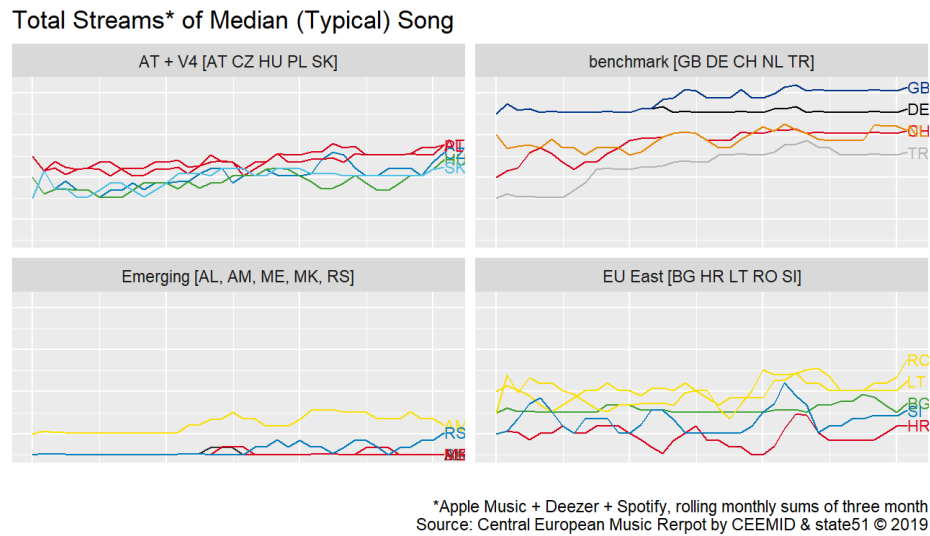

Figure 2.2: Total Monthly Streams of a Typical Song in the United Kingdom and 19 European Markets

2.3 CEEMID-CI Streaming Price Index

When volumes are rising and revenues are flat then prices must be falling. We have seen lowering prices per stream in almost all territories, expressed in GBP terms. Germany's flat price in GBP is the result of a falling euro price offset by the devaluation of the British pound against the euro.

Streaming services logically target first the upper middle class users, and then offer cheaper subscriptions for families or students. The falling unit price follows the profit optimisation of streaming providers.

At first sight we may be surprised by relative high earning potential of poor markets like Albania. We attributed the relatively high remuneration in these countries to a low level of competition (few subscribers, fewer exploited songs) and the effect of various minimum licensing fees.

In fact, we can see the very same effect within the United Kingdom, where the relatively new Apple Music behaves like an "emerging market" provider. In our report we hypothesized two explanations.

- Rightsholders set a minimum licensing fee at the beginning of a new license. In the first months, when there are only a small number of subscribers who stream a small number of songs, the average revenue on each song is relatively high.
- There is a price discrimination present. Each service provider sells for the most interested consumer groups first, then offers various sales, and then rolls out family packages, where the subscription fee per household member is very low.

While it is tempting to single out Spotify as a lowest value provider, Spotify usually generates the highest volumes. The streaming service providers are rolling out subscriptions to different parts of society, and at different costs. For Spotify, the largest marketing cost is access to Apple's in-app purchase system. Spotify has challenged Apple on similar related cases in the UK, in the Netherlands and in EU jurisdictions (Chee, 2021).

In this period, we did not see a very strong logic in how the subscription fees were set by the various streaming providers, though generally speaking, they set higher subscription fees in richer markets.
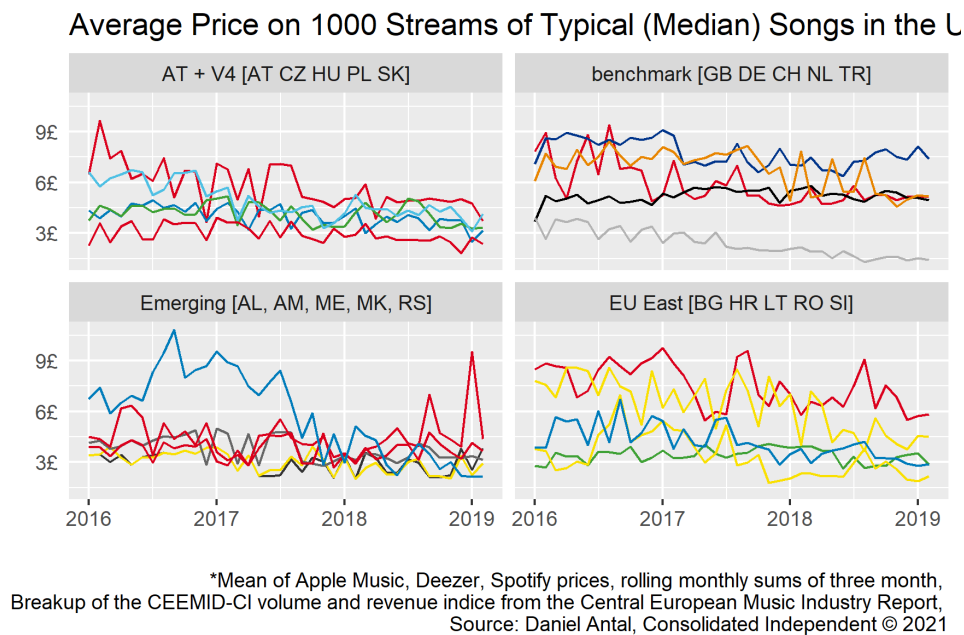
Average Price on 1000 Streams of Typical (Median) Songs in the U

*Mean of Apple Music, Deezer, Spotify prices, rolling monthly sums of three month,
Breakup of the CEEMID-CI volume and revenue indice from the Central European Music Industry Report,
Source: Daniel Antal, Consolidated Independent © 2021

Figure 2.3: Total Monthly Streams of a Typical Song in the United Kingdom and 19 European Markets

Average Price on 1000 Streams of Typical (Median) Songs in the U

*Mean of Apple Music, Deezer, Spotify prices, rolling monthly sums of three month
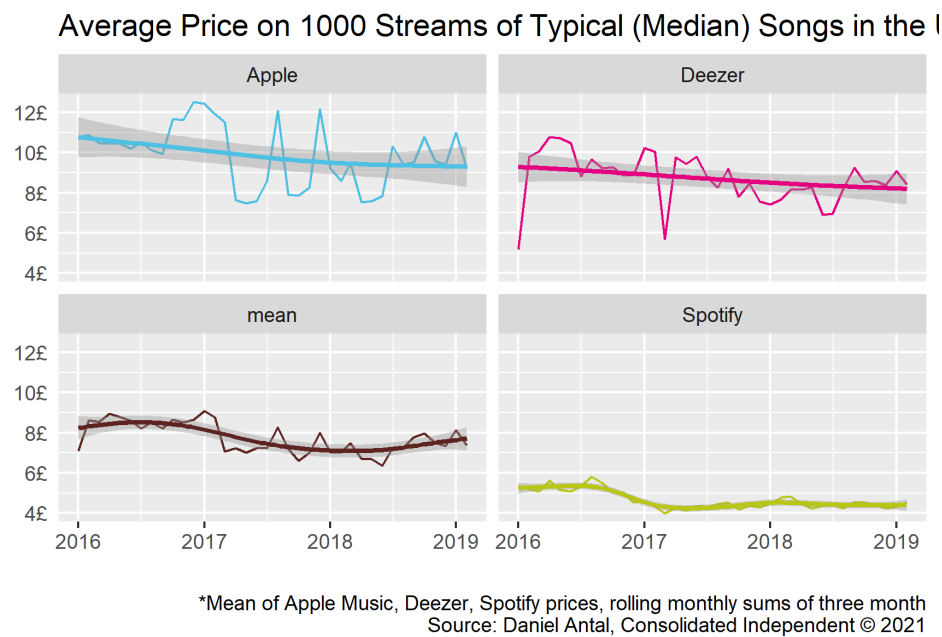Source: Daniel Antal, Consolidated Independent © 2021

Figure 2.4: Quarterly Moving Average Price of a Typical Song in the United Kingdom

Richer markets had higher subscriber bases, higher subscription fees, but also a higher level of repertoire competition—as larger parts of society could afford to buy electronic appliances and pay for subscription fees, they were interested in a broader repertoire of music. In small markets, where usually upper-middle class subscribers were present, the repertoire used was more homogeneous. In the 19 markets, we saw very different levels of interest for domestic and foreign music.

The index values for the United Kingdom can be used as a benchmark for domestic revenues for the UK Music Industry, particulary for smaller rightsholders who are British nationals or British taxpayers. But because of the international competitiveness of the UK repertoire, even these artists are earning foreign revenues. For benchmarking the domestic (as opposed to national) revenues, an internationally diversified portfolio is a better point of reference. The UK music industry often represents foreign artists, too, who may have their main audience outside the UK.

One measure of competitiveness for a recording is its ability to draw revenues for many jurisdictions. Music is a predominantly local business: most musicians have a smaller fan base than a country. Some bands may have followers only in Cardiff, or in Wales; some Welsh bands may be well-known in the entire United Kingdom but not abroad. And some of them are streamed globally. Any artists who can draw revenues from more lucrative markets than the United Kingdom may be better remunerated than the 'typical' British artist. And an artist who has many followers in relatively less lucrative markets, for example, in Eastern Europe, may receive less remuneration on the same number of streams.

2.3.1 Publishing side

We do not have direct information about the publishing side, and we believe that this is more difficult to observe. Some rightsholders, particularly large publishers, are licensing their songs directly to YouTube, Apple Music, Spotify and Deezer, and we do not have access to the terms of these agreements. Other rightsholders are collectively licensed. These licenses are not public.

Generally, the revenues of the publishing side should move parallel to the recording side, but in our experience, because of the lower value of the publishing side, and the hybrid licensing model (direct or collective), far less care has been exerted to correctly document works. Our experience outside the UK shows that uncollected revenues may be significantly higher for the publishing side.

2.4 Unpaid Revenues

In the streaming services, a very significant number of uses are not matched with rightsholders. There are many reasons for this: - There are billions of transactions that have a very low value; the financial motivation to fix small administrative errors is minimal. - Because of the very low expected income for many works and recordings, often the present value of the proper documentation of the work and recording exceeds the likely present value of future revenues. The documentation is patchy. - There are no harmonized metadata standards, and righsholders cannot keep up with the documenting needs of billions of transactions.

The problem of these uses is that they use up a large part of the total revenues (that are financing both royalty payments and operating costs, including identification costs). The cycle of these unpaid royalties is seemingly long, but due to the complexity of the problem, very often the problem is never solved on time. These revenues, after cost deductions, will eventually land in some form of a 'fund' that will re-distribute these revenues to other rightsholders. This effect is a redistribution from less documented and typically small catalogues to large catalogues.

In emerging markets, the problems are so persistent, that often they are not solved over several resolution cycles. We have all reason to believe that the problem is more severe on the publishing side, at least partly because the value of the publishing side is smaller in streaming, and the licensing model is more complex—there are more things that can go wrong, and the revenue pot to solve the problem is smaller (Senftleben et al., 2021).

We do not want to suggest that all factors always move against niche genres and independent labels. This example tries to explain why a very fragmented industry like the UK music industry can face so different revenues and economic landscapes. An empirical examination of this problem is even more difficult than observing the average or typical rightsholder revenues, because the magnitude of the problem is unlikely to be independent of firm size, language, and other characteristics of the repertoire documentation.

2.5 Distortion of Earnings Distributions

In a study prepared by Rasmus Rex Pedersen for Koda (Pedersen, 2020), the Danish rightsholder organization, local artists would have gained from WiMP in Denmark, August 2013 a 9.86% from a user-centric payout. In Norway, the local gain would have been 12.89%. In large repertoires, like the ones of Universal, Sony, WMG in 2013 the wins and gains were no more than 0.8%.

The study prepared by Deloitte for the CNM in France showed that various genres that are more liked by dedicated music fans, such as classical music, blues, and jazz, would have gained on average 0-20%. They have also shown that because much of the redistribution affects the very long tail, very few artists would really feel the change.
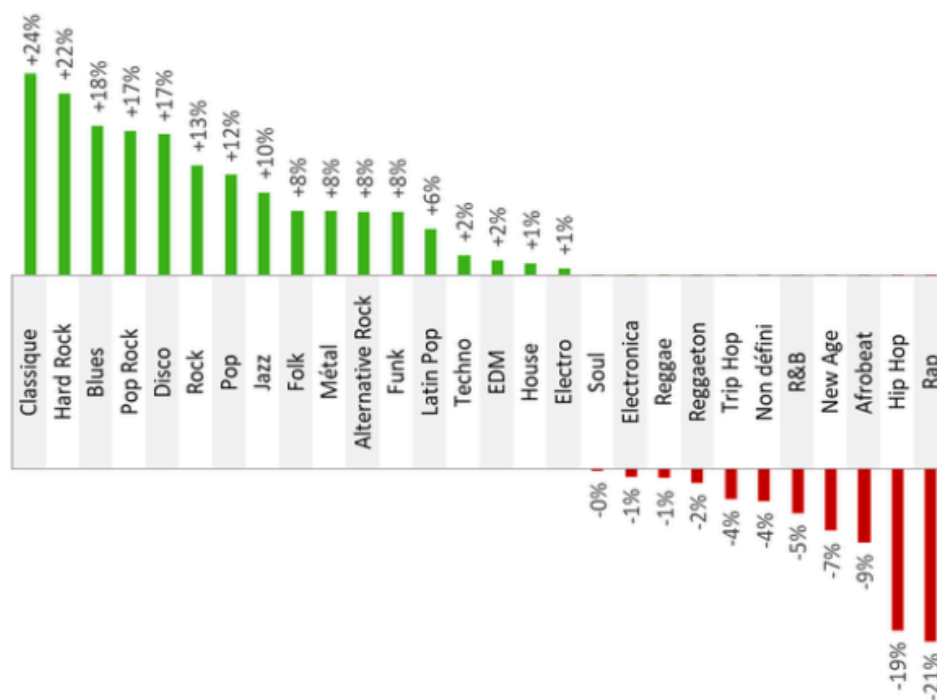

Figure 2.5: Typical difference between pro-rata and user-centric (hypohtethical) distribution, Deloitte 2021.

The difference between the two models is the weighting of subscribers. The pro-rata model divides up every user's subscription fee based on the user's use percentage. The user-centric model divides up the entire subscription fee on the percentage of the total streams, i.e. giving a bigger weight to users who listen to music more. Both distribution models are fair but distribute welfare differently. To use an everyday analogy, the pro-rata model keeps the tips at each waiter and waitress. The user-centric model collects all tips into a jar and divides it up among the staff of the restaurant.

The *pro-rata model* overweights young audiences, who tend to listen to more music, but also tend to choose the cheaper subscription models (ad-supported, family.) The *user-centric* model weighs users by their subscription fees, and they give a large weight to middle-aged and older users who tend to pay for more expensive subscriptions.

Older, and more affluent audiences tend to listen to music that they discovered when they were young

(Holbrook and Schindler, 1989; Holbrook M.B and Schindler R.M, 2013; Stephens-Davidowitz, 2018; Ong, 2019), and which thus tend to be older artists. The pro-rata model favours artists who are listened to by young audiences, in particularl hip-hop, rap, and afrobeat, while the user-centric model favours classical, blues, jazz, disco, and rock, among others, which are more favored by middle-aged or older audiences.

> Consider user A who pays 10 pounds for a service and listens to 10 songs in one month: this user "pays" for each of these songs as if each were worth 1 pound. The more dedicated user B, who listens to 333 songs, "pays" each song at about 0.03 pounds, or 3 cents. In economic terms, if User A does not cancel her subscription, we can say that her subjective valuation of playing a song is at least one pound. For User B, it is at least 0.03 pounds. Would user B be willing to pay 10 pennies per song, and end up paying £ 33.33 for her subscription? We do not know. We only know that she is willing to pay at least 3 cents for listening to one song.

> The **pro-rata model** distributes to a song that was listened to three times by both User A and User B a total of 3 pounds + (3 x 0.03) pounds, all together 3.09 pounds, which after rounding will be equal to 3.1 pounds. A total of £20 was distributed in this case.

> The **user-centric model** divides up the total 20 pounds on 343 song streams, and awards 2x3x(20/343), or 35 cents to the song. Again, this algorithm will distribute a total of £ 20.

This seems fair for the artists, but not for the users. Let's imagine an **artist-centric model** where each user pays a subscription fee pro-rata her consumption. If we would set the price of each stream at 0.058 cents, user A would pay only 0.58 pounds per month for the subscription, while user B would pay 19.42 pounds for the higher consumption. The total subscription fee in this case would still equal to £20 for distribution, and in this case, the song would earn 2 x 3 x 0.0058 pounds, or 3.5 cents.

We could argue that the artist-centric model would be probably most in line with the equitable remuneration principle because each user who enjoys more music pays more. The pro-rata model can be seen as price discrimination model, which allows less price sensitive consumers to pay more for each song. The difference between the pro-rata and the hypothetical artist-centric model is the redistribution of the consumer surplus, i.e., the value that User A was willing to pay 33.3x times more for each song than User B. In the *pro-rata model* the entire consumer surplus is distributed to the song which could "exploit it", while the *user-centric model* shares the consumer surplus among all rightsholders.

The effect is not small though, and it is multiplicative. To stay with the earlier example, when the GBP appreciates in value by 10% against the euro, the artist who has a large fan base in Europe has uncollected revenues of 5% and loses out on the pro-rata model another 15%, receiving 27.33% less revenue on 1000 (German, Slovak, Lithuanian, etc.) streams compared to 1000 fully paid streams in the UK.

The winners and losers are not equally distributed among artists. The top artists usually have dedicated fans, and casual listeners who just listen to whatever is on the charts. They would be losing with the user-centric system. Then there are stars of genres, who have dedicated music fans, and less popular artists, who only have dedicated fans – they would be winners.

Our introductory example was very simple and posed a radical difference, because it was considering only 343 streams. The distribution effects on billions of streams largely cancel each other out, but not entirely. We have empirical estimates to the distribution effect between the pro-rata and the user centric model on artist level and on genre level, and on the level of independent versus major label catalogues. The smaller the level, the larger the standard deviation of the distribution effect. On the level of artists, this may run up to 30% and on the level of genres, which represent hundreds of thousands or millions of songs, the effect is around 20%. On the level of major labels versus independents, or on the level of ten million songs, the effect is in the single digits of percentages, often below 1%.

Most assets earn zero in each royalty distribution period, so they are not affected by the model choice. It seems that the pro-rata model significantly favours a smaller number of artists, genres, and labels, and the user-centric model would benefit many artists by a small percentage. All in all, while 10-30% seems like a big difference, 30% of 1 pounds is only 30 cents, and 30% of 0 pounds is still zero. In my opinion,
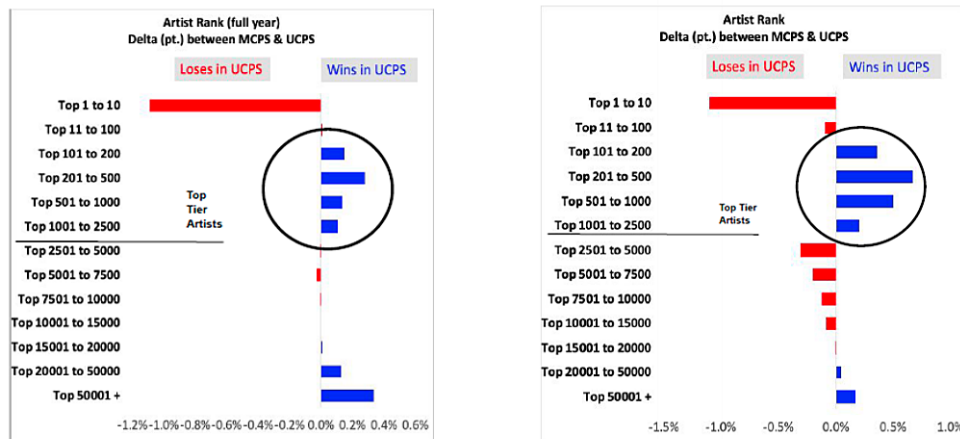
Source: Étude relative à l'impact du passage à l'UCPS par les services de musique en ligne, p14

Figure 2.6: Winners And Loosers With pro-rata and user-centric payouts; distribution effects by artist ranking

the user centric model would benefit medium-sized labels and publishers and smaller countries, but it would have very little tangible benefit for the vast majority of individual artists.

We can infer from the Deloitte study that the majority of the righsholders, and practically all constituent artists of the CI-CEEMID index would be winners of the user-centric system (Deloitte, 2021, p14), as the 'typical' or median song usually comes from artists similar to the median artists, who are in the Top 50001+ range.

Without further data, it is difficult to simulate this effect properly—it is unlikely that the losses and gains are normally distributed among artists, and it is also unlikely that the effect is constant in time. We believe that somebody who loses 15% in a month is likely to be among the losers in most months, but sometimes less, sometimes even more. Because the gains and losses are rooted in listening differences among genres and types of music, we believe that certain artists systematically gain or lose.

Chapter 3

Differences in Music Creators' Earngins: Simulation Results

Our simulations use the CI-CEEMID index as starting points and create hypothetical songs that behave similarly to the 'median' and 'successful' songs. We will expose these hypothetical songs to various genre-, exchange-rate, and other economic scenarios. We will create about 10,000 simulated songs to show why their earnings are so vastly different. Our simulated songs are not actual songs, but they very much resemble a realistic song. The changes in the streaming market environment are either historically recorded changes, or realistically changed simulations (for example, recasting historic exchange rate scenarios.)

1. In subchapter 3.1 we simulate earnings for recordings with different popularities. We create random, but markedly different song audience build-up profiles (quick hits, perennial hits, etc.)

2. In subchapter 3.2 we simulate the effect of different release dates, for example March 2019 and January 2017.

3. Random, but realistic market shares among six countries (the addition of more countries would not give more insights) are presented in subchapter 3.3.

4. The effect of different exchange rates is simulated in subchapter 3.4 for five countries—both using historical exchange rates, and realistic, hypothetical, alternative scenarios.

We show that two identically popular songs released in the same month and with the same number of fans in each country but with different price levels and subject to different exposure to external economic factors can have 30% more or less revenue on the same number of streams. If we consider market share differences between the UK home market and abroad, the differences will be even bigger.

3.1 Different Popularity

We generated realistic streaming volume profiles for smaller and larger bands. We follow them for 60 periods (5 years), but most of them, like real songs, make 95-99% of their revenues in the first 1-3 years.

This means that the average `monthly 100 streams` on a song (6000 over 60 periods) can be much higher, around 1000 in the first month, and very low, as calculated by the CEEMID-CI medium volume index, afterwards. Our scenarios are a bit too optimistic: while we know that most songs tend to have 0 streams per months in the long-tail, we gave them the CEEMID-CI medium volume value (1-9 streams), which is a slight overestimation. In the way in which the CEEMID-CI index is constructed, these values are calculated only from songs that were streamed at all – and many songs that were used to construe the index value in period 2 may have fallen out from the index (in the absence of streams) in period 3. Any song that reached consistently the level of the CEEMID-CI Medium Volume index is far more successful
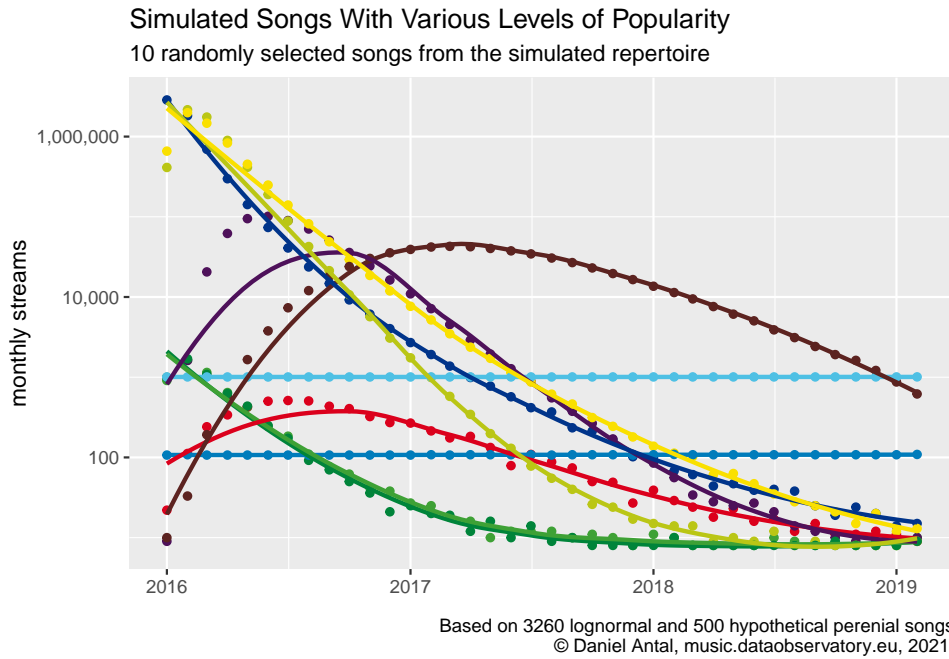
Figure 3.1: Simulated Songs With Various Levels of Popularity

		Earning Level		Earning Differences				
popularity	release period	mean	median	max	min	range	rel. range	stdev
monthly 100	1	20.42	18.39	44.76	0.83	43.94	2.15	12.42
monthly 1000	1	178.72	109.60	476.50	1.77	474.73	2.66	151.79
monthly 10000	1	1819.51	1320.23	4838.16	2.73	4835.43	2.66	1562.45
monthly 100000	1	14217.50	8797.88	48925.19	2.41	48922.78	3.44	16278.45

than the half of the entire observed universe. Yet, we are talking about a penny over a year, so this "optimistic" view does not really alter our conclusions. Rather, it further highlights the fact that most songs do not earn anyway a single penny in a typical month.

So, our popularity level `monthly 100 streams` distributes 6000 streams over 60 periods. Another popularity level of `monthly 10000 streams` distributes 600000 streams over 60 periods in various scenarios usually following the density function. In fact, for discrete time periods, the histogram of left-skewed lognormal pseudo-random numbers repeated 6000-6000000 times.

- Quick hits reach their peak audience (1000, 10000,100000 streams per year) in a territory in 1-3 months.

- Slowly building hits reach their peak audience in a territory in 3-12 months.

- Perennial hits have a stable audience.

Our songs have random streams in each territory following various lognormal distribution density functions, rounded to 0 decimal points. (There are no 1.5 streams, only 1 or 2 streams.)

$revenue_1 = x_1 * p_1$

$revenue_2 = x_2 * p_2$

If we place the annual revenue of the first 1..12 periods into the matrix *X*, then the annual revenue will be

$revenue = X * p$

In our hypothetical scenario, within the different popularity ranges there were 10.2-14.6% *justified differ-*

ences in the first-year earnings in the United Kingdom. These values are lower than the observed 28% variability in streaming values—-the maximum possible difference would have been observed between two unrealistic hypothetical song: one that only had an audience in the best paid, and the other in the worst paid month. Just within the UK, not counting export revenues from foreign countries, a difference in the value of 1000 streams on the level of 10-14% is perfectly normal due to the fluctuation of monthly streaming prices.

3.2 Effect of Release Time

We simulated the effect of different release times with filling up the revenues of our hypothetical songs with 1-25 zero values; they started their 60 simulated periods in period 1, 2, … 25. We used this perfectly realistic simulation to show that new releases in the period 2016-2019 were released in different (and deteriorating) environments. Songs released in 2019 were competing for a far bigger and diverse audience, and with far more songs than songs released in 2016. One of the biggest *justified difference* among righsholders is the release time of their work or recording.
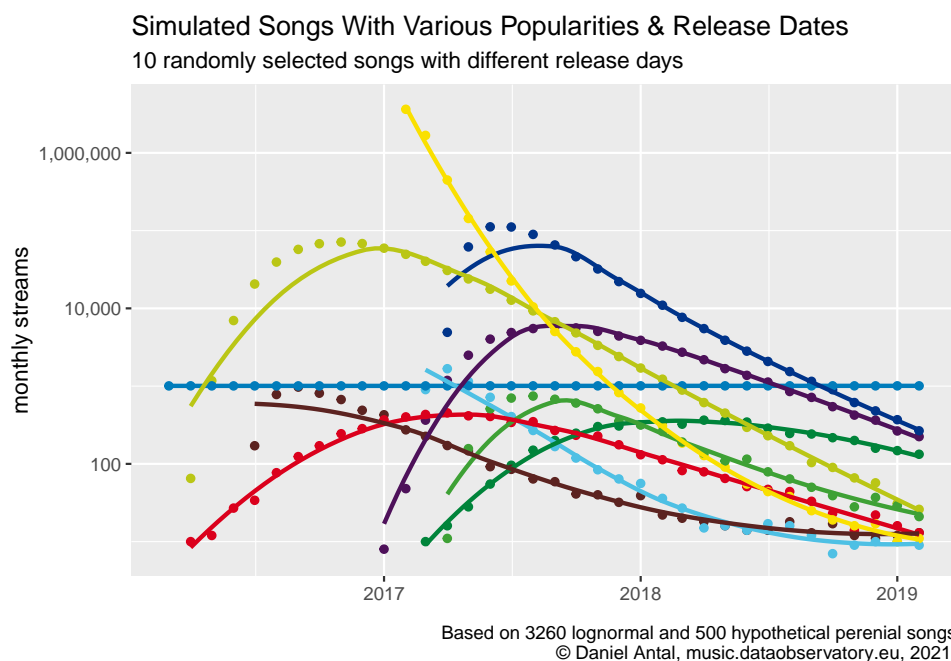

Figure 3.2: Simulated Songs with Various Popularities and Release Dates

As more and more users are joining the streaming markets, and the services are available on newer and newer territories, competition is increasing on the platforms. For example, with licensing to India, potentially hundreds of millions of users emerged who have a distinctly different taste from Northern American or European audiences' tastes. This motivates rightsholders of Indian music to start distributing their new and back catalogue even more in the platforms. The increasing competition leads to a lower likelihood of a song being selected by playlists or users, and because both user tastes and available catalogue is getting more heterogeneous. Furthermore, as all the rich markets are already licensed, new users are entering the system at far lower subscription fees and per stream revenues.

We see a short, transitory period in new markets when only a small number of early users subscribe, typically from the local upper middle class, and music that fits into those local tastes achieves sometimes higher revenues per stream than in the UK. These may be caused by the way licensing agreements are constructed, and initial or minimal payouts are defined. But these opportunities are very short-lived and available for a very limited number of rightsholders.

The 3.3 shows that the price difference over time is relatively small in Germany and the United Kingdom,
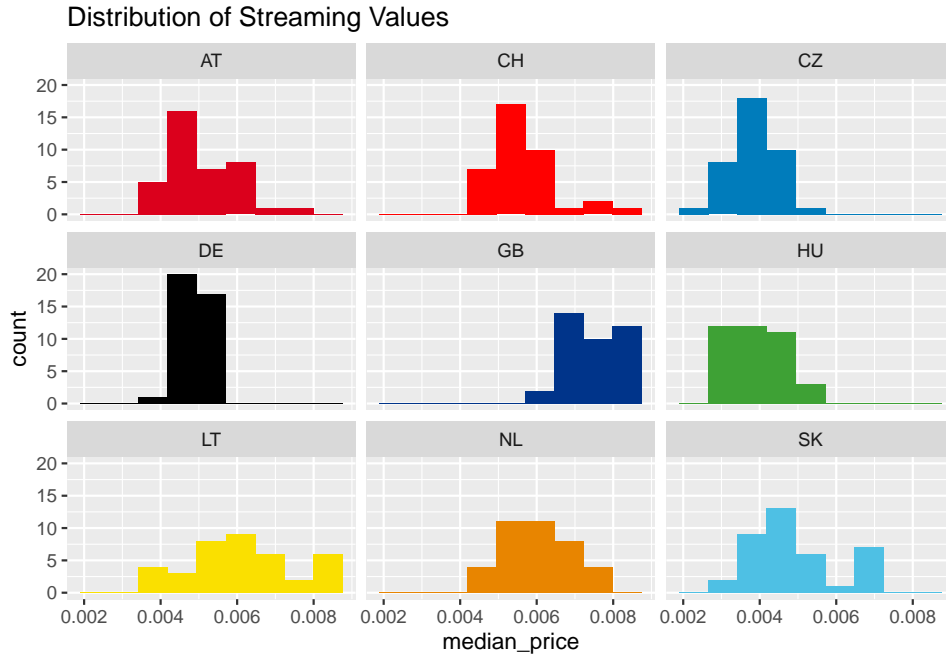
Figure 3.3: Streaming Price Differences in the UK and Select Countries Over Time in the CEEMID-CI Index, Graphic Presentation

but quiet widespread in the emerging market of Lithuania. In this dynamic environment, when a song is released is not a matter of indifference. We will simulate this with placing our hypothetical, randomly generated, but realistic stream counts in different times into the model. A hypothetical song released in December 2017 has a very different market environment in the first critical months than another one in March 2019.

Our model is extended with the release date, which will be one of the 1 … 25 periods of the indexed period. For a song that is released in the first known period the first annual revenue is

$revenue_1 = x_1 * p_1$

…

$revenue_{12} = x_{12} * p_{12}$

and for a song released in the 15th period, the first annual revenue will be the sum of

$revenue_{15} = x_{15} * p_{15}$

…

$revenue_{26} = x_{26} * p_{26}$

Taking only British streaming prices into consideration, the later the release date, the less is the revenue on a similarly popular song.

In the UK, the observed typical prices had a 29% range: this is the 'worst case scenario' difference in the observed period between revenues on the same number of streams. In reality, extreme low and high prices do not last long, and any rightsholder whose song was in use for several months could expect a smaller variability. But regardless of popularity, we can clearly see on the Chart 3.4 that songs released later could expect a lower return on the same number of streams across popularity categories.

The UK in this example is a high-return, low-risk country. We have seen temporarily higher prices in some emerging markets—we believe these are temporary anomalies, related to minimum licensing fees, and lack of competition. (During the observation period, often only Deezer was present in the emerging
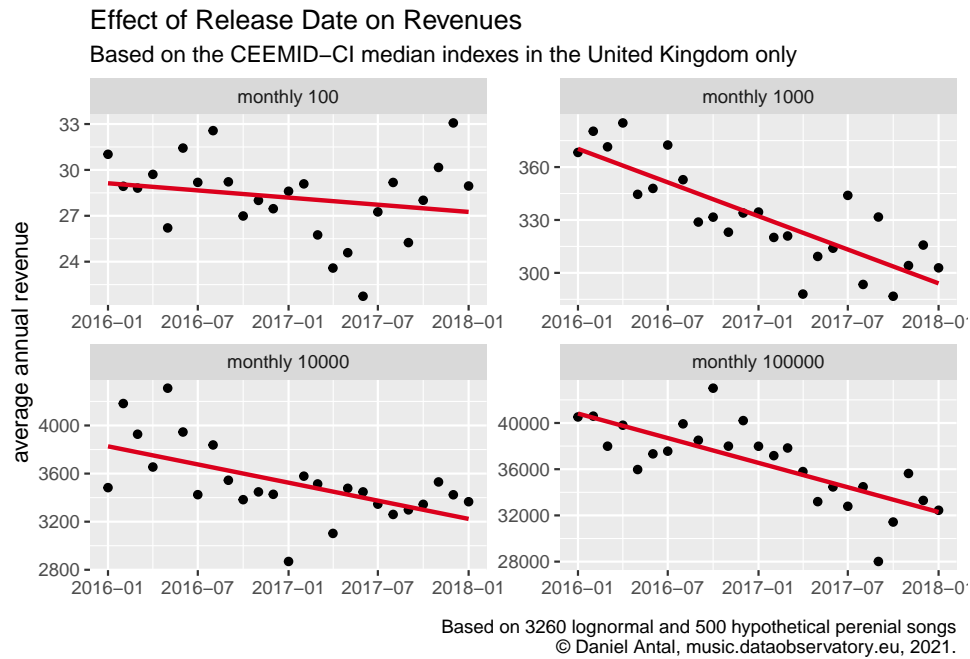
Figure 3.4: Effect of Release Date on Revenues

	Revenue	Risk		
country	mean	sdev	range	absolute deviation %
GB	0.743	0.063	0.217	29.174
HR	0.685	0.206	0.828	120.850
LT	0.612	0.139	0.465	75.965
NL	0.601	0.090	0.352	58.530
CH	0.560	0.085	0.342	61.065
AT	0.509	0.091	0.422	83.069
SK	0.486	0.103	0.357	73.463
DE	0.486	0.035	0.189	38.948
HU	0.390	0.068	0.236	60.517
CZ	0.382	0.063	0.246	64.267
AM	0.345	0.074	0.260	75.161
BG	0.342	0.041	0.163	47.626
RO	0.340	0.116	0.495	145.675
PL	0.266	0.035	0.141	53.052
TR	0.200	0.044	0.172	86.161

country	Revenue	Risk		
	mean	sdev	range	absolute deviation %
HUF	379.615	26.990	111.962	29.494
CZK	31.670	2.964	10.185	32.159
HRK	8.999	0.719	2.793	31.041
CHF	1.350	0.111	0.377	27.911
EUR	1.195	0.086	0.335	28.054
GBP	**1.000**	**0.000**	**0.000**	**0.000**

markets, and was distributed in special packages together with telecom companies.) Among the mature markets, only the German market was less risky than the British one.

3.3   International Competitiveness

Our simulation contains realistic scenarios about earnings in a few advanced and emerging markets besides the UK. While we only simulate results only in a few select countries, we do not believe that a global model would add a lot more insight into our report. The differences in earning potentials among rich markets is not that great, and rightsholders who have a large audience in the United States or other markets can expect similar results. We want to keep our example traceable, and we want to work with a limited number of exchange rates. If we compare the following table with the price differences in 3.2, we will realize that the exchange rate changes in the last 8 years were as great, or even greater as the streaming price change *within the UK*. If a British rightsholder had significant audience in Eastern European emerging markets, the price changes could have been greater than within the UK *only because of the exchange rate risk.* The exchange rate risk against rich, advanced markets in the Eurozone and Switzerland was around the same magnitude as the price variability in the UK.

The exchange rate risk and the foreign streaming price risk, depending on the risk metric used is multiplicative or additive. The absolute price changes in the European markets were about 30-80% in the streaming period, and the exchange rate movements add about another 30%. In the indexed, post-Brexit period the pound was devaluing, while the streaming prices were decreasing. This means that the systematic devaluation of the British pound shielded from about 1/3 of the negative export price changes in Europe. (Systematically, similar effects were felt in the American and Japanese markets, but with different variability.)

Our hypothetical artists have a random market share in the UK, Germany, the Netherlands, Switzerland, Czechia, Hungary (We chose CH, CZ and HU because of their different currencies, the CHF, CZK and HUF.) These are relevant scenarios for the British music industry, which is a global leader. British labels and publishers often sign Dutch artists, for example, who have a larger market in the Netherlands than in the UK. Or, in some of our scenarios, a mainly British rightsholder has a significant fan base in Poland and Hungary. These are all plausible scenarios, but of course, for smaller British labels and artists who do not have a large following abroad, those cases are more interesting where the large majority of the income, say, at least 80% is arriving from the UK.

For the British music industry, the use of the English language is an international competitive advantage in most genres. (Of course, this is not the case with authentic Sami folk or classical music, or instrumental genres.) While a Hungarian-language song will hardly find place on a playlist outside of Hungarian-speakers, who are a very small minority in the United Kingdom, English-language music is dominant on the Hungarian market.

The point we are making here is that artists who are competitive in the rich markets (and the UK is one of them) have better earnings prospects than artists who are focused on emerging or future markets. We simulated this by creating hypothetical divisions of 100% of the plays between six markets: GB, CH, DE, NL, CZ, HU. This way we can show that it does matter where a rightsholder's music is played, as well as the actual exchange rate.
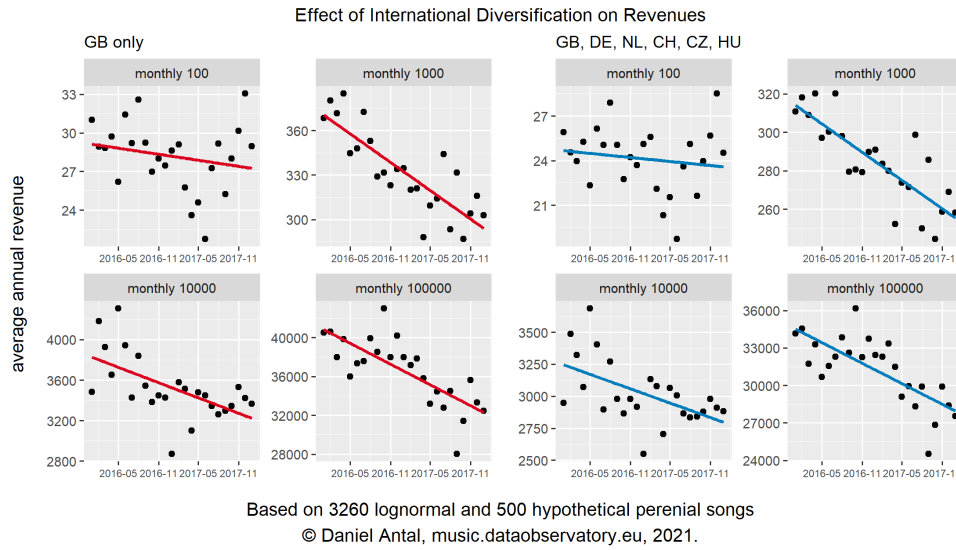
Figure 3.5: Effect of International Diversification on Revenues, Side-by-Side Comparison
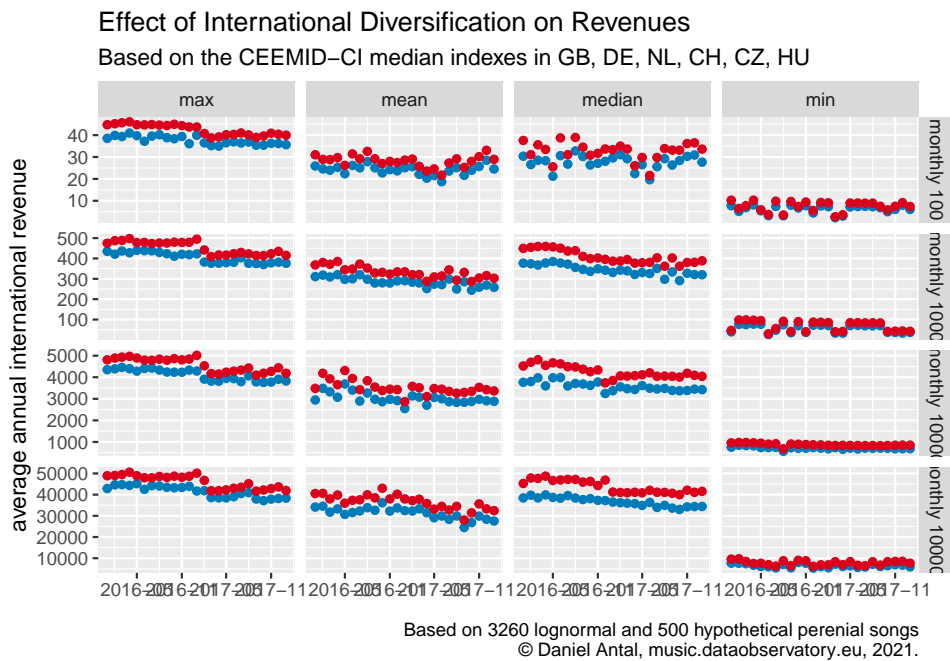

Figure 3.6: Effects of International Diversification on Revenues

The British market is a relatively high value market, and in our simulated portfolios the international diversification (on the same number of streams) always reduced the average or the typical (median) income and brought the best and worst performance down (See Fig. 3.6). If a rightsholder had 1000, 10000 or a million streams, it was better to have it in the UK market than in the other markets.
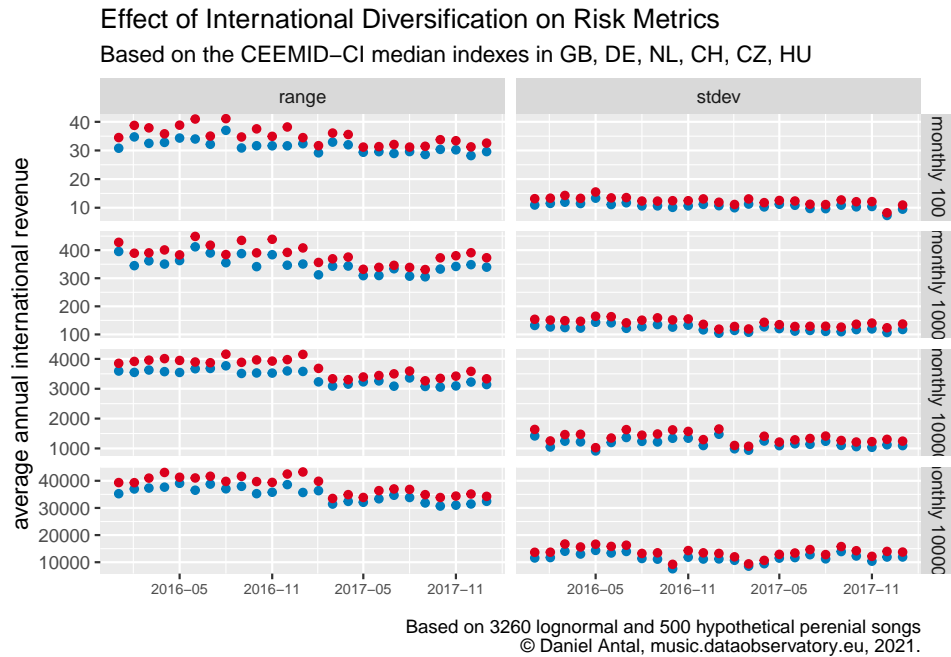

Figure 3.7: Effects of International Diversification on Risk Metrics

The returns on the same number of streams steadily decreased (see Fig. 3.6), because, as we have seen in the 2.3 CEEMID-CI Streaming Price Index, the prices have been decreasing over the period.

However, as expected, international exposure reduced all risk metrics somewhat. Not only did diversification reduce the absolute and standard deviation of a song's revenues, but the foreign markets performed slightly better, and the general decrease on unit price was lower on the same portfolio. Next, we examine the contribution of the devaluation of the GBP in the period against most currencies.

3.4 Exchange Rate Effects

We are going to present earnings in GBP. However, foreign earnings are not made in GBP. When the GBP depreciates, and loses 10% of its value against the CZK, then Czech streams, ceteris paribus, bring in 10% more GBP revenues. The Czech streams are paid from Czech subscriber revenues made in Czech korunas. However, when the GBP appreciates in value against the euro by 10%, British rightsholders immediately lose 10% in GBP terms, because they must buy more expensive pounds as they convert back their German or Dutch revenues.

During our observation period, i.e. the CEEMID-CI Index period, the British pound was systematically depreciating (see Fig. 3.8), but of course, all national currencies had their own movements. The Swiss Frank and the Hungarian Forint had depreciation periods against main currencies, including the British pound; in the latter part of the index period, the GBP was appreciating against the CHF and the HUF.

The depreciation trend helped British artists during this period, and contributed to a great effect on the risk reduction and revenue loss in the portfolios of internationally competitive rightholder assets. The depreciation means that more pounds are given for an EUR or CZK stream revenue. Such trends never last forever – exchange rates do have a strong mean-reversion characteristic. With index values for the years 2020 and 2021, we would certainly see a diminishing value of Hungarian streams, and rather great fluctuation in the value of Swiss streams, for example.
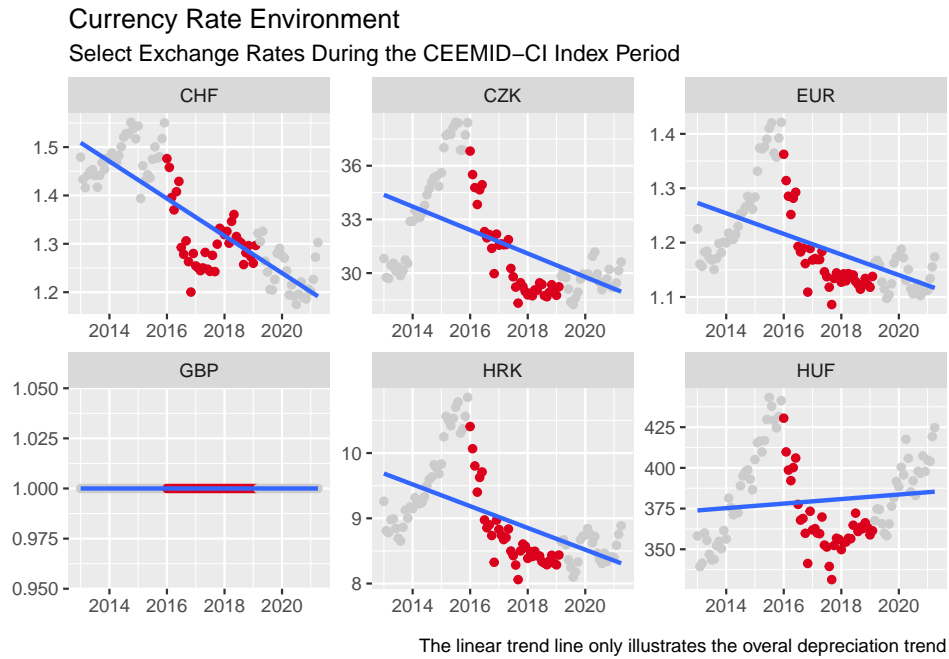
Figure 3.8: Select Exchange Rates During the CEEMID-CI Index Period

We want to show that much of the inequality in GBP earnings is related to exchange rate fluctuation. To stay with this example, when the GBP loses value against the CZK, artists with a Czech fan base win money immediately. When the GBP gains value against the euro, artists with a European (Eurozone-based) fan base lose income in GBP terms, of course, not only in Germany and the Netherlands, but also in further 17 Eurozone countries, and in Montenegro and Kosovo, too, which use the euro unilaterally.

To show the extent of this factor, we created hypothetical but realistic scenarios for our simulated repertoire. Besides using the actual exchange rate used to convert the revenues in the CI-CEEMID index, we used fictitious exchanges rates, too. These rates were simply exchange rates from a varying number of months before or after the actual payout. In this way we created fictitious exchange rate paths that were realistic, because they happened before or after the true payouts. These exchange rate scenarios therefore have a realistic volatility, mean, and median value.

Chapter 4

Conclusions And Recommendations

There is always a hypothetical possibility that organizations with monopolistic power try to corner the market or make the playing field uneven. Because the music industry, in the absence of live performances due to the pandemic, relies critically on streaming revenues, and streaming is managed by large organizations outside of the music industry, any such problems may not even be present within the music industry itself. There are some music royalties in some countries of the world where public bodies, copyright tribunals, and other specialized institutions are taking care of potential payment problems, but music is mainly a lightly regulated business where systematic market problems must be resolved via the application of competition law (Antal et al., 2021).

This is precisely one of the most important ideas behind a permanent music data observatory.

4.1 Ongoing Data Collection and Market Monitoring: The Data Observatory Concept

The music industry requires a permanent market monitoring facility to win fights in competition tribunals, because it is increasingly disputing revenues with the world's biggest data owners. This was precisely the role of the former CEEMID (Artisjus et al., 2014) program, that was initiated by a few collective management societies after a dropped GESAC project. Starting out from three relatively data-poor countries, where data pooling allowed rightsholders to increase revenues, the CEEMID data collection program was extended by 2019 to 12 countries. It was eventually transformed to the Demo Music Observatory in 2020 (Antal, 2021), which is now open for any national rightsholder, stakeholder organization, or music research institute. The idea of this observatory was brought to the UK policy debate on music streaming by the observatory's only (former) British users, via the *Written evidence submitted by The state51 Music Group* to the *Economics of music streaming review* of the DCMS Committee (state51 Music Group, 2020).

> "There are instructive initiatives in other industries in which there is perhaps a clearer and longer standing recognition of the role of economic analysis. This sometimes results in initiatives such as 'Observatories' like the European Market Observatory for Fisheries an Aquatorial Products or the European construction sector observatory [...] These tend to be collaborative endeavours, with a varying mix of government, industry, economists and in some cases funding bodies. [...] To date there have been few if any entities or initiatives for music similar to the above-mentioned observatories. We suggest this is something that policy makers can support and encourage, but which ultimately needs to be driven by the industry itself. [...] This is one reason we have worked with the economist Daniel Antal and his team, in particular on the Central European Music Industry Report 2020. Economists such as Daniel Antal produce data about the music industry that is consistent with international statistical standards and adhere to rigorous data ethics principles, seeking external validation through data and code repositories for underlying data and methodologies."

The data observatory concept is derived from Earth and natural sciences, when often many research stakeholders build large observation stations, such as the Hubble telescope in space, or CERN. Data observatories are managed by a triangular stakeholder base of business, scientific, and policy stakeholders.

The EU has been sponsoring about 60 data observatories, i.e. permanent data collection programs for policy, industrial and scientific research in various domains. These include, for example, the functioning of the milk market, homelessness, the development of alternative fuels, among many others. A similar concept has been embraced by the OECD and some UN bodies, too.

Data observatories are usually managed by a consortium of policy, scientific, and consulting users, or by a public body. In our preliminary research, we have reviewed about 60 observatories sponsored by various EU organizations and added another dozen of data observatories recognized by the Council of Europe, UNESCO, or OECD, and gathered information on defunct observatories that failed to generate a high enough value for money to be continued.

We have identified various problems:

- Almost no data observatory relies on any form of open data, either under the EU Open Data Directive or various definitions of scientific open data. The EU is financing many data gathering initiatives that do not even seem to consider the feel and high quality alternatives of proprietary data.
- The quality of the data is often questionable, plagued with erros in measurement units, currency translation, and wrong labeling.
- The data is often not machine readable, does not conform to tidy data principles, and the use of the data requires significant investment into re-processing in order to be used in statistical or spreadsheet applications.
- Data documentation is not standardized, and often does not meet scientific or other use standards. Metadata is often ad-hoc and does not follow any international or best practices.

The EU Open Data Portal is faced with similar problems. In the case of these open data initiatives, usually metadata quality is high, but the usability of the data is low. The data is processed for the public sector and it is often incomprehensible, lacking documentation that will allow it to be re-processed for new public policy, business or scientific uses.

In the age of big data and AI, the data needs of small and medium enterprises, consulting agencies, universities, and policy-makers is increasing day by day. Yet, only the world's largest companies, Ivy League universities, and rich governments can sustain comprehensive data collection programs. The EU tackles this problem with a financial injection to about 60 data observatories, providing legal access to the re-use of public sector information, and setting up various metadata portals to help accessing these data assets.

Our application aims to significantly improve two serious problems: the EU makes available valuable data in a form that makes its use very difficult for the abovementioned actors; and various EU institutions spend a significant amount of their policy consulting and scientific research budgets to create data programs that do not even consider open data as a source.

4.2 AI Problems

The simplest recommendation systems just follow the charts: for example, they select from well-known current or perennial greatest hits. Such a system may work well for an amateur DJ in a home party or a small local radio that just wants to make sure that the music in its programme will be liked by many people. They re-inforce existing trends and make already popular songs and their creators even more popular.

Spotify's recommendation system (Jacobson et al., 2016) is a mix of content- and collaborative filtering that exploits information about users' past behaviour (e.g. liked, skipped, and re-listened songs), the behaviour of similar users, as well as data collected from the users' social media and other online activities,

or from blogs. Deezer uses a similar system that is boosted by the acquisition of Last.fm – big data created from user comments are used to understand the mood of the songs, for example.

YouTube, which plays an even larger role in music discovery, uses a system comprised of two neural networks: one for candidate generation and one for ranking. The candidate generation deep neural network provides works on the basis of collaborative filtering, while the ranking system is based on content-based filtering and a form of utility ranking that takes into consideration the user's languages, for example. (Covington et al., 2016)

These systems offer a high-level of personalization and usually re-inforce use trends, in turn discriminating users (Werner, 2020; Kraemer and Holden, 2020). Externally validating, or in YouTube's case even understanding how they work would be impossible – YouTube's system uses so many resources and data that replication is impossible outside Google's systems. The deep neural networks are black-box deep learning systems that cannot be fully interpreted by humans.

A commonality across these systems is that they maximize the algorithm creators' corporate key performance indicators. Spotify wants to be 'your playlist to life' and increase the amount of music played during work or sports in the background, during travelling, or active music listening — i.e. maximizing the number of hours spent using it. YouTube and Netflix have similar targets. They are in many ways like commercial radio targets, which want to maximize the time spent listening to the broadcast stream. Radios and YouTube, in particular, have similar goals because they are mainly financed through advertising. For Spotify or Netflix, their key financial motivation is to avoid users' canceling their subscriptions or changing it to different providers, such as Apple or Amazon.

Local content guidelines in public broadcasting, or local content requirements informed by quotas set for commercial broadcasting are similar to utility or knowledge-based recommendation systems. A utility-based recommendation system that targets Slovakness, for example, would prefer from two playlist candidates whichever one has markers of Slovakness in the nationalities of composers, performers, or lyrical content. In this example, a knowledge-based system knows the language or the nationality of a song and creates mixes with a pre-defined Slovak rate.

In a recommendation system many bad outcomes may happen that can eventually lead to lower or no payment for a rightsholder. In these cases, we are not talking about *unjustified* differences of payment as simulated in 3 Differences in Earngins: Simulation Results, but a systematic breach of non-copyright rules that leads to an *unjustified* lower streaming *volume*, and therefore a lower royalty payment on lower volumes:

- It may recommend too few female or small country artists, or start recommending artists with hateful language.
- It may put certain label's music on less visible places.
- It may make the works of major labels easier to find than independent labels.
- It puts less works and recordings on personalized lists than local content guidelines (applied in about 90 territories of the world, see (Stein et al., 2012)) would require from competing local television or radio stations.
- Your personalized list is filled with misleading track/artists names that are hunting for accidental streams much like click-hunting text content in search engines.

These *potential* undesirable outcomes are sometimes illegal, and they may go against non-discrimination or competition law. They may undermine national or EU-level cultural policy goals. They may make Welsh artists earn significantly less than American or English artists.

In our work in Slovakia, we reverse engineered some of these undesirable outcomes. Spotify's and YouTube recommendation system have at least three major parts, which employ machine learning from sources which include metadata:

1. The user's history. Is it the user's history that is sexist, or might the training metadata database be skewed against women?

2. The works characteristics – are Henry Purcell's works as well documented for the algorithm as Taylor Swift's or Drake's?

3. Independent information from the internet. Does the internet write less about women artists?

More often than not, the biggest problem is that the algorithm is learning from data that is historically biased against women or biased for British and American artists, or data that is only discoverable in the English language about relatively popular works that journalists have been covering. Metadata plays an incredibly important role in supporting or undermining the general music education, media policy, copyright policy, or competition rules. If a streaming providers's algorithm does not know the music that music educators or parents find suitable for teenagers, then it will not recommend that music to your children. Parental control algorithms may filter out harmful content for children, but will not play a more proactive role without deliberate involvement of public bodies into documenting children-friendly works.

Until now, in most cases, it was assumed that it is the artists or their representatives' duty to provide high quality metadata; and in mass uses, like in radio or television broadcast, collective management organizations were taking care of the process with appropriate data and IT knowledge. But streaming services usually bypass (with some exceptions) collective management organizations. Small organizations and individual, self-publishing authors do not have an appropriate level of data literacy to provide relevant metadata. What's worse, they cannot scale up the production and validation of metadata. Metadata errors can easily be captured by machine learning algorithms if those algorithms are trained on hundreds of thousands of work/recording records – a precious resource that only large labels, publishers, or collective rights management organizations possess. (Senftleben et al., 2021)

In some cases, it is well understood that some public investment is needed to maintain metadata records for each new technological innovation. While the works of Henry Purcell are eternal, played and re-recorded centuries after his death, neither he nor his heirs receive copyright royalties. If we want to correctly document his works for newer and newer AI-driven technologies, somebody must do this investment. Investment into documentation of cultural heritage, for example, by specialist public libraries is not a novel idea. However, extending this public service to the long-tail of living artists, or heirs who still enjoy copyright protection may be necessary if we want to give them an equal chance on streaming services. As explored in this report, often works earn only pennies over years, which does not even cover the costs of documenting these works. If rightsholders believe that works and recordings in the long-tail should have an equal chance to be re-discovered, heard and paid, then considerable public or philanthropic investment may be needed.

Appendix

Technical Notes

The simulations were made in the R language (R Core Team, 2021) using the *Tidyverse* packages (Wickham et al., 2019) of *dplyr*, *tidyr* and *purrr* (Wickham et al., 2021b; Wickham, 2021; Henry and Wickham, 2020) for data processing and *kableExtra* for the tables (Zhu, 2021).

For visualisation, *ggplot2* and *scales* was used (Wickham et al., 2021a; Wickham and Seidel, 2020) . The document was compiled with *rmarkdown*, *bookdown* and *knitr* (Xie et al., 2018; Xie, 2021a,b).

Acknowledgements

Bibliography

AEPO-ARTIS (2016). Give Performers Their Fair Share!

AEPO-ARTIS (2017). Comments on the Proposed Directive on Copyright in the Digital Single Market.

Antal, D. (2020a). Central And Eastern European Music Industry Report 2020.

Antal, D. (2020b). Feasibility study on promoting slovak music in slovakia & abroad.

Antal, D. (2021). Launching our demo music observatory.

Antal, D., Fletcher, A., and Ormosi, P. (2021). Music streaming: Is it a level playing field? *CPI Antitrust Chronicle*.

Artisjus, HDS, SOZA, and Candole Partners (2014). Measuring and Reporting Regional Economic Value Added, National Income and Employment by the Music Industry in a Creative Industries Perspective. Memorandum of Understanding to Create a Regional Music Database to Support Professional National Reporting, Economic Valuation and a Regional Music Study.

Chee, F. Y. (2021). Apple faces EU charges over Spotify complaint - sources.

CISAC (2017). CISAC position paper on the transfer of value.

Covington, P., Adams, J., and Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, page 191–198, New York, NY, USA. Association for Computing Machinery.

Deloitte (2021). Étude relative à l'impact du passage à l'UCPS par les services de musique en ligne.

EUR-Lex (2019). Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. Text with EEA relevance.

Europe Economics & IVIR (2015). Remuneration of authors and performers for the use of their works and the fixations of their performances.

Flignor, P. and Orozco, D. (2006). Intangible asset & intellectual property valuation: A multidisciplinary perspective.

GESAC (2015a). The Copyright Directive. Addressing the transfer of value online that harms authors.

GESAC (2015b). Regulatory environment for platforms, online intermediaries, data and cloud computing and the collaborative economy. English version.

Henry, L. and Wickham, H. (2020). *purrr: Functional Programming Tools*. R package version 0.3.4.

Holbrook, M. B. and Schindler, R. M. (1989). Some Exploratory Findings on the Development of Musical Tastes. *jconsrese Journal of Consumer Research*, 16(1):119–124.

Holbrook M.B and Schindler R.M (2013). Commentary on "Is there a peak in popular music preference at a certain song-specific age? A replication of Holbrook & Schindler's 1989 study". *Musicae Scientiae Musicae Scientiae*, 17(3):305–308.

IFRS (2011). IFRS 13 Fair Value Measurement.

ILO (1951). C100 - Equal Remuneration Convention, 1951 (No. 100).

Jacobson, K., Murali, V., Newett, E., Whitman, B., and Yon, R. (2016). Music personalization at spotify. In *Proceedings of the 10th ACM Conference on Recommender Systems*, RecSys '16, page 373, New York, NY, USA. Association for Computing Machinery.

Kraemer, D. and Holden, S. (2020). Spotify, apple music, deezer and YouTube found recently hosting racist music.

Moore, F. (2016). The value gap - the missing beat at the heart of our industry.

Ong, T. (2019). Our musical tastes peak as teens, says study.

Pedersen, R. (2020). A meta study of user-centric distribution for music streaming.

Puca, A. and Zyla, M. L. (2019). The intangible valuation renaissance: Five methods.

PwC (2008). Valuing the use of recorded music.

R Core Team (2021). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Senftleben, M., Margoni, T., Antal, D., Bodó, B., Gompel, S. v., Handke, C., Kretschmer, M., Poort, J., Quintais, J., and Schwemer, S. F. (2021). Ensuring the visibility and accessibility of european creative content on the world market: The need for copyright data improvement in the light of new technologies. *SSRN.*

state51 Music Group (2020). Written evidence submitted by The state51 Music Group. economics of music streaming review. response to call for evidence.

Stein, S., Brock, S., and Inc., C. G. (2012). On quotas as they are found in broadcasting music.

Stephens-Davidowitz, S. (2018). The songs that bind. *The New York Times Sunday Review.*

Werner, A. (2020). Organizing music, organizing gender: algorithmic culture and spotify recommendations. *Popular Communication*, 18(1):78–90.

Wickham, H. (2021). *tidyr: Tidy Messy Data.* R package version 1.1.3.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., and Dunnington, D. (2021a). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.* R package version 3.3.5.

Wickham, H., François, R., Henry, L., and Müller, K. (2021b). *dplyr: A Grammar of Data Manipulation.* R package version 1.0.7.

Wickham, H. and Seidel, D. (2020). *scales: Scale Functions for Visualization.* R package version 1.1.1.

WIPO (1996a). WIPO Copyright Treaty.

WIPO (1996b). WIPO Performances and Phonograms Treaty.

Xie, Y. (2021a). *bookdown: Authoring Books and Technical Documents with R Markdown.* R package version 0.24.

Xie, Y. (2021b). *knitr: A General-Purpose Package for Dynamic Report Generation in R.* R package version 1.34.

Xie, Y., Allaire, J., and Grolemund, G. (2018). *R Markdown: The Definitive Guide.* Chapman and Hall/CRC, Boca Raton, Florida. ISBN 9781138359338.

Zhu, H. (2021). *kableExtra: Construct Complex Table with kable and Pipe Syntax.* R package version 1.3.4.