Increasing returns to information in digital music downloads

Néstor DUCH-BROWN and Bertin MARTENS
IPTS, JRC, EC

DRAFT - Preliminary and incomplete
(This version 11.10.2013)

Abstract: This paper studies the distributional dynamics of digital music downloads using data for 18 countries during the period 2006-2011. The main objective is to test if the hypothesis of increasing returns to information also holds in digital markets, since it is well documented for traditional (physical) cultural markets. We find that increasing returns to information are present in almost every country analysed and in every year of the sample used. The results lead to the conclusion that increasing returns to information is a strong and general feature of demand for digital music. However, the value of the coefficient significantly decreases towards zero when a temporal analysis is carried out. Autocorrelated growth in digital downloads is becoming less important in a period when music is mainly distributed on-line. To explain this trend, we relate the degree of concavity of the distribution with information about the share of the top 100 songs and that of the long tail. We find evidence of coexistence of the superstar and long-tail effects. In such a context, both large and small ranked products outgrow middle-rank products.

Keywords: Digital music downloads, Pareto law, increasing returns to information
JEL-codes: C46; D12; L82

1. Introduction

The cultural and entertainment industries are of considerable economic importance in most countries. In particular, the film, music, and book industries have features that are of interest to economists and policy makers, as evidenced by the growing and diversified literature (Caves, 2000; Ginsburgh and Thorsby, 2005). However, the "digital revolution" is deeply transforming these industries. In the recorded music in particular, over the last decade or so, the digitisation of content along with the development of the Internet have had major impacts on the industry, both in the supply and in the demand side.

Until very recently, the literature relating digitisation and the music industry has concentrated on the connection between digitally-enhanced piracy and the observed decline in sales during most of the last decade[1]. This relationship has attracted a lot of research[2], and very few papers have in turn analysed other issues of the music industry that are equally important as to understand the profound transformation digitisation is bringing to the industry[3].

The literature devoted to the analysis of the digital economy suggests that one of the main effects of digitisation on markets is related to the role of information as a catalyst of demand. In the case of cultural industries, and music in particular, the Internet enhances decentralised promotion (word-of-mouth) in detriment of the centralised promotion through traditional mass media (Godes and Mayzlin, 2004; Chevalier and Mayzlin, 2006). Decentralized promotion is actually at the basis of the Long Tail theory (Anderson, 2006) stating that the digital revolution will benefit less popular artists (niches) rather than superstars[4]. Additionally, the broader literature on the economics of entertainment and cultural industries (cinema, books, music, and theatre, among other activities) has focused on product diversity issues and the superstardom theory that seeks to explain the highly concentrated distribution of sales among artists. Rosen (1981) argues that small differences in talent are amplified, whereas Adler (1985) stresses that imitation among consumers could explain the rise of stars without talent. Hamlen (1991, 1994), Chung and Cox (1994), Strobl and Tucker (2000) and Krueger (2005), among others, have conducted empirical tests of these conjectures for the

[1] Until the end of the 90's, physical piracy (bootleg CDs and private copies on blank CDs) was not considered as a big problem. Since Napster, and the advent of P2P, digital piracy has been blamed, at least from the perspective of sector associations, as the main cause for the decline in music sales from an historical maximum in 1999 and up to 2006, when sales began to recover mostly due to digital sales (IFPI, 2013).

[2] Some authors have found a causal effect of piracy on sales displacement (Liebowitz, 2006a ad 2006b), while others have not (Oberholzer-Gee and Strumpf, 2007). Many more draw more cautious conclusions (Blackburn, 2004; Hong, 2004; Peitz and Waelbroack, 2004; Rob and Waldfogel, 2006; Aguiar and Martens, 2013).

[3] Some of these are Bhattacharjee et al. (2006), Maffioletti and Ramello (2004) and Liebowitz and Watt (2006) who investigate the efficiency of policy measures designed to discourage piracy; Bourreau et al. (2009), analysing the impact of digitisation on the music industry new business models.

[4] Additionally, digitisation challenges traditional business models. In creative industries, intellectual property rights (IPR) represent the legal framework that ensures creators a fair return on the commercial exploitation of their cultural production. However, in the digital era, copyright might be difficult to enforce, and new business models might be called for (Bourreau et al., 2009).

traditional (physical) music industry, but research analysing the digital side of the music industry is scarce (citations!!) [5]. In this paper, we analyse the effects of information transmission on music downloads in the digitisation era and its effects on market demand dynamics, in order to assess if it promotes or dampens the emergence of superstars.

One of the biggest problems encountered in the analysis of the cultural industries is related to understanding the way supply and particularly demand operate in markets of experience goods[6]. What makes the sequential discovery of demand (and adaptive supply) difficult to model and understand is the complicated distributional dynamics they produce. In these industries, audiences make hits or failures, not by revealing pre-existing preferences, but by product discovery mechanisms that remain a black box even for industry specialists[7]. Once they have detected a product they like, consumers make a discovery and spread this information to their social network (family, friends, colleagues, etc.). Reviewers and critics do something similar. Information is transmitted to other consumers and demand develops dynamically over time as audience sequentially discovers and reveals its willingness to consume the product. Hence, perhaps the most interesting issue involved in the music industry is the extent to which the transmission of information affects demand. This phenomenon has been documented in the case of traditional (physical) cultural industries to some extent, with evidence emanating mostly from the film industry.

We explore empirically digital music demand dynamics. These dynamics are important as they may imply a sales (downloads) distribution with remarkable uncertainty in outcomes, a critical observation that has implications for the structure of the industry, for firm's conduct and performance, and for the policy dimension as well. Empirical evidence on traditional cultural industries suggests that the word-of-mouth opinion shared between consumers may produce increasing returns to information. In the case of music, for example, this would indicate that the fact that the top ranking albums (songs) seem to earn the majority share of total sales in any given week is an indication of a convex relationship between ordinal ranking and total sales.

A number of studies have assumed that cultural products' sales/revenues evolve according to a particular statistical process and have then tested for the implied distribution. There appear to be some similarities between the way in which particular music recordings gain popularity,

[5] Other authors have dealt with the issue of music production diversity and examine whether we can observe a trend toward the homogenization of musical production and a negative impact of market concentration on diversity. In particular, see Peterson and Berger (1975, 1996); Burnett (1992); Lopes (1992); Alexander (1994, 1996) and Dowd (2004).

[6] According to Kretschmer et al. (1999), cultural goods fall under the category of credence goods, i.e. goods for which quality is rarely learned, even after consumption. They base their argument on the following test: "*Is there a basis for claiming redress if the quality of a product disappoints after purchase? For sub-standard holidays there might be (experience good), for a bad movie there might be not (credence good)*".

[7] This is commonly known as the "*nobody knows anything*" (Caves, 2000).

and the ways in which this occurs for movies, theatrical performances or DVD's. In each case, for example, word-of-mouth recommendations can play an important role. The hypothesis of increasing returns to information has been tested empirically for the film industry by De Vany and Walls (1996), Walls (1997), Hand (2001), and McKenzie (2008). Maddison (2004) has considered Broadway theatrical productions, while Walls (2010) analysed the market for DVD's. For the music industry, empirical studies such as Hamlen (1991, 1994), Chung and Cox (1994), Strobl and Tucker (2000), Krueger(2005), Giles (2006, 2007), Elliot and Simmons (2011), while not always addressing the hypothesis of increasing returns to information, provide evidence consistent with it in some sense.

The decision to buy a music record is shaped by social influences. The choice of an album from the shelves of a traditional brick and mortar music retailer was often the outcome of a decision process where personal attitudes towards music were combined with a bundle of information from peers' experiences through word-of-mouth on the one hand, and public sources like reviews, media or specialised magazines, on the other. Social interaction and information transmission were, therefore, powerful forces shaping the economics of cultural industries in general. Has this changed now that the main form of retailing for cultural/entertainment goods has switched to computer-mediated transactions (e-commerce)?

Advances in information and communication technologies have magnified the power of crowds (Chen, 2008). The emergence of the Internet has enabled consumers to form technology-mediated communities through which they can exchange opinions and experiences regarding companies, products, services, and almost every topic easier and much faster than traditional word-of-mouth. Additionally, the emerging online economy provides consumers with easy access to numerous choices. Unlike traditional face-to-face retail environments, in which products can be seen and touched and customers can consult salespersons, transactions occurring in a computer-mediated communication environment provide no opportunities for experiencing a product or for face-to-face consultation before making a purchase. Hence, consumers' decision making in electronic markets is a more complex process. When information is scarce, people often infer information from the actions of others. This tendency results in herd behaviour, a situation in which everyone is doing what everyone else is doing (Banerjee, 1992). For example, consumers frequently select popular brands because they believe that popularity indicates better quality[8]. Such imitative behaviour can lead to the formation of informational cascades, which occur when individuals follow previous behaviour of others disregarding their own information (Bikhchandani, et al., 1992). Our results show that digital music downloads are also subject to a positive information feedback among audiences that is captured by a Bayesian demand process[9] that

[8] In a famous and seminal contribution, Becker (1991) showed that when two restaurants exist beside one another, customers often pick the one with more seats occupied.

[9] De Vany and Walls (1996) showed that this stochastic process can be explained by Bose-Einstein distributional dynamics.

can produce superstars (Rosen, 1981), path dependence (Arthur, 1989), information cascades (Bikhchandani et al., 1992) or herd behaviour (Banerjee, 1992)[10].

The paper is organised as follows. In the next section, we describe the data and the methodology used in order to detect the presence of increasing returns to information in digital music downloads. Section 3 presents and discusses the results of the analysis. Finally, section 4 concludes.

2. Data, empirical methodology and results

2.1 Data

We use yearly data on digital music downloads for 18 countries for the period 2006-2011. Data comes from Nielsen and includes all songs downloaded from major online retailers such as iTunes, Amazon.com and other music e-retailers in each market (country) during the year of reference. The database contains 6,752,059 different songs from 1,375,892 different artists for a total of more than 92 million observations that account for 7.6 billion downloads in the period under study.

A departure from previous studies concentrating in a single market (country), with this data we are able to analyse simultaneously different countries, among them the most important markets for recorded music and leading countries in internet penetration and e-commerce as well. Moreover, we have data to characterise the whole distribution, not only the top 50, top 100 or some partial upper-tail part of the distribution, as in previous studies. Additionally, we can distinguish the distribution of downloads at the song level and at the artist level. As the data shows (tables 1 and 2), the long-tail is extremely long: an important share of both songs and artists has only one download in the entire period of six years. Overall, 17% of total songs in the database have been downloaded only once, whereas the share of artists with only one download is 10%. Since not all songs and artists are downloaded in every country, the distribution of shares differs by countries. For instance, in Poland, 75% of songs and 44% of artists registered only one download while in the US the shares were only 27% in the case of songs and 12% for artists. Another feature of the heavy-tailed distribution of downloads resides in the difference between the average downloads by song or artists with their maximum values and in the share of songs/artists with at least 1000 downloads, ranging in the case of songs from a minimum of 0.01 in Poland to 1.6 in the US. The figures for artists

[10] Economists and policy makers have begun to recognise the importance of the actions of other agents in the decision-making process. Herding is the deliberate mimicking of the decisions of other agents. Examples of mimicry range from the choice of restaurant, fashion and financial market participants, to academic research. Herding may conjure negative images of irrational agents sheepishly following the actions of others, but such actions can be rational under asymmetric information and uncertainty.

are similar; again the minimum corresponds to Poland with 0.01 and the maximum to the US with 7.3%.

< Insert tables 1 and 2 around here >

Summary statistics for yearly downloads and yearly ranks are provided in table 3 for songs and table 4 for artists. From both tables the idea of a heavily right skewed distribution with large variance is again apparent. The severity of the skew is evident from the calculated means consistently being significantly greater than the medians. For instance, the mean of the yearly downloads ranges from 92.6 in 2006 to 74.5 in 2011 while the median remains constant at 2. In terms of ranking, it can be observed that the average ranking of a given song is quite far away from the top ranked hits and again, quite above the medians for all the years in the database.

< Insert tables 3 and 4 around here >

Unfortunately, we only observe downloads and we don't have data on revenues. Most studies dealing with demand dynamics in cultural industries use data on revenues, but Elliott and Simmons (2011) in the case of UK music records and Maddison (2004) for Broadway performances use quantities and not revenues[11]. Moreover, we have yearly data, when the appropriate time-unit for music is the week (at least, this is the time unit used for elaborating popularity charts[12]). In any case, the analysis of the distribution of digital music downloads can shed some light to further understand some of the effects of digitisation on the demand for music.

Table 5 reports the top five downloaded songs and artists of the database where it can be seen that *I gotta feeling*, a song released on May 21, 2009[13] by *The Black Eyed Peas*, achieved the highest ranking in terms of cumulative downloads with a total of almost eleven million, representing 0.142% of total downloads. For this particular song, 60% of downloads were done in 2006 while 73% were from US consumers. On the artists side, *Rihanna*[14] was the top ranked with a total close to 60 million cumulative downloads in all the countries of the

[11] Given the pricing schemes prevailing in online music distribution, we expect a strong correlation between sales (downloads) and revenues.
[12] For instance, the record of most weeks at number one (16) was reached by Mariah Carey and Boyz II Men in the mid 90's with the song *One Sweet Day*. The record of most weeks in the top 10 charts was 32 while the record of most total weeks on the top 100 is 76 weeks (almost one year and a half). See http://en.wikipedia.org/wiki/List_of_Billboard_Hot_100_chart_achievements_and_milestones#Most_total_weeks_on_the_Hot_100
[13] Wikipedia, consulted on august 20, 2013 (http://en.wikipedia.org/wiki/I_Gotta_Feeling). This hip hop group with both *Boom boom pow* and *I gotta feeling* hold the chart record for 26 consecutive weeks at number one in the Billboard charts.
[14] Although she started her career in 2004, the big commercial success arrived in 2007 with her album *Good Girl Gone Bad*. See http://en.wikipedia.org/wiki/Rihanna.

database, representing 0.8% of total downloads. In her case, the song *Disturbia* concentrated 9.2% of her total downloads, with the top 5 songs accounting for 41% of downloads. As in the previous case, most downloads came from the US (70%).

< Insert table 5 around here >

These features of an extremely skewed distribution arise primarily because of the extreme events in the database. This would suggest a convex downward relation between songs' ranks and their downloads as can be seen in figure 1, where we have plotted the whole distribution in normal scale (upper row) and in double logs scale as well (bottom row) for songs (left column) and artists (right column). The visual logarithmic transformation of the distribution showed in the bottom part of figure 1 suggests a non-linear relationship between the variables, both for songs and for artists. In a similar vein, figures 2 (for songs) and 3 (for artists) show the heavy skew of the respective distributions by partitioning the data into the top 50, top 500, top 5000 and top 50000 cumulative downloads. We now formally test these distributions.

< Insert figures 1, 2 and 3 around here >

2.2 The Pareto law and deviations

The empirical methodology used in this paper involves a simple test of a scaling phenomenon commonly known as Pareto's law[15]. This law states that the probability of occurrence of an event starts high and peters out. Thus, a few events occur very often while many others occur rarely. Paretian behaviour is generally the result of a stochastic process in which all possible outcomes are a priori equally likely. In the context of the digital music industry, songs (or records) are naturally ranked by their number of downloads (sales). This rank-size relation is extremely useful as a graphical device as it accentuates the tails of the empirical distribution. In particular, the appearance of a linear relationship in a double log scale can be interpreted as the sign of a power law distribution, which corresponds to the asymptotic limit of the Pareto distribution. It follows that in a power law distribution the tails fall to the estimated power, which leads to much heavier tails than other models commonly used in describing demand attributes in monopolistically competitive markets[16]. In a nutshell, if the digital downloads data are power law distributed, the average values of downloads (and hence profits) are dominated by few megahits (in the case of songs) or superstars (in the case of artists).

[15] Pareto's law (1897) was originally used in economics as a description of income distributions but the scaling phenomena has also been observed in many other areas in natural and social sciences.
[16] For instance the Gaussian or the exponential distributions.

The Pareto law[17] implies the following relationship between downloads and rank:

$$SR^{\beta} = A \tag{1}$$

where A and β are constants, S denotes size(downloads in our case), and R denotes rank. The natural logarithmic transformation of (1) yields the following equation

$$ln\, S = ln\, A - \beta\, ln\, R \tag{2}$$

So on a log-log scale, β measures the slope of a straight line. Simon (1955) showed that the Pareto law can be derived given two assumptions: i) Gibrat's law, stating that the growth rate of sales is size independent and; ii) a constant entry of new songs (artists). Hence, the Pareto law has a natural interpretation for the music industry: an increase in sales affects future growth through the information sharing between those individuals who have bought the song and potential listeners, but the effect of the increase will diminish as time goes on due to saturation of the potential audience and the entry of competitors (release of new albums/songs).

Later on, Ijiri and Simon (1974) observed that empirical rank-size distributions frequently deviate from the Pareto law by exhibiting strong concavity. They derived analytically the size distribution when growth rates can be autocorrelated[18], and they found that positively correlated growth rates lead to a downward concave relationship. They suggested that the curvature of the distribution may be quantified by using the following equation:

$$ln\, S = ln\, A - \beta\, ln\, R + \gamma\, (ln\, R)^2 \tag{3}$$

The curvature of the distribution is concave downward if the coefficient γ is negative and convex downward if it is positive. Strong downward concavity of the sales-rank distribution would indicate autocorrelated growth in revenues and this is the predicted effect of increasing returns to information[19]. This hypothesis is consistent with the superstar phenomenon (Rosen, 1981) where small differences in products can become magnified into huge differences in final outcomes, an explanation also consistent with information cascades (Bikhchandani et al., 1992) in which individuals place relatively greater weight on the information provided by a song's previous listeners. If a song enjoys increasing returns to information then its sales would be autocorrelated. A song that has enjoyed recent growth is more likely to grow further

[17] The Pareto law appears to hold in many unconnected areas. For example, Steindl (1965) found it applies to the relationship between firm size and rank in an industry; Zipf (1965)

[18] Vining (1976) has shown that autocorrelation cannot solely account for the departure from the Pareto curve but depends on both autocorrelation and the rate at which past growth is discounted.

[19] In the context of the film industry, De Vany and Walls (1996) show that the increasing returns to information causes some movies to become hits and others to become bombs through the feedback of the information dynamic.

than a song whose growth occurred further in the past. The more people who have listen to the song, the more information there is available to potential consumers. And then the information cascade occurs.

The implication of such a finding is that song downloads may grow in a manner that is related to their relative performance. Finding of $\gamma > 0$ (convex downward) or $\gamma < 0$ (concave downward) may signify negative or positive autocorrelated growth respectively. The previous cited studies have rejected a linear Pareto rank-size relation and found statistically negative estimates of γ. These results have been interpreted as evidence of 'increasing returns to information' based on the model of Ijiri and Simon (1974). That is to say, a song/artist that has enjoyed recent growth is more likely to grow faster than a song/artist whose growth occurred further in the past.

In the empirical ground, introducing a squared term into the equation could lead to problems of multicollinearity, since the ln(rank) and its squared term will be highly correlated. However, in the strictest sense, multicollinearity refers only to linear relationships between variables. In any case, one approach to the potential problem of multicollinearity in models with a polynomial term is to subtract the continuous variables from their means. When this was carried out on our data it produced no difference in the results. In addition, in order to correct for possible heterosckedastic errors, the White procedure to compute standard errors was adopted in every regression. Additionally, we performed some robustness checks with robust estimation procedures that take into account the existence of outliers and also corrects for the presence of heteroskedasticity in the error term. We did not find any substantial modification from the OLS results reported in what follows.

2.3 Results

Equation (3) is estimated using OLS where S is taken as each song/artist digital downloads and R is the rank. For the estimations we proceed sequentially. First, we concentrate in the results at the song level. In this case, we first perform regressions for the whole universe of songs in four dimensions: by year and country; by country with time aggregation of downloads; by year with aggregation by countries; and finally aggregating by year and country. In these regressions we introduce time and country dummies where appropriate (Table 6). Then, we estimate the regressions year by year controlling for country specific effects by means of dummies (Table 7). The objective of these regressions is to analyse the evolution of the parameter of interest (γ)[20]. Finally, we performed country level regressions aggregating controlling for year effects, with the intention to detect differences in information transmission in the different countries. The results are shown in table 8.

[20] We performed alternative regressions with the aggregated data. The results were unaltered, although the estimates of γ were somehow higher, starting at -0.193 in 2006 down to -0.147 in 2011. However, in this case, the R^2 were lower for each than those reported in table 7.

< Tables 6, 7 and 8 around here >

The results for the songs' digital downloads considered simultaneously, reported in Table 6, show a remarkable similarity between the regressions that used individual songs' downloads and the regression that used different forms of cumulative downloads by time (specification 2), country (specification 3) and overall (specification 4). These results suggest that the rank-downloads relationship for songs tends to depart from the linear Pareto distribution in a direction that may suggest autocorrelated growth as defined by Ijiri and Simon (1974) and has been noted by De Vany and Walls (1996). That is to say, the $[\ln(\text{Rank})]^2$ coefficient is observed to be significant and negative.

The results are very similar to those obtained in the previous studies cited. Although our estimations of γ are somehow far from what we observe in the film industry –for instance, estimates in the literature range from -0.34 in Hand (2001) to -0.54 in the study of McKenzie (2008)- they are quite adjusted to the results obtained in the physical music industry by Giles (2007) who obtained an estimate of -0.11 and Elliott and Simmons (2011) who found a parameter 0f -0.11. An important difference between or study and others is the amount of information used. Here, we are using the whole distribution of songs, whereas previous analyses have only analysed a restricted, and normally the upper part, distribution. The implications from these results suggest that digital music downloads are right skewed with a mean that is dominated by big hit songs. The top songs overall (as well as by year or by country) earn a disproportionate large share of total (year, country and overall) downloads and this leads to a downwardly convex relation between rank and downloads.

In order to confirm these results, we performed additional tests of the increasing returns to information hypothesis. In table 7, results coming from year-by-year regressions controlling for country specific fixed-effects are shown. The results derived from the evolution of the increasing returns to information hypothesis show that the value of the estimated γ parameter is decreasing in time. This occurs in a period in which digital music downloads have been growing in importance, replacing traditional physical music sales as the main source for revenue for the industry. How this result is related to digital information cascades or computer-mediated communications remains a topic for future research.

Results by country, with the introduction of year dummies, show important differences in the intensity by which the increasing returns to information are present in the different countries of the database. Here, we find that Poland is the only country showing positive and statistically significant estimates of the γ parameter, a result that would imply decreasing returns to information (songs with more previous downloads are less likely to be more downloaded), somehow reversing the hypothesis[21]. With this exception, the rest of the

[21] This could be due to different reasons. First, data for Poland is restricted to the period 2008-2011. Second, it is the country with the lowest number of observations, which may imply a scale effect in the results.

countries analysed show negative and statistically significant estimates of the γ parameter, although with important differences in its value. For instance, the strongest effect of the increasing returns to information hypothesis is found in the US, the country with the highest number of downloads in the data. In this case, the estimated parameter is -0.128, higher than those obtained in the previous literature and close to our own highest estimates in tables 6 and 7.

In a different exercise, we aggregate the song information by artist and we perform again the regressions, now at the artist level. Tables 9, 10 and 11 replicate the procedure used to test the increasing returns to information hypothesis for songs of tables 6, 7 and 8 but with the data at the artist level. In this case, the γ parameter was estimated significantly negative in all the different regressions suggesting there is still sufficient curvature in the log Pareto model to warrant the inclusion of the quadratic term. This implies that the theoretical model of autocorrelation of downloads is highly valid for the superstar phenomenon and provides stronger empirical support for the increasing returns to information hypothesis.

<Tables 9, 10 and 11 around here >

Although the nature and mechanics of information cascades are different in the physical (face to face and expert recommendations) and in the digital worlds (recommendation systems, blogging), their effects seem to be similar, at least in terms of the propagation of information concerning consumption decisions of experience goods. Even if we are not able to go into the details of the specific effects of information cascades or herd behaviour in analysing the dynamics of demand for digital music downloads, hopefully future research will help to clarify the mechanisms through which computer-mediated information transmission will shape consumers' decision making.

5. Explaining demand dynamics in the digital music markets

In the last section we documented a significant concavity in the size-distribution of digital music downloads for 17 countries. As discussed by Ijiri and Simon (1974), Vining (1976), and De Vany and Walls (1986) among others, these deviations from pure power law behaviour could arise because of many factors. Among the most frequently used in the literature are autocorrelation of the growth rates of the different products or the instability in the entry and exit rate of varieties. However, an important implication of the curvature of the Pareto distribution is the deficiency of large events and the excess of middle-sized events. In other words, middle-ranked products are over represented vis-à-vis large and small rank products.

In the previous section we also documented a decrease in the intensity of the deviation over time (or the degree of concavity). This is consistent with a process by means of which both high-rank and low-rank products grow more than proportionally than middle ranked products, then gradually reducing the curvature and also the observed upward size variance for middle sized events[22]. If highly ranked products increase their share in total downloads, then this is evidence for the superstar effect (Rosen, 1981). On the other hand, if low-ranked products increase their relative importance, then there is evidence in support of the long-tail phenomenon (Anderson, 2006). Hence, in our data we find that the superstar and the long-tail effects coexist and together they produce a re-scaling of the distribution that resembles progressively a pure Power Law (i.e. a straight line in a double log scale).

With few exceptions, the existing literature is divided between the proponents of the long-tail phenomenon and those that advocate the superstar effect. Some researchers find evidence to support the superstar effect because lower consumer search costs for price information are lower for hit products compared to niche products (Ghose and Gu, 2007) or because online recommendation systems are biased towards more popular products since they are based on historical data and hence they may help reduce the sales diversity (Fleder and Hosanagar, 2009). On the other hand, other studies show that lower consumer search costs contribute to the reduced concentration of sales of popular products, resulting in the long tail phenomenon (Clemons et al. 2006, Hervas-Drane 2009, Maryanchyk, 2008, Oestreicher-Singer and Sundararajan 2009).

The long tail indicates a shift of demand from the hits to the niches, but the very popular products can still dominate market demand at the same time. Hence, the long tail and superstar phenomena are not necessarily conflicting and could coexist. For example, Elberse and Oberholzer-Gee (2007) found empirical evidence for the coexistence of the long tail and superstar phenomena by examining overall video sales through both online and offline channels. In the same line of argument, Tucker and Zhang (2007), using an empirical comparison of consumers' click-through behaviour between the catalogue and Internet channels, showed that sales from superstar products are enhanced by attracting new demand without affecting the demand for niche products. Zhou and Duan (2012) also demonstrate that the long-tail and the superstar effects can occur simultaneously. In their case, they suggest that product variety weakens the effect of online reviews and this reinforces the relative importance of extremely-ranked online software downloads.

In order to explain the decay in the deviation from the Pareto Law estimated in the last section, we regress the estimated coefficients for the $(ln\,R)^2$ term in equation (3) –which measures the intensity of the deviation from the power law- on a set of variables that are related to both demand and supply factors. Our main purpose is to assess if the reduction in the upward size variance of middle ranked downloaded songs with respect to both small and large rank products has an effect on this trend. In order to do so, we computed the share of the top 100 songs in each country and year and use that as a proxy for the superstar effect. In the same vein, we computed the share of the songs that have 5 or less downloads in order to account for the long-tail phenomenon. A positive and significant coefficient on these variables would indicate a significant effect on the reduction of the estimated concavity of the

[22] Size variance refers to the ratio of actual size to the theoretical size from the Pareto distribution.

Pareto distribution and hence a transition towards a pure power law. Results are shown in table 12.

< Table 12 around here>

This simple specification is shown in the first column of table 12. We progressively introduce other variables in order to control for other possible explanatory factors and test the robustness of the results. Column (1) of table 12 shows that both the share of the top 100 downloaded songs and the proportion of downloads that come from songs in the long tail (defined as the share of songs with 5 or less downloads) per country and year contribute significantly to the reduction in the deviation from the power law. Hence, superstar and long-tail effects occur simultaneously, with the effect associated with the long-tail greater than the one for the superstar.

In a next step, we follow Zhou and Duan (2012) and incorporate a measure of product variety to check whether an expanded catalogue plays a role in the process. The variable is defined as the number of active[23] songs in the online retailer's catalogue in every country and, as shown in column (2) of table 12, it has also a positive and significant impact. According to these authors, higher variety dilutes the impact of information transmission via recommendation or rating systems since negative recommendations cancel out with positive recommendations more easily. Moreover, they suggest that this weakening effect is more significant on popular products, since the likelihood of being recommended (both positively and negatively) is also greater than for niche products. Hence, the more product choices, the less the consumers depend on individual product information (reviews or recommendations). In such a setting, information feedback becomes pure herding, i.e., uninformed consumers imitate the actions of others that have decided first.

In order to check the consistency of the results, we incorporate several additional variables that could also eventually play a role in the process. First, we average country and year the popularity measure given by Google trends of the main online music retailer (iTunes) as a proxy for usage of recommendation systems. Its estimated coefficient is negative, implying that online recommendation systems fuel information feedbacks that promote more autocorrelated growth, and hence, more concavity. The same happens with the introduction of the variable Facebook, which was also constructed as the average popularity of this social network as given by the data taken from Google trends. The only difference is that its inclusion was justified by the idea of controlling online word-of-mouth through social interactions between consumers and peers, a different process from just reading rations or recommendations online. Finally, we also introduce the cumulative downloads in each country in order to control for potential network effects. Network effects indicate that the value of a product increases with the network size. In the context of music downloads, this variable helps to control the increase in the number of consumers that switch to digital markets. In all specifications, the coefficients associated to the variables that proxy the

[23] This is songs that have been downloaded at least once, per country and year.

superstar and the long-tail effects remain positive and significant, providing some robustness and consistency to the results.

6. Conclusions

This paper has considered the "increasing returns to information" hypothesis using a new dataset on digital music downloads for 18 countries for the period 2006-2011. We find evidence of a departure from Pareto's law in the success of digital downloads in 18 countries for the period 2006-2011 which is consistent with the phenomenon of increasing returns to information. Digital downloads that have enjoyed recent success are more likely to stay at the top of the charts than are hits whose success occurred at an earlier time.

It has been long recognized that individuals make decisions not as rational and atomistic utility maximisers, but are actively influenced by the tastes and decisions of others. The impact of others' actions influencing individual decisions has been studied under a variety of phenomenon labelled as 'herding' (Banerjee, 1992). In an influential paper Bikhchandani et al. (1992) discuss how such phenomenon result from 'informational cascades' where a large number of individuals converge in their behaviour, leading to large-scale popularity in a trend or idea.

Consumers use the evaluations of others as an indicator of product quality while making their decisions in a context of incomplete information or uncertainty about the attributes of the good to be consumed. This situation becomes more obvious in the face of difficult and ambiguous conditions, such as computer-mediated communication environments. The uncertainty of online retail environments can increase consumer reliance on the opinions of others regarding products. The emergence of the Internet has made it important to understand the potential of online herd behaviour or digital information cascades in influencing consumer decisions. Although herd behaviour has long been studied in traditional retail environments (Lascu & Zinkhan, 1999), influences on online herd behaviour are a fairly recent topic of investigation in retail marketing[24] in general, and in digital cultural markets in particular.

Many markets have historically been dominated by a small number of best-selling products. The Pareto Principle, also known as the 80/20 rule, describes this common pattern of sales concentration, and applies particularly well to cultural or media industries, where popularity plays an important role towards success (superstar phenomenon). However, information technology in general and Internet markets in particular have the potential to substantially

[24] Previous studies have investigated herd behaviour in digital auctions (Dholakia, Basuroy, & Soltysinski, 2002; Stafford, Kilburn, & Stern, 2006) and software downloading, and bid numbers and download counts have been used by consumers to indicate quality (Hanson & Putler, 1996).

increase the collective share of not so popular products, thereby creating a longer tail in the distribution of sales.

The property of non-predictability of final demand for cultural goods implies that ex-ante knowledge of consumers' preferences and the intrinsic merit of artistic works may not be sufficient to predict the final market configuration. The complex interaction of sequential discovery of quality through word-of-mouth, reputation effects, advertising and publishers' distributional strategies entail that the selection among many alternative outcomes is driven by the accumulation of many small random sequential events. Among them, the spreading of information on quality by early buyers is predominant. Therefore, social interactions must be taken into account to properly understand the music industry.

It is true that an understanding of informational cascades does not by itself compel any specific set of policy prescriptions. But it does suggest that policymakers need to have a sense of not only of economics but also of the power of social interactions, and their potential role in shaping them.

References

Adler M. (1985), Stardom and Talent, American Economic Review, 75(1): 208-212.

Aguiar, L. and B. Martens (2013), Digital Music Consumption on the Internet: Evidence from Clickstream Data. Institute for Prospective Technological Studies Digital Economy Working Paper 2013/04.

Alexander P.J. (1994), New Technology and Market Structure: Evidence from the Music Recording Industry, Journal of Cultural Economics, 18(1): 113-123.

Alexander P.J. (1996), Entropy and Popular Culture: Product Diversity in the Popular Music Recording Industry, American Sociological Review, 61(1): 171-174.

Anderson C. (2006), The Long Tail, London: Random House.

Arthur, W. (1989) Competing Technologies, Increasing Returns, and Lock-In By Historical Events. *Economic Journal* 99(394), pp. 116-131.

Banjeree, A. (1992), A Simple Model of Herd Behaviour, Quarterly Journal of Economics, 107, 797-817.

Bhattacharjee S., R.D. Gopal, K. Lertwachara and J.R. Marsden (2006), Impact of Legal Threats on Online Music Sharing Activity: An Analysis of Music Industry Legal Actions, Journal of Law and Economics, 49(1): 91-114.

Bickhchandri, S., Hirshleifer, D., and Welch, I. (1992), A Theory of Fads, Fashion, Custom and Cultural Change as Information Cascades, Journal of Political Economy, 100, 992-1026.

Becker, G. (1991), A note on restaurant pricing and other examples of social influences on prices. Journal of Political Economy, 99: 1109-1116.

Blackburn D. (2004), On-line Piracy and Recorded Music Sales, mimeo, Harvard University.

Bourreau M., M. Genesollen and F. Moreau (2009), The digitization of the recorded music industry: impact on Business models and scenarios of evolution. Ref..

Burnett R. (1992), The Implications of Ownership Changes on Concentration and Diversity in the Phonogram Industry, Communication Research, 19: 749-769.

Caves R.E. (2000), Creative Industries: Contracts between Art and Commerce, Harvard University Press, Cambridge.

Chen, Y. (2008), Herd behaviour in purchasing books online. Computers and human behaviour 24: 1977-1992.

Chevalier J. and D. Mayzlin (2006), The Effect of Word of Mouth on Sales: Online Book Reviews, Journal of Marketing Research, 43(3): 345-354.

Chung K.H. and R.A.K. Cox (1994), A Stochastic Model of Superstardom: An Application of the Yule Distribution, Review of Economics and Statistics, 76: 771-775.

De Vany, A. and Walls, D. (1996), Bose-Einstein dynamics and adaptive contracting in the motion picture industry, The Economic Journal, 106, 1493-1514.

Dholakia, U. M., Basuroy, S., & Soltysinski, K. (2002). Auction or agent (or both)? A study of moderators of the herding bias in digital auction. International Journal of Research in Marketing, 19, 115–130.

Dowd T.J. (2004), Concentration and Diversity Revisited: Production Logics and the U.S. Mainstream Recording Market, 1940 to 1990, Social Forces, 82: 1411-1455.

Elliott, C. and R. Simmons (2011), Factors determining UK album success. Applied Economics, 43:30, 4699-4705.

Giles, D. (2006), Superstardom in the US popular music industry revisited. Economics Letters 92: 68-74.

Giles, D. (2007), Increasing returns to information in the US popular music industry. Applied economic letters 14(5-6): 327-331.

Ginsburgh, V. and D. Thorsby (2005), Handbook on the Economics of Art and Culture, Elsevier, Amsterdam.

Godes D. and D. Mayzlin (2004), Using Online Conversations to Measure Word-of-Mouth Communication, Marketing Science, 23(4): 545-560.

Hamlen W. (1991), Superstardom in Popular Music: Empirical Evidence, Review of Economics and Statistics, 73(4): 729-733.

Hamlen W.A. (1994), Variety and Superstardom in Popular Music, Economic Inquiry, 32: 395-406.

Hand, C. (2001), Increasing Returns to Information: Further Evidence from the UK Film Market, Applied Economics Letters, 8, 419-421.

Hanson, W. A., & Putler, D. S. (1996). Hits and misses: Herd behavior and online product popularity. Marketing Letters, 7, 297–305.

Hong S. (2004), The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster, mimeo, Stanford University.

IFPI (2013) Recording Industry n numbers. The recorded music market in 2012. London.

Ijiri, Y. and Simon, H. (1974), Interpretations of Departures from the Pareto Curve Firm-Size Distributions, Journal of Political Economy, 82(2), 315-332.

Kretschemer, M., G. M. Kilimis and C. J. Choi (1999), Increasing returns and social contagion in cultural industries. British Journal of management 10: 61-72.

Krueger A.B. (2005), The Economics of Real Superstars: The Market for Concerts in the Material World, Journal of Labor Economics, 23(1): 1-30.

Lascu, D.-N., & Zinkhan, G. (1999). Consumer conformity: Review and applications for marketing theory and practice. Journal of Marketing Theory and Practice, 7, 1–12.

Liebowitz S. and R. Watt (2006), How Best to Ensure Remuneration for Creators in the Market for Music? Copyright and Its Alternatives, Journal of Economic Surveys, 20(4): 513-45.

Liebowitz S. (2006a), File-Sharing: Creative Destruction or just Plain Destruction?, Journal of Law and Economics, 49(1): 1-28.

Liebowitz S. (2006b), Testing File-Sharing's Impact by Examining Record Sales in Cities, Management Science (forthcoming).

Lopes P.D. (1992), Innovation and Diversity in the Popular Music Industry, 1969 to 1990, American Sociological Review, 57(1): 56-71.

McKenzie, J. (2008), Bayesian Information transmission and stable distributions: motion picture revenues and the Australian box office. Economic Record 84(266): 338-353.

Maddison, D. (2004), Increasing Returns to Information and the Survival of Broadway Theatre Productions, Applied Economic Letters, 11 (10), 639-644.

Maffioletti A. and G.B. Ramello (2004), Should We Put Them in Jail? Copyright Infringement, Penalties and Consumer Behaviour: Insights from Experimental Data, Review of Economic Research on Copyright Issues, 1(2): 75–89.

Oberholzer-Gee F. and K. Strumpf, 2007, The Effect of File Sharing on Record Sales: An Empirical Analysis, Journal of Political Economy, 115(1): 1-42.

Pareto, V. (1897) Cours d'economie politique, Rouge, Paris.

Peitz M. and P. Waelbroeck (2004), The Effect of Internet Piracy on Music Sales: Cross section Evidence, Review of Economic Research on Copyright Issues, 1(2): 71–79.

Peterson R.A. and D.G. Berger (1975), Cycles in Symbol Production: the Case of Popular Music, American Sociological Review, 40(2): 158-173.

Peterson R.A. and D.G. Berger (1996), Measuring Industry Concentration, Diversity, and Innovation in Popular Music, American Sociological Review, 61(1): 175-178.

Rob R. and J. Waldfogel (2006), Piracy on the High C’s: Music Downloading, Sales Displacement and Social Welfare in a Survey of College Students. Journal of Law and Economics 49(1): 29–62.

Rosen S. (1981), The Economics of Superstars, American Economic Review, 71(5): 845-58.

Simon, H. (1955), On a class of skew distribution functions. Biometrica, 52:425-440.

Stafford, M. R., Kilburn, A. J., & Stern, B. B. (2006). The effects of reserve prices on bidding behaviour in online auctions. International Journal of Internet Marketing and Advertising, 3(3), 240–253.

Steindl, J. (1965) Random Processes and the Growth of Firms: A Study of the Pareto Law, Hafner, New York.

Strobl E. and C. Tucker (2000), The Dynamics of Chart Success in the UK Pre-recorded Popular Music Industry, Journal of Cultural Economics, 24(2): 113-134.

Vining Jr., D.R. (1976), Autocorrelated Growth Rates and the Pareto law: A Further Analysis, Journal of Political Economy, 84(2), 369-380.

Walls, D. (1997), Increasing Returns to Information: Evidence from the Hong Kong Movie Market, Applied Economics Letters, 4(5), 187-190.

Walls, D. (2010), Superstars and heavy teils in recorded entertainment: empirical analysis of the market for DVD’s. Journal of cultural economics 34: 261-279.

Zipf, G. (1949), Human Behaviour and the Principle of Least Effort, Hafner, New York.

Table 1. Data description: songs

	Obs.	Downloads by song			Share of songs with	
		Min	Avg.	Max	1 download	More than 1000 downloads
Austria	2,392,627	1	8.1	35,829	57.9	0.06
Belgium	3,272,883	1	10.7	51,435	55.7	0.11
Canada	8,071,002	1	36.3	588,081	43.9	0.48
Switzerland	4,484,358	1	11.1	86,145	51.8	0.11
Germany	8,096,139	1	26.1	391,256	45.5	0.28
Denmark	2,477,841	1	14.3	67,612	55.7	0.16
Spain	2,722,095	1	12.4	173,188	57.2	0.10
Finland	1,183,042	1	4.8	8,929	63.6	0.03
France	6,267,545	1	21.7	230,125	48.0	0.27
United Kingdom	11,600,000	1	53.8	973,060	41.0	0.55
Ireland	1,955,614	1	12.3	47,439	57.4	0.13
Italy	3,708,194	1	13.7	134,664	54.4	0.16
Netherlands	3,059,281	1	7.9	51,029	58.2	0.07
Norway	2,590,004	1	10.8	83,083	55.5	0.11
Poland	121,667	1	2.2	2,184	74.5	0.01
Portugal	892,679	1	4.7	16,577	64.7	0.03
Sweden	2,582,481	1	7.7	26,757	59.2	0.07
United States	27,400,000	1	219.7	5,755,773	26.9	1.57

Table 2. Data description: artists

	Obs.	Downloads by artist			Share of artists with	
		Min	Avg.	Max	1 download	More than 1000 downloads
Austria	266,621	1	73.1	197,099	36.3	0.9
Belgium	318,521	1	110.5	313,290	33.5	1.2
Canada	563,405	1	520.7	3,001,984	26.2	2.4
Switzerland	403,070	1	124.0	526,567	31.1	1.3
Germany	584,912	1	361.0	2,435,097	26.4	2.1
Denmark	257,895	1	137.0	443,971	35.7	1.2
Spain	282,317	1	119.7	618,560	35.1	0.9
Finland	148,315	1	38.4	50,090	41.2	0.6
France	500,941	1	272.0	1,563,828	27.9	1.8
United Kingdom	727,985	1	858.5	8,120,392	22.5	3.0
Ireland	212,406	1	113.6	358,830	37.9	1.2
Italy	350,893	1	144.6	678,279	32.4	1.1
Netherlands	315,524	1	76.7	200,706	34.8	0.9
Norway	263,007	1	106.0	383,867	35.9	1.1
Poland	27,845	1	9.6	9,808	44.1	0.1
Portugal	127,719	1	32.5	46,624	41.4	0.5
Sweden	267,703	1	74.5	190,932	36.2	0.9
United States	1,106,151	1	5447.6	41,830,897	11.9	7.3

Table 3. Song year downloads and year rank summary statistics

Downloads	mean	median	std. dev.	min	max
2006	92.6	2	3724.7	1	1,935,974
2007	89.7	2	4339.4	1	2,713,920
2008	90.8	2	4812.8	1	3,419,836
2009	83.4	2	5145.9	1	4,676,087
2010	74.7	2	5210.1	1	4,346,572
2011	74.5	2	5302.1	1	5,755,773
Ranks	**mean**	**median**	**std. dev.**	**min**	**max**
2006	516,416.9	254,123	595,565	1	2,303,965
2007	746,331.0	381,156	834,210	1	3,322,808
2008	965,636.3	502,300	1,072,277	1	4,273,554
2009	1,186,742.0	644,743	1,310,788	1	5,271,065
2010	1,356,930.0	744,463	1,497,010	1	5,982,655
2011	1,423,570.0	808,081	1,554,785	1	6,271,424

Table 4. Artist year downloads and year rank summary statistics

Downloads	mean	median	std. dev.	min	max
2006	2,330.7	17	42,776.9	1	3,936,364
2007	2,234.0	15	50,383.3	1	8,408,452
2008	2,190.6	14	56,119.8	1	12,853,485
2009	1,938.0	12	61,317.5	1	17,171,609
2010	1,573.3	9	54,546.9	1	14,721,056
2011	1,467.0	8	60,538.2	1	21,332,656
Ranks	**mean**	**median**	**std. dev.**	**min**	**max**
2006	143,299.0	143,165.5	82,658.8	1	271,520.5
2007	225,594.5	226,896.0	130,115.4	1	426,818.0
2008	299,001.5	296,217.0	172,416.5	1	563,221.5
2009	390,307.5	385,988.0	224,975.4	1	729,824.0
2010	480,229.0	488,107.5	276,611.2	1	888,861.5
2011	549,624.0	550,427.5	316,321.1	1	1,007,001.0

Table 5. Overall top 5 songs and artists – cumulative downloads and shares

Song	Downloads	% of total	Artist	Downloads	% of total
I gotta feeling	10,790,889	0.142	Rihanna	59,670,687	0.783
Poker face	9,132,363	0.120	Lady Gaga	47,032,472	0.617
Just dance	7,966,527	0.105	Taylor Swift	45,239,384	0.594
Party rock anthem	7,930,531	0.104	Eminem	40,610,786	0.533
Boom boom pow	7,787,011	0.102	Katy Perry	38,794,463	0.509

Table 6. Estimates of the Song Rank-Downloads relationship

	By year and country	Aggregation		
		By year	By country	By year and country
	(1)	(2)	(3)	(4)
ln(rank)	0.758***	1.620***	1.815***	3.356***
	(0.000940)	(0.00328)	(0.00473)	(0.0159)
$[\ln(\text{rank})]^2$	-0.0960***	-0.138***	-0.146***	-0.202***
	(3.75e-05)	(0.000129)	(0.000178)	(0.000572)
Constant	3.650***	2.267***	4.031***	-3.035***
	(0.00585)	(0.0207)	(0.0312)	(0.110)
Observations	92,919,358	25,899,550	20,274,656	6,752,059
R-squared	0.984	0.991	0.987	0.987

Note: specification (1) includes time and country dummies; specification (2) includes country dummies; specification (3) includes time dummies. Robust standard errors in parentheses. *** denotes significant at the 1%, ** at the 5% and * at the 10%, respectively.

Table 7. Estimates of the Song Rank-Downloads relationship by year

	2006	2007	2008	2009	2010	2011
ln(rank)	1.213***	1.046***	0.984***	0.896***	0.746***	0.706***
	(0.00434)	(0.00334)	(0.00288)	(0.00257)	(0.00223)	(0.00214)
$[\ln(\text{rank})]^2$	-0.124***	-0.113***	-0.108***	-0.101***	-0.0934***	-0.0916***
	(0.000184)	(0.000138)	(0.000116)	(0.000102)	(8.78e-05)	(8.38e-05)
Constant	2.140***	3.184***	3.706***	4.154***	4.909***	5.220***
	(0.0253)	(0.0201)	(0.0177)	(0.0161)	(0.0141)	(0.0135)
Observations	7,213,178	11,234,039	14,425,149	18,146,582	20,241,147	21,659,263
R-squared	0.990	0.992	0.992	0.993	0.994	0.994

Note: all specifications include country dummies. Robust standard errors in parentheses. *** denotes significant at the 1%, ** at the 5% and * at the 10%, respectively.

Table 8. Estimates of the Cumulative Song Rank-Downloads relationship by country

	Austria	Belgium	Canada	Switzerland	Germany	Denmark	Spain	Finland	France
ln(rank)	-0.0342***	-0.103***	0.292***	0.149***	0.376***	-0.209***	-0.243***	-0.129***	0.244***
	(0.00308)	(0.00111)	(0.00190)	(0.00289)	(0.00359)	(0.00140)	(0.00449)	(0.00230)	(0.00253)
$[\ln(\text{rank})]^2$	-0.0581***	-0.0566***	-0.0772***	-0.0646***	-0.0769***	-0.0550***	-0.0491***	-0.0534***	-0.0727***
	(0.000143)	(4.97e-05)	(7.78e-05)	(0.000126)	(0.000148)	(6.46e-05)	(0.000205)	(0.000115)	(0.000106)
Constant	7.850***	9.238***	8.879***	7.679***	7.793***	9.589***	9.100***	7.283***	8.016***
	(0.0165)	(0.00622)	(0.0115)	(0.0165)	(0.0217)	(0.00760)	(0.0244)	(0.0115)	(0.0150)
Observations	2,392,627	3,272,883	8,071,002	4,484,358	8,096,139	2,477,841	2,722,095	1,183,042	6,267,545
R-squared	0.999	0.998	0.996	0.998	0.996	0.999	0.997	0.998	0.997

	United Kingdom	Ireland	Italy	Netherlands	Norway	Poland	Portugal	Sweden	United States
ln(rank)	0.248***	-0.201***	-0.241***	-0.259***	0.0174***	-1.406***	-0.140***	-0.264***	1.557***
	(0.00211)	(0.00108)	(0.000739)	(0.00112)	(0.00338)	(0.0159)	(0.00450)	(0.00161)	(0.00389)
$[\ln(\text{rank})]^2$	-0.0737***	-0.0567***	-0.0505***	-0.0456***	-0.0626***	0.0433***	-0.0531***	-0.0459***	-0.128***
	(8.39e-05)	(5.04e-05)	(3.23e-05)	(5.06e-05)	(0.000155)	(0.000970)	(0.000232)	(7.45e-05)	(0.000144)
Constant	9.728***	9.276***	9.859***	9.383***	7.837***	7.668***	6.770***	9.164***	4.752***
	(0.0132)	(0.00575)	(0.00421)	(0.00619)	(0.0183)	(0.0632)	(0.0217)	(0.00871)	(0.0262)
Observations	11,616,435	1,955,614	3,708,194	3,059,281	2,590,004	121,667	892,679	2,582,481	27,425,471
R-squared	0.994	0.999	0.999	0.999	0.999	0.990	0.999	0.999	0.988

Note: all specifications include time dummies. Robust standard errors in parentheses. *** denotes significant at the 1%, ** at the 5% and * at the 10%, respectively.

Table 9. Estimates of the Artists Rank-Downloads relationship

		Aggregation		
	By year and country	**By year**	**By country**	**By year and country**
	(1)	(2)	(3)	(4)
ln(rank)	0.665***	1.491***	1.641***	3.402***
	(0.00188)	(0.00665)	(0.00913)	(0.0367)
$[\ln(\text{rank})]^2$	-0.124***	-0.157***	-0.167***	-0.238***
	(8.91e-05)	(0.000295)	(0.000387)	(0.00148)
Constant	6.007***	4.851***	6.686***	-0.0619
	(0.00998)	(0.0374)	(0.0536)	(0.225)
Observations	15,919,107	6,725,230	4,176,105	1,375,892
R-squared	0.984	0.988	0.977	0.977

Note: specification (1) includes time and country dummies; specification (2) includes country dummies; specification (3) includes time dummies. Robust standard errors in parentheses. *** denotes significant at the 1%, ** at the 5% and * at the 10%, respectively.

Table 10. Estimates of the Artists Rank-Downloads relationship by year

	2006	2007	2008	2009	2010	2011
ln(rank)	1.110***	1.022***	1.013***	1.025***	0.905***	0.844***
	(0.00901)	(0.00758)	(0.00680)	(0.00658)	(0.00555)	(0.00521)
$[\ln(\text{rank})]^2$	-0.163***	-0.149***	-0.145***	-0.141***	-0.132***	-0.128***
	(0.000467)	(0.000376)	(0.000329)	(0.000310)	(0.000258)	(0.000239)
Constant	5.429***	5.835***	6.081***	6.000***	6.441***	6.754***
	(0.0431)	(0.0380)	(0.0350)	(0.0347)	(0.0297)	(0.0281)
Observations	1,125,264	1,829,813	2,358,876	3,066,853	3,571,274	3,967,027
R-squared	0.990	0.990	0.991	0.991	0.992	0.993

Note: all specifications include country dummies. Robust standard errors in parentheses. *** denotes significant at the 1%, ** at the 5% and * at the 10%, respectively.

Table 11. Estimates of the Cumulative Artists Rank-Downloads relationship by country

	Austria	Belgium	Canada	Switzerland	Germany	Denmark	Spain	Finland	France
ln(rank)	0.215***	0.290***	0.315***	0.465***	0.576***	-0.00660*	-0.0804***	0.123***	0.398***
	(0.00536)	(0.00585)	(0.00430)	(0.00690)	(0.00731)	(0.00372)	(0.00819)	(0.00544)	(0.00573)
$[\ln(\text{rank})]^2$	-0.0972***	-0.103***	-0.108***	-0.109***	-0.117***	-0.0910***	-0.0811***	-0.0951***	-0.108***
	(0.000279)	(0.000297)	(0.000202)	(0.000341)	(0.000345)	(0.000193)	(0.000422)	(0.000304)	(0.000274)
Constant	7.989***	8.541***	9.971***	7.681***	8.384***	9.604***	9.363***	7.488***	8.598***
	(0.0258)	(0.0289)	(0.0230)	(0.0349)	(0.0386)	(0.0181)	(0.0396)	(0.0245)	(0.0299)
Observations	562,197	691,495	1,352,558	904,630	1,420,553	550,285	597,926	287,470	1,163,575
R-squared	0.994	0.993	0.993	0.993	0.991	0.996	0.993	0.994	0.993

	United Kingdom	Ireland	Italy	Netherlands	Norway	Poland	Portugal	Sweden	United States
ln(rank)	0.650***	0.0157***	0.0494***	0.238***	0.117***	0.215***	0.0988***	0.166***	1.260***
	(0.00723)	(0.00326)	(0.00516)	(0.00672)	(0.00467)	(0.0437)	(0.00669)	(0.00619)	(0.00787)
$[\ln(\text{rank})]^2$	-0.123***	-0.0947***	-0.0884***	-0.0964***	-0.0951***	-0.103***	-0.0922***	-0.0945***	-0.154***
	(0.000331)	(0.000172)	(0.000258)	(0.000343)	(0.000242)	(0.00298)	(0.000384)	(0.000322)	(0.000342)
Constant	9.129***	9.271***	9.306***	8.179***	8.640***	4.037***	7.070***	8.417***	8.854***
	(0.0393)	(0.0157)	(0.0258)	(0.0330)	(0.0224)	(0.159)	(0.0289)	(0.0298)	(0.0452)
Observations	1,856,971	443,933	753,784	666,489	563,726	37,314	241,165	563,455	3,261,581
R-squared	0.989	0.996	0.993	0.993	0.996	0.981	0.995	0.994	0.981

Note: all specifications include time dummies. Robust standard errors in parentheses. *** denotes significant at the 1%, ** at the 5% and * at the 10%, respectively.

Table 12. Estimates of the determinants of the decay in the Pareto curvature

	(1)	(2)	(3)	(4)	(5)
Top 100	0.106***	0.0820***	0.0941***	0.0914***	0.0868***
	(0.0162)	(0.0172)	(0.0168)	(0.0171)	(0.0168)
Long tail	0.209***	0.161***	0.182***	0.176***	0.175***
	(0.0486)	(0.0443)	(0.0414)	(0.0425)	(0.0421)
Variety		0.000838***	0.000995***	0.00104***	0.000965***
		(0.000294)	(0.000293)	(0.000265)	(0.000257)
iTunes			-0.000407**	-0.000482***	-0.000594***
			(0.000168)	(0.000172)	(0.000156)
Facebook				-6.08e-05*	-6.86e-05**
				(3.47e-05)	(3.28e-05)
Cum.downloads					0.00334**
					(0.00166)
Constant	-0.121***	-0.111***	-0.110***	-0.107***	-0.156***
	(0.00888)	(0.00814)	(0.00792)	(0.00840)	(0.0265)
Observations	102	102	102	102	102
R-squared	0.981	0.984	0.985	0.986	0.987

Note: all specifications include time and country dummies. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Figure 1. Distribution of songs and artists: the long-tail and Pareto's Law

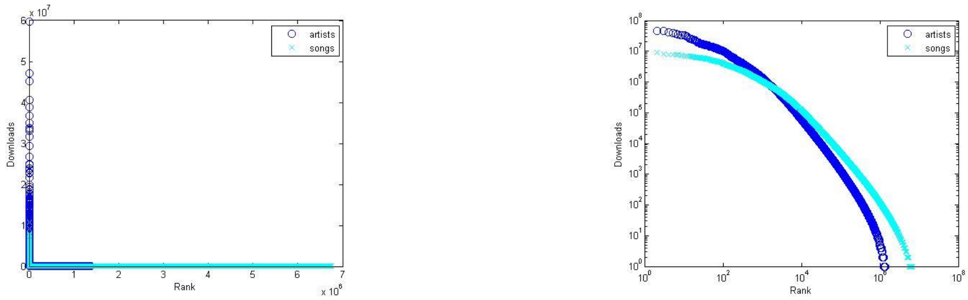
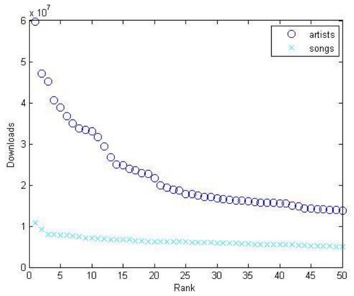
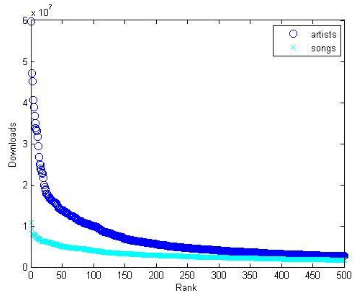


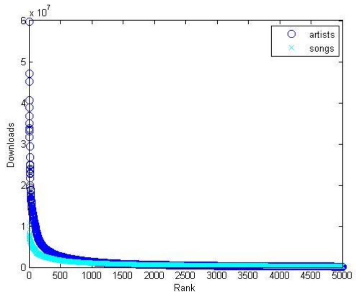
Figure 2. Distribution of songs and artists: cumulative downloads vs. cumulative rank



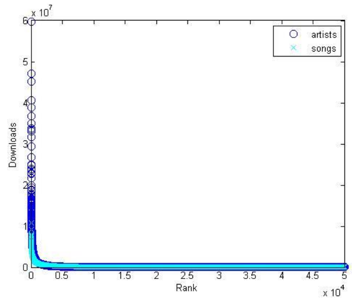
a) Top 50



b) Top 500

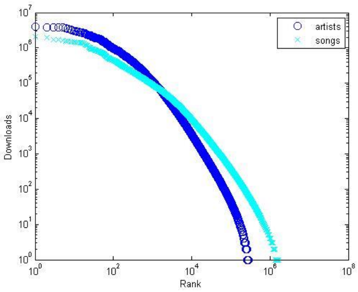


c) Top 5,000

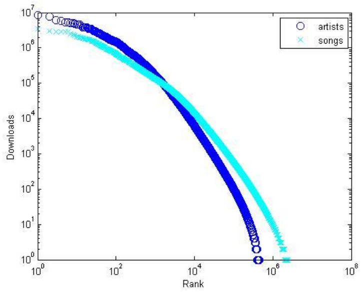


d) Top 50,000

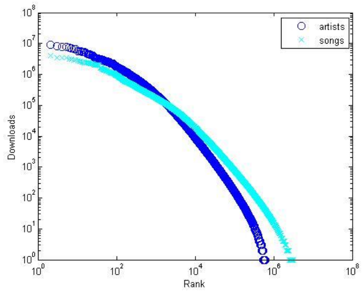
Figure 3. Distribution of cumulative songs and artists: downloads vs. rank by year



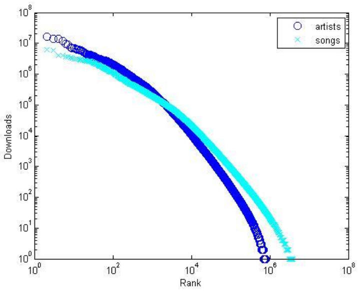
a) 2006



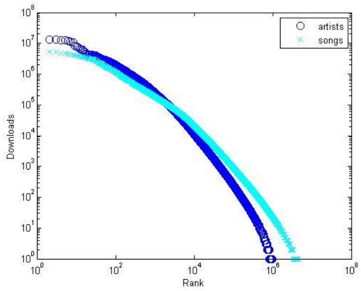
b) 2007



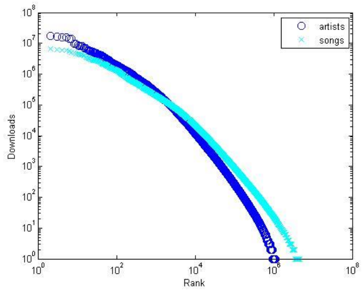
c) 2008



d) 2009

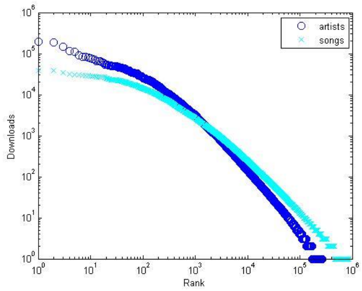


e) 2010

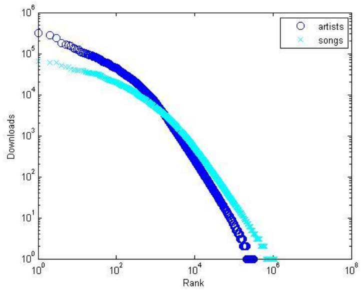


f) 2011

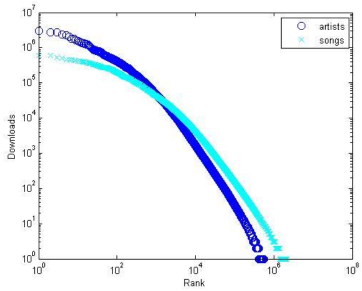
Figure 4. Distribution of cumulative songs and artists: downloads vs. rank by country



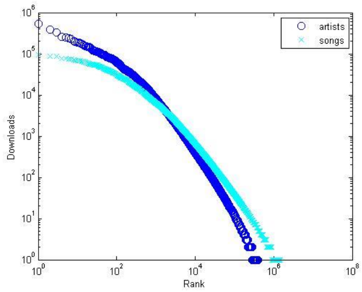
a) Austria



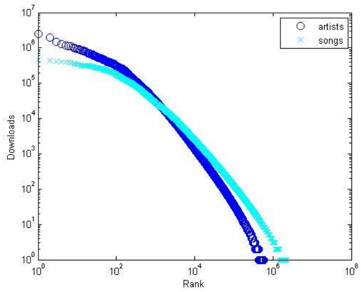
b) Belgium



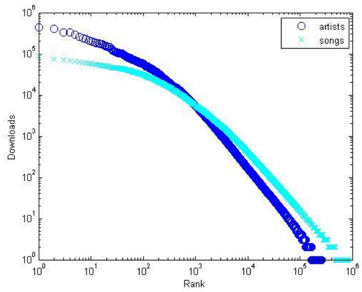
c) Canada



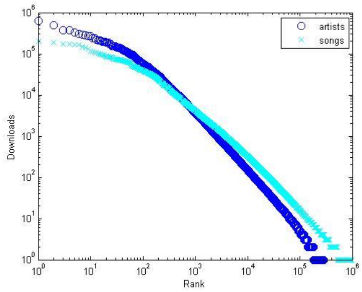
d) Switzerland



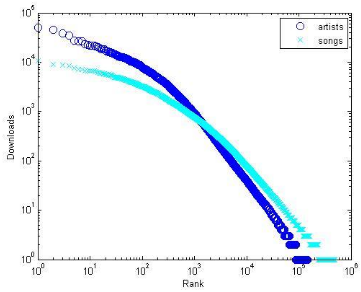
e) Germany



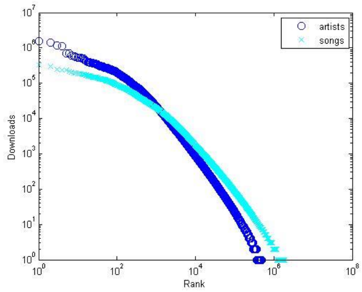
f) Denmark



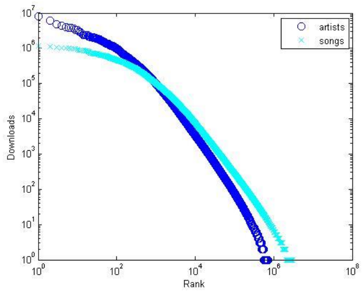
g) Spain



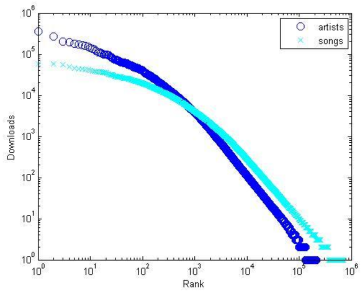
h) Finland



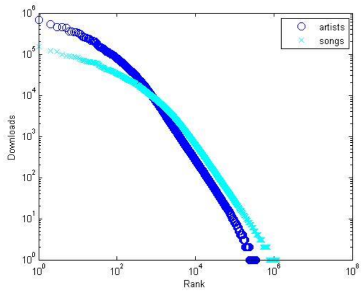
i) France



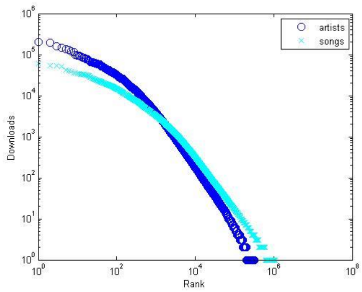
j) United Kingdom



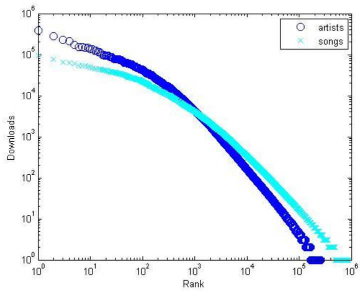
k) Ireland



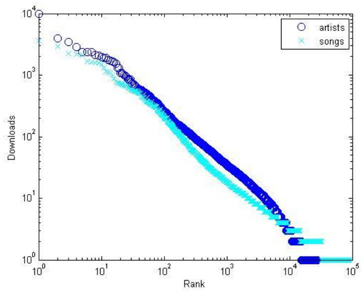
l) Italy



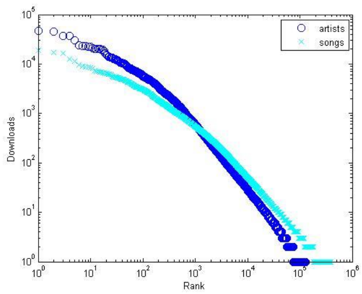
m) Netherlands



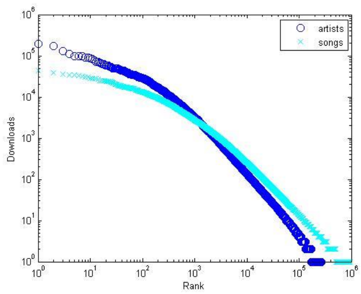
n) Norway



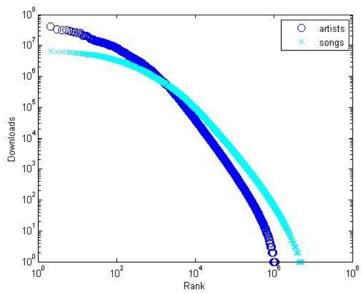
o) Poland



p) Portugal



q) Sweden



r) United States