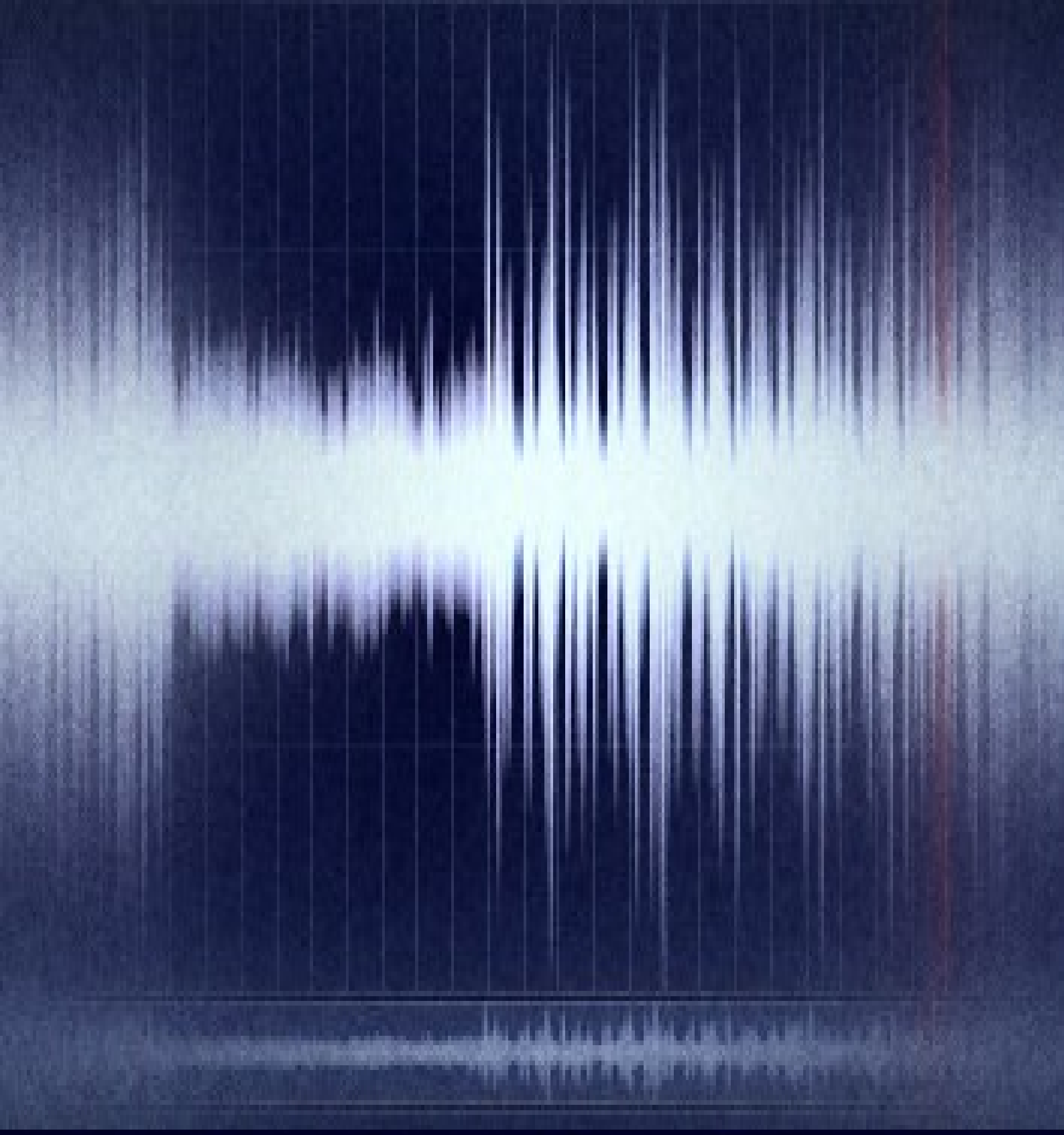


Off the Beaten Tracks

A Study of the "Long Tail" in the Dutch Music Market

José Mora-Jiménez

Off the Beaten Tracks

A Study of the "Long Tail" in the Dutch Music Market

José Mora-Jiménez
325584

Erasmus University Rotterdam
Master Cultural Economics and Cultural Entrepreneurship

Tutor: Christian Handke

June 2010

Abstract

Among the most recent ideas about the future of digital sales for information goods is the "Long Tail Theory" proposed by Chris Anderson in 2004. This theory claims that in a digital world, hits and superstars will cede space to niche products. Therefore, for the future of the music industry and other industries such as books and video, niche products could be more important than selling millions of copies of a hit product. There are relatively few studies concerning the long tail theory. Evidence found still remains contradictory and the theory as a whole remains somewhat untested.

This thesis aimed to explore the long tail hypothesis by studying the Dutch music market. The market was divided into CD's and digital albums. In the market for CD's, when looking at concentration, there is no evidence that the growth in online distribution has had an effect into the overall sales' distribution. In the digital albums' market, the market size does not allow to speak of the long tail as a profitable business model.

Nonetheless, it is argued in this thesis that the long tail should be studied beyond the profitability of the tail. The hypothesis touches many other aspects such as the importance of niche content online and differences in the composition of sales online, aspects that cannot be measured looking only at concentration. If the relative importance of niche content is taken into account, then, the ideas behind the long tail are supported by the data used in this thesis. Different comparisons were drawn between CD's and digital albums, the outcome of these comparisons points towards a better performance of niche content in the digital market.

The mixed results obtained showed that there are still areas to be researched. This thesis provided with evidence for some of these areas, for example the performance of niche content in a digital market and how e-tailers might price niche content in order to profit from the tail.

Index of Contents

1

Introduction

1.1 Description of the topic

A decade ago, buying a music album was in many respects, a different process from what it is today. In order to buy a record, a consumer had to go the nearest record store and choose from the titles that were available at the moment. If the record the consumer was looking for was not available in that store, he/she had only two options: either continue the search in another store or abandon the search altogether. Continuing the search implied additional time and search costs, maybe even travel expenses and there was no guarantee of success in the end.

Today this is no longer necessary, online stores make it possible for consumers in almost any part of the world to make purchases online and have their records shipped right to their addresses, geographical barriers have become blurred in the online market. Today, if the nearest record store does not have the CD I want to buy, I can simply go home and buy it from Amazon or any other online store that has it available. To make the whole process easier, there are now websites dedicated to scan and find products from many different online stores, so that consumers do not have to navigate through different websites, allowing them to compare prices and different products in one site.

Available music formats have also changed. A decade ago, the CD was the standard, the cassette and the LP were still present on the market, but it would not take long before these formats disappeared from the market. Similarly, today digital formats are increasingly taking over the market, once dominated by physical products. With the appearance of digital formats, a consumer can today download an album instantly, without waiting times and without shipping costs. In recent years, buying music can even be done from a mobile phone anywhere at any time.

Without a doubt, during the last decade digitalization of media content and the increasing use of the internet have brought significant changes in the way we search, buy and consume information goods such as music, video and books. These developments may have the potential to change the market structure for these products, the role of intermediaries and the business model the involved companies should use, among other aspects.

In the music industry, online sales have increased significantly and today they account for around 20% of total music sales (IFPI 2009: The record industry in numbers). The book industry is

also starting to venture into digital sales, e-readers are now gaining market and the offer of digital titles is increasing. Moreover, online video is also a growing market, where consumers can rent movies and watch them online from their homes.

Digitalization has yet another consequence, it has increased the number of individual offerings that are available in the market. Since the cost of adding more titles has been significantly reduced for e-tailers, they can afford to carry out more content, as a consequence, a consumer has considerable more options today than he/she had in the past.

The increasing importance of digital distribution has opened a space for theoretization over the direction that music industry will follow, costs are clearly different for suppliers due to new distribution channels and consumers have easier access to an increasing number of products.

Up until recent years, record companies and music retailers relied on a limited number of superstars as their main source of income, therefore all efforts in terms of production and marketing went into finding and launching new artists with potential to conquer the market. However, this business model was meant to serve a market where physical constrains forced suppliers to choose among different products they could make available. This scenario changes when brick-and-mortar stores can be replaced by online stores offering digital products.

Among the most recent ideas about the future of digital sales for information goods is the "Long Tail Theory" proposed by Chris Anderson in 2004. This theory claims that in a digital world, hits and superstars will cede space to niche products, products that have a very specific and therefore reduced market. The internet and digitalization of media content provide the conditions for these products to reach more consumers and for retailers to include more variety into their stock. He claims that for the future of the music industry and other industries such as books and video, niche products could be more important in terms of revenue than selling millions of copies of a hit product.

This theory has several implications for market structure, consumer welfare and the future of the cultural industries, however it remains a debated topic. The digital music market is one of the most fruitful areas for research due to the constant increase in market share for these products. Moreover, this industry has the potential to fulfill more closely all the characteristics described by Anderson in his book.

1.2 Relevance of this study

So far, there are only a handful of empirical studies putting the theory to the test, results have turn out to be inconclusive. Different studies have reached different conclusions, one supporting the long tail, the other rejecting it. Despite the fact that Anderson's ideas have been discussed since 2004, the topic has not ceased to be relevant, and there is plenty of room for research, given the

contradictory results obtained so far.

For this thesis, the long tail hypothesis will be studied from a broader perspective than the one used by most studies so far. Most studies of the long tail have focused in measuring market concentration. However, by only measuring concentration, other outcomes of online distribution and digital sales are not even considered, despite being also a fundamental part of what the long tail hypothesis is.

For this reason, measures of market concentration will be complemented with other measures that will capture fundamental differences that exist between the digital and the physical market. By approaching the long-tail hypothesis from these different angles, this thesis will attempt to unveil information that has been overlooked by other research done on this topic. This information will give a greater insight into the role of e-tailers in the market and how the digital market might be different from the market for CD's.

1.3 Aim and research question

The main purpose of this thesis is to test whether or not there is evidence of the effects described by the long tail hypothesis in the market for CD's and MP3 files in the Netherlands. This study of the music market will be done in a broader way than only looking at one aspect or relaying on one single measurement. The long tail effect will be defined in a broader sense, not only looking at sales, but also at the inner complexities and differences in the digital and physical market. This research was carried out with the following objectives in mind:

- To gather secondary data on the music market: top selling titles, number of firms, sale's concentration levels, among others.
- Analyze and compare the differences in sale's concentration for the best selling titles between digital/online retailers and the offline retailers.
- Make an analysis of the trend of the music market either towards a long tailed distribution or to more concentration around hits.
- Study other secondary trends in the market: diversity, differences in composition of the "head", industry structure in terms of recording companies.

Bearing these considerations in mind, the main research question this thesis will address is:

Is there evidence of a long tail effect in the music market in the Netherlands?

In order to answer this question, the fundamental issue that has to be tackled is: what is understood by *long tail effect?.* A tail effect can be measured in different ways, according to the

angle from which we look at the market. For this reason, this research question will be subdivided into different sub-questions that will attempt to cover the different aspects surrounding this hypothesis:

- *Is the market share of hits lower for digital products?*
- *Is the overall demand for records more equally distributed in digital markets?*
- *Are the best-selling records different in digital format than in physical format?*
- *Does niche content sell better in digital format?*

1.4 Thesis structure

This thesis will be structured in the following way. The second chapter will describe the different theories that have been used to explain the market outcomes of the music market. This chapter will be focused on two contrasting interpretations of the market: the superstar effect and the long tail effect. It will also give an overview of the results that have been obtained by different empirical studies made in recent years.

The third chapter will provide an overview of the music market in The Netherlands. The overall trend in this market will be described, this overview will include sales, number of stores and market distribution among different products.

Next, a fourth chapter will explore the different ways on which the long tail hypothesis can be explored. This phenomenon can be explored at different levels, these levels will be described and the pros and cons of each method will be discussed. It will also describe the methods that will be used in this thesis. Additionally, it will also describe the difficulties that one might encounter when trying to define and measure what the *long-tail effect* is. This discussion is fundamental for any data analysis since it will clarify the reasons why a certain method was chosen.

The results will be divided into two chapters. The fifth chapter will cover the study of the market for CD's and the sixth chapter will focus into the digital market and the differences between this market and the market for CD's.

The final chapter will provide conclusions. This chapter will come back to the questions raised through the whole thesis and will try to answer the main research question. The main findings of the thesis will be summarized and the paths for future research will be addressed.

2

Market for music: Superstars or long tail?

This chapter will address two different scenarios that have been used to describe the music market: superstars and the long tail. Superstars have been up until recent years the most common way to understand the music market. Therefore, record companies have tried to find the next big superstar in order to dominate the market.

Recently, ideas like the Long tail hypothesis have questioned this practice, arguing that in a digital world, hits will lose importance. However, the same digitalization process that is supposed to reduce market concentration can be a reason for even more concentration, both views predict a different scenario as a consequence of the same process. The long tail hypothesis, being the main topic of this thesis will be discussed in more detail.

2.1 The superstar phenomena

Superstars arise in the market when a "relatively small number of people earn enormous amounts of money and dominate the activities in which they engage" (Rosen, 1981). There are different reasons that have been used to explain the emergence of superstars. The first reason for this dominance is the imperfect substitutability of the product of superstars by others. The product of superstars should have greater quality in order for this to happen and this quality should be observable by consumers.

The other reason used is the development of technological improvements that facilitate endless reproduction, one single artist might be able to serve, at least from a production point of view, the whole market. Internet and digital files should according to this view, result in even greater income for superstars since they can now reach even more audiences.

MacDonald (1998) explain the process of star formation by dividing the process in two stages, first artists' quality is observable by audiences, those artists obtaining good reviews and positive feedback obtain greater market share in a second period, since they pose a smaller risk to audiences than newcomers, the consumption capital for this artist increase and the benefit of consuming the same artist increases.(Towse, 2008: 431). Superstars rise because at the beginning they are better known than other artists, artists who's quality is yet unknown for consumers. Consumers gain understanding of the music style of an artists, which additionally allows them to enjoy them more. New artists pose a higher degree of uncertainty for consumers, this puts them

in disadvantage in front of those artists already known and whose style is already assimilated.

Additional factors that facilitate the emergence of superstars are network externalities and informational cascades. Direct networks externalities have been defined as "those generated through a direct physical effect of the number of purchasers on the value of a product" (Liebowitz & Margolis, 1988). In case of a product as music we can also take as an externality the benefit a listener receives from being able to communicate with other listeners of the same artist (Kretschmer, et al, 1999). Frank and Cook (1995) attribute winner-takes-all markets to the social nature of individuals (Elberse & Oberholzer-Gee, 2008), people need to share their experiences and tastes with others. Therefore, when music is consumed, some of its value is created by social interaction.

Furthermore, informational cascades arise from observing the choices of others and obtaining information about artists' quality from their choices. Someone observes that many have bought an album, without knowing their reasons for this choice, however, he assumes that their choice reveals information about the quality the product might have. Given the little information contained in the observed behavior, cascades are fragile and prone to change when receiving new information (Hirshleifer, 1994).

This might explain why some artists' popularity is temporal , since the choice made by consumers was not based on quality judgments. The possibility that informational cascades arise, open the possibility for superstars to emerge even if they do not posses a product of higher quality.

There are some limitations to the superstar idea, tastes are heterogeneous and a couple of superstars will not satisfy all costumers, this is one of the questions posed by the long tail hypothesis, what would happen to the market if by removing physical space constraints an if infinite variety could be offered?

2.2 The long tail hypothesis, an alternative view for the online market

The "long tail" term was coined by Anderson after observing data from Rhapsody, an online music retailer. After plotting the graph of this data, he observed that the demand curve took a shape where a few titles captured a large number of sales, while a large number of titles sold a few units, however demand never became zero, even after the title ranked number 400 000. This large group of titles selling few units gave the distribution of sales a long flattening curved shape. This is known in statistics as a long-tailed distribution (Anderson,2006:10). The head represents the most popular items, selling many units, while the tail represents items selling a few units.

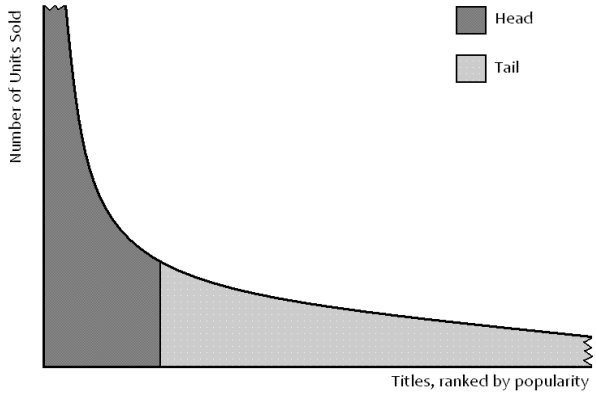


Figure 2.1 : The long tail

Anderson summarizes his idea in the following terms: "The theory of the Long Tail can be boiled down to this: Our culture and economy are increasingly shifting away from a focus on a relatively small number of hits (mainstream products and markets) at the head of the demand curve, and moving towards a huge number of niches in the tail. In an era without the constraints of physical shelf space and other bottlenecks of distribution, narrowly targeted goods and services can be as economically attractive as mainstream fare." (Anderson ,2006: 52)

The long tail hypothesis as stated by Anderson, can summarized in the following points:

- In the digital era, where there are no shelf space constraints, the cost of carrying out a bigger number of niche products is reduced almost to zero.
- With the possibility of selling/buying through internet, consumers are no longer restricted to the physical area where they live, making it easier for any product to find a customer in any place of the world.
- Given these two conditions: low cost of storing a wider range of products and the lack of physical barriers to reach potential customers, a new model of making business focused on the tail would become profitable.

2.2.1 How to benefit from the long tail?

In order to enjoy the profits of a long-tailed approach, Anderson proposes certain "rules":

- **Make everything available:** It does not matter if the items are no longer sold in regular stores or if they had low sales when launched for the first time, in the digital market this "misses", as he calls them, have still the potential to find costumers. Given the large number of these items, when grouped they can become a contender of the few best selling items.

- **Help me find it:** It is not enough to offer greater variety, these products should find customers, therefore, companies should facilitate consumers finding products via tools that lower their search costs. These tools come in the form of customers reviews, allowing consumers buying to see what others consumers buying the same item have bought, and allowing preview of the content of the product.

The first rule creates a long-tailed stock, products are made available that won't be found in regular brick-and-mortar stores, they are either back catalog items or highly specialized content. Moreover, these items could be items that never reached a physical store, since they were produced without the support of a big corporation that would cover the cost of marketing and distribution. This an additional aspect of digitalization that contributes according to Anderson to a long tailed economy. The easy access to the tools of production makes more items available, as authors publish their own books without the need of a publishing company or musicians become successful without having a deal with a major record label.

The second rule in theory drives consumers down the tail, by bringing to the front tail-products that otherwise would have been difficult to discover by consumers. By a few clicks in a computer, a consumer might jump from an item ranked in the top 100 to one ranked below the top 10000 or further down.

2.2.2 The different types of retailers and their economic cutoff point

Anderson calls retailers that target the long tail as aggregators: "a company or service that collects a huge variety of goods and makes them available and easy to find, typically in a single place" (Anderson, 2006:88). Anderson calls companies dealing with physical products such Amazon, hybrid retailers, and the ones selling digital products pure digital retailers.

The difference in practical terms between these two types is that hybrid retailers have a limit on how much they can venture down the tail, in case of physical goods, "somebody's got to store them somewhere before they're sold" (Anderson 2006: 90). In contrast a pure digital retailer can go as far down the tail as it wants, since storage can grow almost without a limit and without a significant increase in costs.

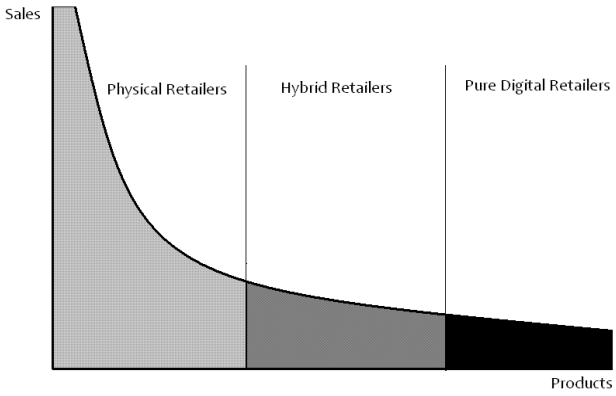


Figure 2.2: *Different types of retailers and the economic cut-off point*

2.2.3 The 80/20 rule and the long tail

The Pareto principle also known as the 80/20 rule, has been used to describe different phenomena such as distribution of wealth. It was notice by Pareto how the 80% of wealth fell in the hands of only 20% of the population. In the case of markets such as music, the 80/20 rule would mean that 20% of products account for 80% percent of revenues.

For an online long tail retailer the distribution according to Anderson will be different than the 80/20 rule. For an online retailer having 10 times as much inventory as one offline retailer, the 20% from the offline retailer turns into only 2% in the case of the online retailer, this 2% still will account for 50% of sales, the next 8% (the 80% of the offline retailer) account for 25% of sales.

The 90% of inventory available only online will account for a 25% extra revenue when compared with the offline retailer. According to Anderson, the biggest benefit of the long tail, appears when we analyze profits: "Because of the low cost of inventory, the margins for non-hits can be far higher in long tail markets than in traditional bricks-and-mortar" (Anderson, 2006: 133). This means that in the head, mainly made out of new releases, there are high acquisition costs, while the tail, having everything from old titles to independent artists acquisition costs are low, making the profit a company can make from long tailed products higher than for hits.

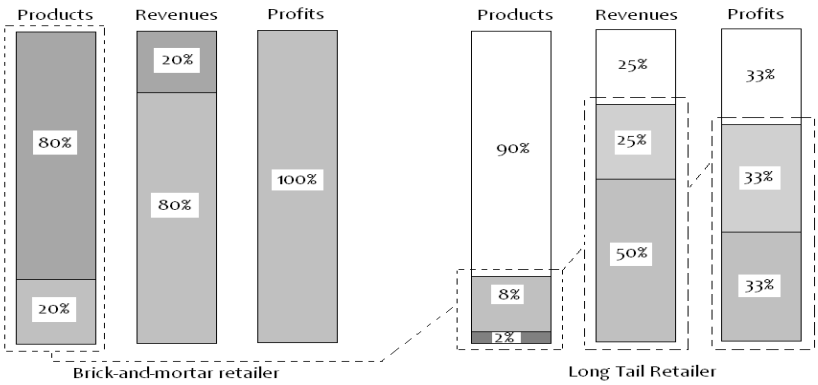


Figure 2.3: *Revenues and profits from the long tail*

2.2.4 Changes in the demand structure?

There are questions regarding demand that arise from the long tail hypothesis. The first question is the effect increased variety has on hits' sales. Will hits sell less due to increased variety? According to Anderson, lowering search costs for niche products by means of recommendation tools, filters and samples will shift demand towards niche products, *reducing the share of hits*.

Another question that the long tail hypothesis raises is whether more variety increases demand or only shift it towards niche contents. Anderson does not give an answer to this matter, he claims it would depend on the sector whether demand increases or not.

Individual demand might not increase since time and attention continue to be scarce: there is a limit to the amount of products one individual can consume. For example music consumption occurs in time, therefore, consumption cannot be increased only by increasing variety, in this case, consumption will be better suited to individuals' taste thanks to the increased availability of products However, there is a limit to how much one individual will increase their expenditure. It is important to bear in mind that increases in sales for an e-tailer can be due to increases in individual demand but also due to aggregation of widespread consumers that were not buying products before.

An additional issue that is briefly discussed by Anderson in his book is whether prices should increase of decrease for long tailed products. To answer this question he makes the distinction between "need" and "want" markets. A "need" market is where a customer know what he wants to buy and is not able to find it except online, these markets will be more price insensitive. On the other hand in "want" markets it is possible to encourage consumers to try an unknown product (Anderson, 2006: 138-139).

2.3 The two scenarios and the role of search costs

There are two potential extreme outcomes from the digitalization process in the music market: superstar domination or an increased variety being consumed. For the supporters of a superstar market, online markets will further increase the power of superstars since they will be able to reach even more consumers and a bigger share of the market. The need for socialization and the lower search costs for hit-products will lead consumers to focus on a small number of artists.

On the other hand, for supporters of the long-tail idea, consumers prefer to buy products that are tailor-made to their personal taste. This was not possible in a brick-and-mortar world due to physical constrains. Although Anderson does not say that superstars will disappear completely, he claims that consumers will consume more and more of lesser-known products, resulting in a less dominated market. One of the main reasons Anderson gives for this shift is the lower search costs for online products due to recommendation tools.

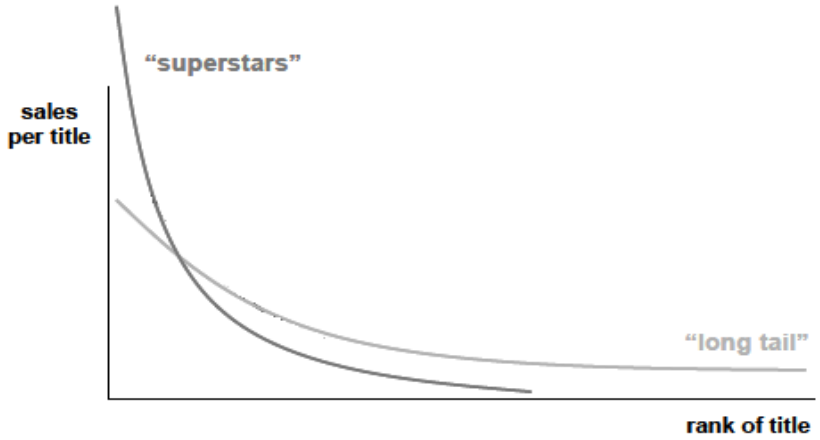


Figure 2.4: *Superstars and the Long tail (Source: Elberse, A., F. Oberholzer-Gee. 2008)*

Search costs play a fundamental role in determining which scenario will materialize in real markets. As it was described above, risk and incomplete quality information are factors that could prevent consumers from buying niche content. Online, search costs for obscure titles can be lowered, consumers can obtain information about these products with ease. Not only previewing content is becoming a standard practice by online retailers, recommendation tools and peer reviews provide more information about quality for consumers to make their choice.

On the other hand, this availability of information, in a digital world where thousands of new products appear each day, might also have an "over-supply" effect on consumers. Consumers might find so much information that get discouraged and confused by the immense range of options. (Elberse & Oberholzer-Gee, 2008:7)

Without empirical data it seems impossible to take a stance on which of the two markets will be arising in the future, when Anderson's book was published there was not much data available except from the data Anderson himself had used. From 2006 onwards and as the long tail idea became known more data and research started to emerge.

2.4 The evidence so far

The number of papers dealing with the long tail has been growing since 2006, when the book was published. However, the results are still inconclusive, some evidence seems to support the theory, other seems to contradict it. This section will review the most important studies made so far on this topic.

Brynjolfsson et al. (2007) investigated the effect lowering costs has on sales by studying a retailer selling via two different channels: online and by catalog. Featured products on a catalog can be easily found by consumers while they have to make a bigger effort to locate in the catalog pages for non-featured products. In contrast, online there are search engines and recommendation systems, making it easier to find non-featured products. As consumers acquire experience with online channels their search costs are also reduced. Using a random sample from both channels they estimate Lorenz curves and Gini Coefficients.

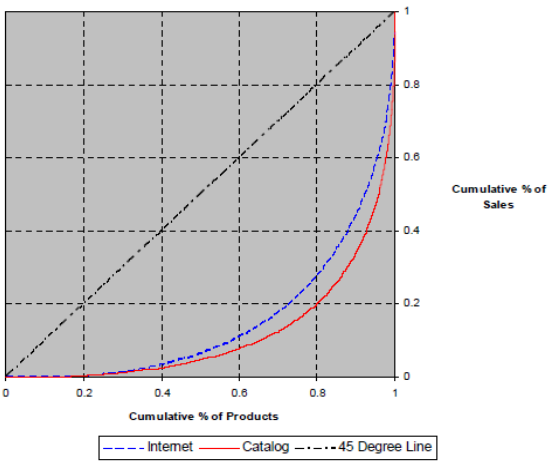


Figure 2.5: *Gini Coefficient for different distribution channels (Source: Brynjolfsson et al. 2007)*

The Lorenz curve and the Gini coefficient are frequently used as measures of inequality in income distribution, the Gini coefficient is the ratio of the area between the Lorenz curve and a 45 degree line, this line represent an equal distribution among all products . The more concentrated sales are, the more the Lorenz curve moves away from the 45 degree line and the higher the Gini coefficient becomes. The results show a higher concentration for the catalog channel with a value

of 0,77 , for the online channel Gini coefficient had a value of 0,70. Since it could be argued that consumers are significantly different online than offline a sample made with consumers buying in both channels was made, returning similar results.

This study was focused on the effect that search and recommendation tools have on sales. However, in this case both markets have the same number of individual offerings, therefore, it does not test the effect of recommendation tools might have on stocks of different size. Despite this fact, the study does prove that appropriate recommendation tools, might shift demand from hits to niches.

However, not all results support the long tail hypothesis, Elberse & Oberholzer-Gee (2008) using a data from home-video sales in the US for the 2000 to 2005 period, find that the number of titles in the top 10% of sales declined more than 50%, a sign of a winner take's all market.

Although there is a move towards the tail: "at the tail end, we find that there is a rapidly increasing number of titles that never, or very rarely sell-the long tail appears incredibly flat" (p 24). The market is becoming more concentrated because fewer hits attract the majority of the market, niche titles are still selling few units, which makes the middle of the distribution flatter, they also find evidence that there is over-supply of titles.

Tan & Netessine (2009) studied data from Netflix, an online video rental company, containing customers' reviews. In their study they analyzed hits and niches both in absolute terms (top 100) and relative terms (top 10%). They found that from 2000 to 2005 there is no significant evidence of the long tail. When measured in absolute terms, consumption decreases for hits, while when measured in relative terms, hits consumption increases over time.

They conclude that new movies appear faster than consumers can discover them. Moreover, only a small group of heavy users, users who consume more than the average consumer, are the ones likely to venture down the tail.

Another study that challenges the long tail was made by Elberse. (2008), according to this study distributions are quite concentrated towards hits, for example, Rhapsody's distribution of music sales is dominated by the top 10% of titles accounting for 78% of plays (Rhapsody is an online music streaming service).

She also studies the music industry by using Soundscan data, she finds that concentration towards hits is higher for digital tracks than for physical CD's. However, when describing Rhapsody's results, Elberse makes the remarks that "one percent of a million is still 10 000, far more than the number of titles most U.S. Radio stations play in a given year...and equal to the entire music inventory of a typical Wall-Mart store" (Elberse 2008:2)

A study that tested the market share of hits was published in Billboard magazine (November 14, 2009: 24-28). Using data of Nielsen Soundscan from 2004 to 2009, the authors

found results that on one hand point towards more concentration and on the other hand towards a long-tailed effect in sales. When it comes to individual tracks, the market in the US has become more concentrated.

The market share of the top 10 tracks increased from 2.1% in 2004 to a 3.1% in 2009. The same trend can be seen for the top top 10 , top 100 or top 200. The market share of the top 200 increased in the same period from 14.5% to 18.7%. However, there is some evidence of the long tail when it comes to albums and digital albums. Although music sales have been declining at all levels, hit albums were affected the most.

Sales of the top 5000 albums dropped 40.5% while albums in the tail declined 27.4%. In the case of digital albums, the top 5000 accounted in 2008 for 64.7% while the top 5000 in overall album sales accounted for 70.2%. Furthermore, hit digital albums have been losing market share.

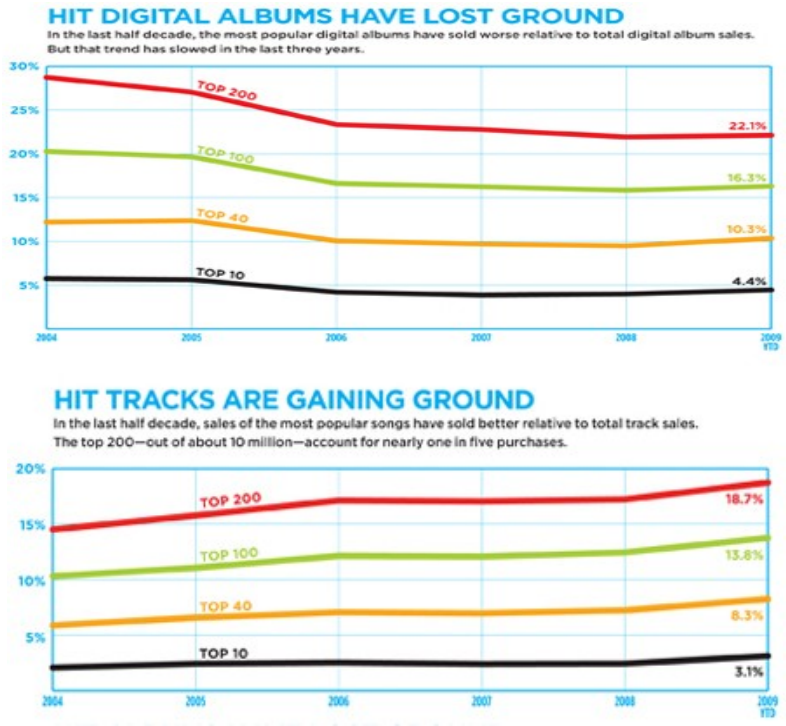
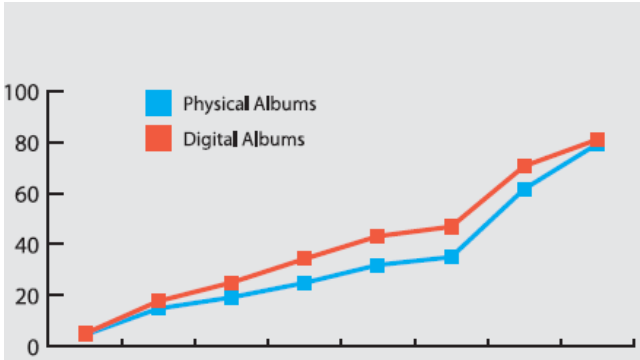


Figure 2.6: Digital hits in the US (Source:Billboard Magazine November,2009)

It is important to remark that even though Nielsen Soundscan is the most comprehensive source for information for the U.S. music market, the fact that the study covers only five years limits its explanatory power. Besides the reduced number of years analyzed, using only the top 200 does not describe if demand is shifting towards the tail or only being shifted to titles ranked from, for example, top 200 to top 500. Although this is a shift of demand, it could hardly be called a shift to the tail, since those albums would still be part of what Anderson would consider as head.

A similar study was published in the UK by the Entertainment Retailers Association (ERA). However, they don't look to the trend over time, but rather compare the differences in concentration levels between digital and online products for the year 2008.



% of sales at different chart positions				
	Singles		Albums	
% of market taken by	Physical	Digital	Physical	Digital
Number One	22.7	2.6	4.9	4.4
Top 10	49.3	11.9	17.6	14.7
Top 20	62.3	17.0	24.9	19.1
Top 40	72.9	22.4	34.2	24.8
Top 75	80.2	27.3	43.1	31.8
Top 100	82.3	29.5	46.8	34.9
Top 1000	90.0	50.5	70.6	61.6
Top 5000	91.5	59.7	81.0	79.3

Figure 2.7: *The long tail in the British market 2008 (Source: Entertainment Retailers Association Yearbook 2008)*

From these data it can be seen how the difference between physical and digital is minimal at the number one position, however there is a difference that start to grow between the two markets when we move to the top 10, top 20 and so on. For example at the top 40 there is a difference of 10%, that difference continues until the top 100. However, the distance begins to disappear and by the top 5000 titles, both concentration levels are virtually the same. Here we can see how important is the way head and tail are defined. If the head would have been defined at the top 100 there would have been an apparent difference between online and offline products.

However, if head is defined at the top 5000, there is no difference between the two markets. Without the whole data it is impossible to tell the true shape of the curve in both cases. However, it could be said that online, products from the top 100 to the top 1000 benefit from a long-tailed effect.

Moreover, the rest of the tail could be said to be flat for both markets, since after the position 5000 the rest of available titles account for only a 20% of sales. To put the "flatness of the tail" into perspective, in 2008 the number of albums tracked by the Official Charts Company was of 216 527. This means that many albums did not sell any copies, as the report states "surprising is the relatively small range of albums sold digitally – at 53,785 titles it is less than a third of the total available".

This is one important aspect for the existence of the long tail effect: is the increased variety being consumed? If consumers are not venturing down the tail then the market stays the same, maybe concentration levels are shifting mildly, however some evidence shows that the effect is only a small fraction of the potential effect desired by increasing variety.

In this respect two studies have deeply questioned the long tail idea. One made by Will Page, an economics of the MCPS-PRS Alliance (a royalty collecting society) who found that of 1.23 million of available albums, only 173000 were ever bought. This means that only 15% have been bought. (Times Online, Dec. 22, 2008) Another study with similar results was made by 24-7 Entertainment regarding the mobile music industry, according to this company out of 4.3 million tracks available for download to their users 3,68 millions have never being downloaded. (http://www.longtail.com/2008/10/does-the-long-t.html)

However, at this point it is important to recall the rule proposed by Anderson: help me find it. There are many options where to buy and download files, but are they all doing a good job driving consumers down the tail? Do all of them use recommendation tools? That type of information is usually omitted from the press notes. Mobile phones today might not be suitable for such systems or might not be the place where users search for new music.

2.5 The importance of how head and tail are defined

From different articles it can observed how one of the reasons why results are contradictory is the way on which "head" and "tail" are defined. Anderson prefers to define the head in absolute terms, while other researchers prefer to define head and tail in relative terms, in other words a percentage of total offerings. Which option is the correct one?

In some studies the market seems to be moving towards more concentration if percentages are used to define head and tail, however, if the top 200 titles are used as a measure of concentration, then the market seems to be shifting towards the tail.

Anderson has defended his theory from researchers claiming that the market is becoming more concentrated instead of less concentrated by saying that: "...trying to define "head" and "tail" in percentage terms is meaningless in a market with unlimited inventory, because the denominator can grow infinitely large...Let's say you have 1,000 items and the top 100 (10%) account for 50% of the sales. Then you add another 99,000 items to the catalog, and the sales of that top 100 fall to just 25% of the total, while it takes another 900 items to make up the next 25%. I would say that demand has shifted down the tail, because those top 100 items have dropped from half the market to just a quarter of it and the rest of the demand is spread over more items."(http://www.longtail.com/the_long_tail/2006/07/factchecking_my.html).

From Anderson's perspective absolute numbers should be used to measure the head and tail, others see percentages as a better way to describe the market. How to measure concentration and how to define head and tail are fundamental questions that have to be answered in order to test the long tail hypothesis. Therefore these questions will be discussed into detail in the chapter concerning the methods to be used in this thesis.

3

The music market in the Netherlands

This chapter will give an overview of the developments that have occurred in the recorded music market in The Netherlands during the last years. The market developments in the Netherlands cannot be understood without placing it into the context of the global music market, since the changes that have been experienced by the Dutch market are forces that have been shaping the global market for a decade. For this reason an overview of the evolution in the global market and the digital market will be given. After having described the global market, the specific characteristics of the dutch market will be addressed.

3.1 Evolution of the Worldwide music market

The music world experienced profound changes in the last decade due to the appearance of digital formats and the widespread use of the internet. A look at the 2000 IFPI report shows that the music market was divided between singles, LPs, Cassettes, CD's and mini discs:

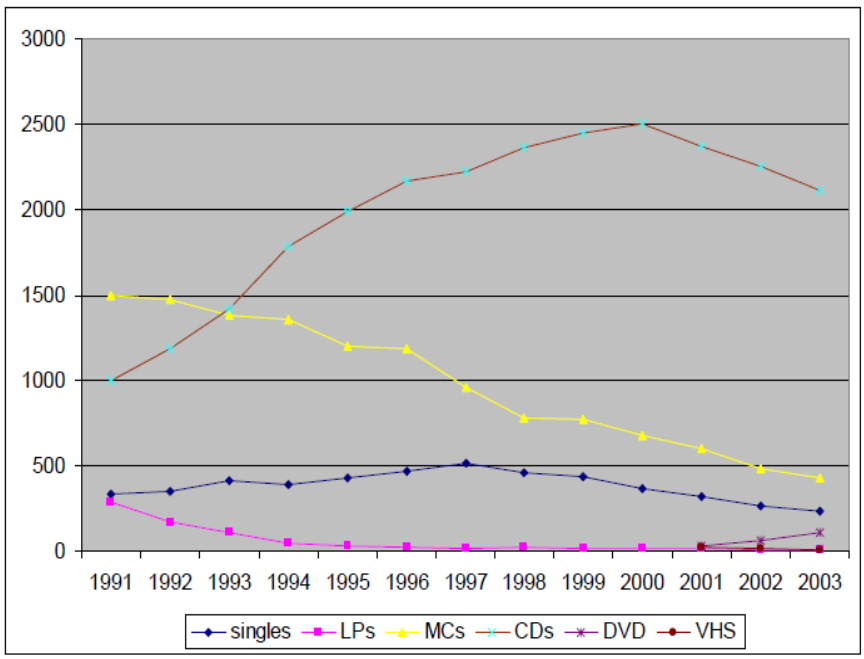


Figure 3.1 : The worldwide music market trend

(Source: *Peitz, M. and Waelbroeck, P. 2005. "An Economist's Guide to Digital Music")*

By now, formats like the cassette and the mini disc have disappeared from the market, and new ones have come in their place, such as mp3 files. The CD took over the market between the years 1990 and 2000 , it can be observed how sales from other formats declined while CD sales grew.

Between 2000 and 2001 the market experienced a decline in sales in all formats. Among the causes that have been mentioned are the creation of Napster in 1999 and the use of file-sharing technologies in 2001 (Peitz & Waelbroeck, 2005: 366) Piracy issues aside, digitalization also opened the door to new ways to distribute and sale music. By 2003 there were about a million tracks available on over 90 music sites worldwide and download sales accounted for over $30 million in the US alone. (IFPI: World Sales Report 2003).

According to the IFPI "2003 was the break-through year for online music services, as record companies expanded their licensing agreements across a wide variety of online retailers, offered consumers greater flexibility of track usage, began licensing the catalog of major international acts and shortened the gap between off-line and online releases" (IFPI online music report 2004)

In 2004 the number of online retailers selling digital music increased to more than 230 worldwide , and services like Napster and Itunes became established. Record companies started to see significant revenues from digital sales and new models of distribution began to be used, services such as *mobile* distribution.(IFPI online music report 2005)

The market continued to grow in 2005: "Some 420 million single tracks were downloaded in 2005, up more than twenty times on two years ago. That excludes the entire business of music on mobile phones, a market which is not far behind music downloads in value. Together in 2005, these two new distribution channels took record company revenues from digital sales to an estimated $US 1.1 billion globally, tripling in value compared to 2004." (IFPI online music report 2006:3) The market continued to grow steadily in the next years and in 2009 digital sales accounted for 25% of all music sales globally (IFPI Digital Music Report 2010).

Table 3.1: *The changes in the digital market from 2003 to 2009(source: IFPI:The Recording Industry in Numbers 2009)*

Digital Music: Charting Change

	2003	2009
Licensed music services	Less than 50	400+
Catalogue available	1 million	11m + tracks
Industry's digital revenues	US$20m	US$4.2 billion
% of industry's revenues from digital channels	Negligible	27%

As it can be seen, the growth in this market has occurred at all levels: number of tracks available, number of online stores and market share obtained. However, although the digital market has been on the rise, the overall music market has been decreasing since 2000. In the five years between 2004 and 2009 the total music market shrank by 30%, a 940% increase of digital sales in the same period could not offset the decrease of physical sales.

3.2 The Music Market in The Netherlands

The Netherlands ranked number 10 position in the worldwide ranking of sales. It is important to notice that the share of digital sales is one of the lowest from the list, accounting for only 6% of sales, on the other hand performing rights account for 18%, the highest of the table. It can be observed that although in the top 10 globally, when it comes to digital sales The Netherlands falls to the 19th place.

Table 3.2: *The world's top music markets(Source:IFPI-The recording industry in numbers 2009)*

Recorded Music Sales
Top 20 Markets

		TRADE VALUES				MARKET SPLIT			RETAIL VALUES	
POSITION	COUNTRY	US$ M	LOCAL CURRENCY	M	% CHANGE	PHYSICAL	DIGITAL	PERF. RIGHTS	US$ M	LOCAL CURRENCY
1	US	4,976.8	USD	4,976.8	-18.6%	63%	36%	1%	8,597.8	8,597.8
2	Japan	4,109.0	JPY	425,159.5	0.9%	78%	20%	2%	5,600.7	579,504.0
3	UK	1,845.4	GBP	1,015.0	-2.5%	79%	14%	7%	2,430.3	1,336.7
4	Germany	1,627.8	EUR	1,106.9	-4.2%	87%	8%	5%	2,355.4	1,601.7
5	France	1,049.6	EUR	713.7	-11.4%	77%	15%	8%	1,559.6	1,060.5
6	Canada	456.3	CAD	488.2	-8.1%	79%	17%	4%	600.9	642.9
7	Australia	389.2	AUD	467.0	-6.0%	84%	13%	3%	578.4	694.1
8	Italy	326.1	EUR	221.7	-17.0%	82%	9%	9%	462.8	314.7
9	Spain	302.4	EUR	205.6	-7.8%	78%	10%	12%	409.9	278.7
10	Netherlands	271.9	EUR	184.9	-9.9%	76%	6%	18%	373.5	254.0
11	Brazil	221.8	BRL	408.2	8.1%	81%	13%	6%	328.3	604.0
12	Russia	220.8	RUB	5,492.2	-2.4%	95%	3%	2%	410.2	10,201.0
13	Switzerland	192.5	CHF	207.9	-5.4%	90%	7%	3%	243.6	263.1
14	Belgium	192.1	EUR	130.6	-4.8%	76%	10%	14%	268.9	182.9
15	Austria	147.3	EUR	100.2	-9.9%	84%	7%	9%	272.9	185.5
16	Mexico	145.9	MXP	1,629.5	-22.0%	89%	11%	0%	231.5	2,586.0
17	Sweden	143.2	SEK	944.0	-6.9%	84%	8%	9%	215.7	1,421.4
18	South Korea	140.6	KRW	155,111.9	16.0%	40%	60%	0%	294.8	325,078.2
19	India	140.4	INR	6,151.9	6.3%	64%	21%	16%	179.3	914.5
20	Denmark	124.5	DKK	635.0	-9.5%	73%	16%	12%	179.3	914.5
	Other	1,391.5	–	–	-5.0%	78%	15%	8%	2,205.8	–
	Total	**18,415.2**	**–**	**–**	**-8.3%**	**75%**	**21%**	**4%**	**27,824.6**	**–**

Source: IFPI

Table 3.3: *The Netherlands in the world's sales ranking(Source:IFPI: The recording industry in numbers 2009)*

World Ranking	
Physical sales	11
Digital sales	19
Performance rights	6
Total market	10

3.2.1 Music sales in the last two decades

The market in the Netherlands follows in general terms the trends described for the global music industry. The decline of sales becomes evident from 2000 onwards, before the year 2000, sales' pattern is irregular with some years reaching sales close to 40 million. Starting from the year 2000 sales have decreased steadily going from 34,5 million albums sold in the year 2000 to 18 million in 2009, a change of -42 % in units sold. As mentioned before, the causes for this decrease in sales are not entirely clear, although online piracy is often mentioned as one of the main causes, the possibilities for buying digital products can also be used as an explanation.

Table 3.4: *Album sales in millions of units (Source NVPI)*

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
CD	35	39,2	33,2	34,6	34,6	35	34,4	36,9	34,2	33,5	34,1	32,4	28	25,1	23,4	20,2	19,4	19	18	17
Total	41	43,2	35,6	36,5	36,2	36,4	35,4	37,7	34,8	34	34,5	32,5	28,3	25,5	23,7	20,6	19,5	19	18	18

In contrast, while physical sales have been decreasing, digital sales have gained ground. Digital downloads increased by ten million in the period from 2005 to 2008, there was no overall data available for 2009. It can also be observed how singles have almost disappeared from the market, probably partly explained due to a shift of demand from singles to mp3 files.

Table 3.5: *Sales per carrier from 2005-2009 in millions of units*
(Source NVPI -"De Entertainmentbranche"reports 2005-2008)

Year	2005	2006	2007	2008	2009
Albums	20,6	19,5	19	17,99	17
Singles	2	1,7	1	0,68	0,52
Downloads	4,3	10,1	12	14,1	

3.2.2 The growth in online distribution

When it comes to online distribution of physical CD's, the role of the internet and online distribution has been growing. The first data available dealing with online distribution is from the year 2001, at that point in time, internet stores accounted for only 2% of music sales. The percentage had grown to a 17% in 2008. There was no data available for 2005 and 2009. Without the data of 2005 it is not possible to establish in which year did online sales began this upwards trend, however, in 2006 it could be seen how the online channel had almost tripled its presence in the market. The trend continued in the following two years.

Table 3.6: *Percentage of CDs sold via the online channel (Source: NVPI year reports 2005-2008)*

Year	2001	2002	2003	2004	2005	2006	2007	2008
%	2,00%	5,00%	3,00%	4,00%	–	11,00%	14,00%	17,00%

On the other hand, the number of "CD stores" (CD-zaken, are stores selling CD and video and games, therefore there are other sectors included into this category) has declined over the years. There has been a reduction of 230 stores in the period from 2005 until 2009. (Hoofdbedrijfschap Detailhandel website: http://www.hbd.nl)

Table 3.7: *Number of Physical CD stores (Source: Hoofdbedrijfschap Detailhandel www.hbd.nl)*

Year	2005	2006	2007	2008	2009
# of stores	850	790	710	690	620

3.2.3 Sellers of digital music in The Netherlands

In 2009 there were 28 online stores in The Netherlands where mp3 files could be bought or streamed .Out of this number, only seven offered more than a million tracks from both majors and indies: Itunes, Planet Music, Glandig Music, Mediagigant, Surf2music, Toost, Jaha. There are some cases where no information was available about the number of tracks. However, some stores might also surpass the one million mark, stores such as 7digital, one of the biggest online stores and Last Fm. There are also stores that have a very limited supply, between 200 and 10 000. There are only 3 online stores serving the cellphone market, however , all of these offer more than a million tracks. (NVPI: Het digitale aanbod van beeld en muziek)

3.3 Summary

Overall, online channels are gaining importance in the market, not only as a distribution channel for physical CD's but also as a source of digital files, a growing force in the music market. It cannot be said that demand is simply shifting from physical to digital, the decrease in physical sales is not compensated through digital sales: overall sales and revenues continued to decrease, however, the growing importance of online distribution and digital music cannot be denied.

Moreover, for the Dutch market, digital and online sales are a recent phenomena, starting to gain significance only after 2004, in comparison with the global market, the Dutch digital market is lagging behind. According the the IFPI, the market share for digital sales was of around a 6% (IFPI: The recording industry in numbers 2009) This situation is different from other countries where the 6% mark was already surpassed years before, countries like the USA, or the U.K.

4

Definition and Measurement of the Long Tail

4.1 The difficulties of measuring the long tail effect

4.1.1 Multiplicity of outcomes

At first sight the long tail hypothesis seems to be simple: make everything available because adding up thousand of niche titles online will turn out to be profitable. However, the deeper one goes into the topic, the more difficult it becomes to define the long tail hypothesis on simple terms.

The difficulties become evident when reading Anderson's book more into detail since it touches many aspects not directly related to this main hypothesis: changes in the Pareto distribution, increase in the number of back-catalog sold, better performance of niche albums online, among others. For example, the book begins with the story of an unsuccessful book that became a best-selling book due to the power of online recommendation systems; in this case we are not speaking of the tail anymore, but of the rise of a new hit that used to belong to the tail before.

The problem we encounter when trying to refer to the long tail using all these arguments is that these predictions are *not necessarily dependent on each other* and what is even more important, each of these outcomes may have a different cause. The long tail as it is described in Anderson' book, is a complex of causes and effects; to summarize it in terms of profits and the number of titles being bought, turns out to be a simplistic way to approach a phenomenon that is much more complex.

The difficulties arise especially *when trying to measure* the presence of the long tail effect on the market. One of the difficulties of testing the long tail hypothesis is that it comprises many different predictions. As it is described in Anderson's book, the long tail effect encompasses so many different outcomes that it is difficult to either accept or reject it unless the long tail effect is defined in a more narrow sense or unless every claim made in his book is tested. That has made the long tail an elusive phenomenon, when one argument fails to support the long tail, another one provides with evidence supporting it.

The outcomes of the long tail can be explained through different causes: the role of lower

search costs online, the effect of more variety available or simply an increase in the size of the market reach due to online sales. For these reasons, if we take only one of these outcomes as a proof of the existence of a "long tail effect" in the market, we would be only referring to one aspect of what the long tail idea is.

As it was described in Chapter 2, Anderson's hypothesis relies mainly on the economic potential of the tail, a tail where *adding millions of niche sales would rival sales of hits.* Based on this prediction he builds a business model for the e-tailer of the future and describes a new market structure that challenges the importance of superstars and big record companies. It is for this new scenario that all the other outcomes are used either as a proof or as a result of the long tail effect.

One effect described in the book has to do with the *additional* sales that niche content available *only* online will bring. These additional sales are the ones grabbing market share from hits and the ones increasing variety consumed. However this outcome can be the consequence of two different situations: either the same group of consumers lowered their consumption of hits or new consumers were reached due to the lack of physical barriers online, and these new costumers are consuming niche content. In one case, consumers are shifting from hits towards niche content, in the other, consumers are not changing their preferences, only new consumers were reached thanks to the lack of physical barriers online.

In either case, if we take the number of titles consumed as a measure of the long tail, it seems that the number of titles consumed online should increase. However, what if the number of different titles consumed online is lower than offline, but still hits are losing market share? In this case, the tail might be shorter online, but consumers might be consuming more titles that were hardly bought offline or that were not available via physical stores. This effect could be attributed to the "long tail filters", the online tail might be shorter but it can be fatter. If we take the fact that hits are losing market share as the essence of the long tail effect, then this example clearly shows a long tail effect.

Another case could be that the hits are more important online, their market share is bigger than in the physical world. This would be a market without the presence of the long tail effect if we look only at concentration. However, what if when studying the composition of those hits, they were only offline-tail-content? Or content only available online? In this case, variety offered online does not necessarily requires to alter the demand curve for hits, but only to change which artist becomes a hit.

4.1.2 How to define what is the real long tail effect?

Based on the examples given above, there are different ways on which the presence of the long tail can be defined:

- ***The tail becomes longer***: additional titles are added at the end of the distribution, these additional sales could come from new consumers or could come from consumers shifting from hits to niche content. In this case, a tail-effect means *more titles in the distribution.*

- ***The tail becomes thicker*****:** This outcome has to do with a shift from hits to niche products, but in contrast with the previous definition, the tail needs not to be longer, as long as the market share of hits is decreasing and the market share of niche content is increasing. In this case a tail-effect means *less market share from hits.*

- ***Content available online only becomes part of the items sold***, either as head or as tail: If the composition of the online demand includes content that is only available via internet, it could be said that there is presence of the long tail, consumers are buying items "off the beaten path", thanks to filters and recommendation tools. In this case the long-tail effect means *different titles in the distribution.*

It is important to note how all of the three definitions of the long tail effect leave unaltered the idea that retailers should make *everything available* online and that profit can be made by selling niche content. No matter if additional titles are added up to the sales or if only there is a shift the distribution of these sales, the importance of *niche content* seems to be the common element when defining what a long tail effect is.

If we consider the definition of "tail" to be closely related to the idea of niche content, a "tail effect" becomes then a broader concept that encompasses aspect like differences in artists online-offline and performance of previously "tail" content into digital markets. Measuring only the market share of hits for example, will not properly describe if there are additional titles being consumed or only a shift towards already consumed niche titles.

Therefore, in this thesis, these three ways of defining the long tail will be explored in order to have a wider understanding of the outcomes that have occurred in the Dutch music market due to digital sales.

4.1.3 The difficulties of comparing market concentration from different markets

Defining how to measure concentration is an important aspect of any study of the long tail hypothesis. What is usually left out of the analysis is that either measuring concentration in absolute terms or in percentages, these measurements will be influenced by the overall shape of the distribution.

Long tailed distributions are also sometimes called power law distributions. In contrast with many other phenomena where most values peak around a typical value, for example those following a normal distribution, these distributions are heavily right skewed. The power law distribution can be represented by the equation:

$$p(x) = Cx^{-\beta}$$

The exponent of the power law is *minus Beta* (-β), this exponent is a measure of the magnitude of concentration, the higher the value of β, the more skewed the distribution will be. As an example, let us imagine a music market where there are 10 hits and 100 different titles. Having a *Beta* of value 1 (thus, power exponent -1), the market share of those ten hits will be 56,5%, or if we take those hits as a percentage of total offerings, 10% of the market account for 56,5% of sales. If *Beta* increases to a value of 2 (thus, power exponent -2), then the market share of the same ten hits increases enormously, now top titles account for 94,8% of sales. The result in a graphical representation can be seen in Figure 4.1

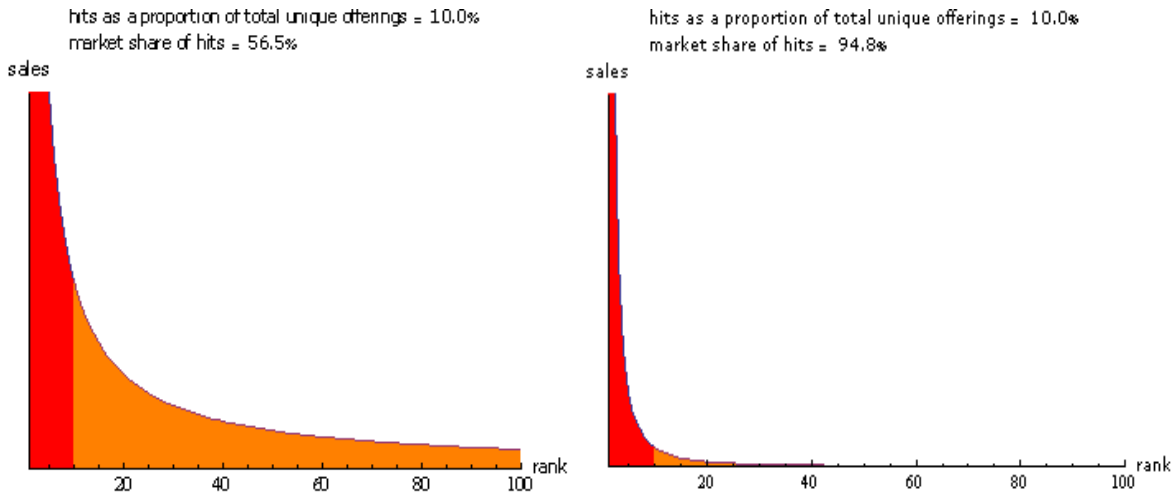


Figure 4.1: *Power curve and the effect of a higher power exponent (made using Wolfram Mathematica Player,"The Long Tail" from The Wolfram DemonstrationsProject: http://demonstrations.wolfram.com/TheLongTail/)*

There are two ways for hits to lose market share, either more titles are consumed or the value of Beta becomes lower. For a market to show a long tail effect, one of these two different processes should be present. When the same market is studied over time, both processes will have

as a result that hits will lose market share.

However, when two different markets are studied, aspects such as size, and shape of the distribution have to be taken into account in order to make any comparison. Comparing two different markets represents a bigger challenge than comparing the same market over time. One of the challenges involved is how to measure concentration: by means of using percentages or by means of using the top "x" titles.

These challenges will be explored by means of three different examples. For these examples a power curve will be assumed to be the underlying shape of the market, although in practice it is much more complex to determine which type of equation fits better any given distribution, for example a log normal or exponential equation might be a better choice for a specific distribution, however it is not in the scope of this thesis to explore the differences and how to choose from one of these possibilities. What this assumption will provide is a reference point to describe the overall shape of the curve in terms that can be used to compare different curves disregard of size or volume. The approximation curve will be plotted in the graphs using a thinner line.

4.1.3.1 Example 1: Consistent result measuring concentration when using top titles or percentages

The following example consist of two short data sets having the same number of offerings but different sales across titles. The sales for Market A are more concentrated towards hits than for Market B. In this case, both the analysis by means of percentages and the analysis by means of top titles provide the same outcome. It can be seen that both top titles and a percentages indicate a higher concentration for Market A, in this case, the Top 2 and the 10% of titles are the same: 97% in market A vs 62,9% in Market B.

Table 4.1: *Example 1*

	A	B
Titles offered	20	20
Top 1	**77,78**	39,57
Top 2	**97,22**	62,9
Top 3	**99,16**	76,51
10% titles	**97,22**	62,9
20% titles	**99,28**	84,29
Beta	**2,68**	2,53
R^2	0,88	0,85

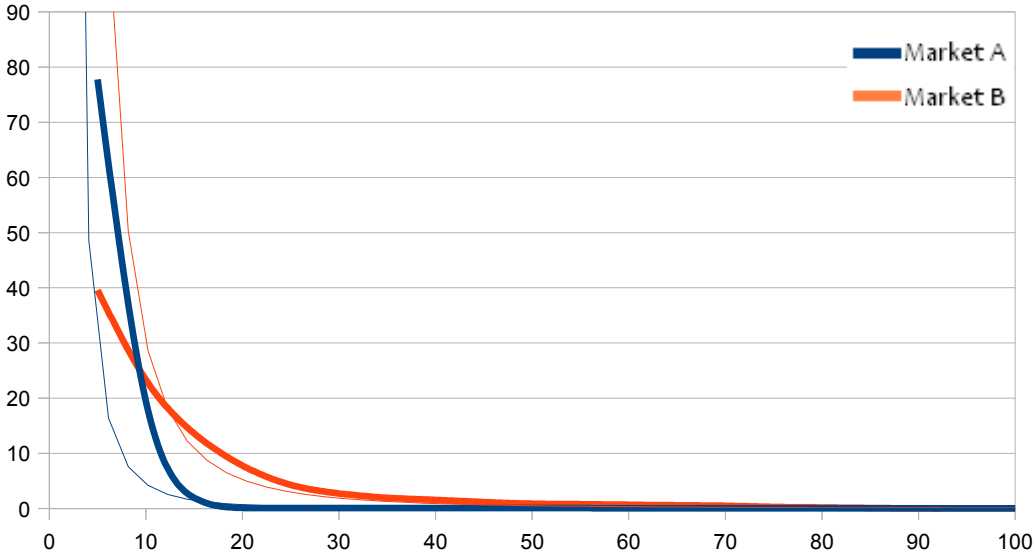


Figure 4.2: *Example 1*

If we consider both markets to follow a power law distribution, then the power exponent of each of them could be also used as a way to measure concentration. After comparing both exponents, Market A turns out to be more concentrated since it has the highest value for *Beta* (2,68). The R^2 value represents the goodness of fit of the power curve to these particular curves.

An intuitive conclusion that can be made out of this simple example is that in markets of an equal or very similar size, both top titles and percentages might be good measurements of concentration, and therefore can be used to identify a long tail effect.

4.1.3.2 Example 2: Different results percentages/top titles

The second example was made using markets of a different size. In this case, the difference in size and shape of the curves resulted into opposite results when using top titles and percentages as a way for analysis.

Table 4.2: *Example 2*

	A	B
Titles offered	50	20
Top 1	45,41	44,89
Top 2	71,65	72,95
Top 3	83,76	84,18
10% titles	**94,36**	72,95
20% titles	**98,5**	90,91
Beta	2,59	2,3
R^2	0,94	0,97

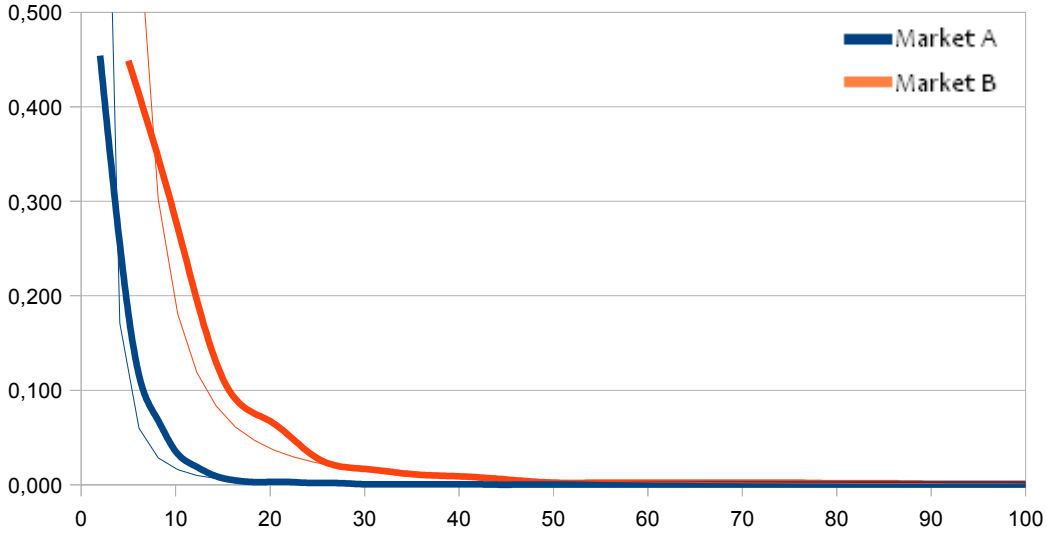


Figure 4.3: *Example 2*

If we were studying the market and only the top 3 titles were used as a way to measure which market is more concentrated, then Market B would appear to be more concentrated than Market A. However, if the 10% of titles were used, then Market A would appear to be the most concentrated market.

However, percentages are in this example a better way to describe the market. What percentages will show in this case, is that as we move further down the tail, titles in the middle or the tail have more market share as well. If a power law equation is approximated, Market B has a lower beta coefficient: 2,3 with a goodness of fit of 0.97, thus less concentrated. Percentages are consistent with this result.

In this example it could be said that two outcomes of the long tail hypothesis are present at the same time. The first one is a longer tail and flatter tail, which reduces the share of hits (Market A) and the second one has a thicker but shorter tail (Market B), which makes the distribution more even. A conclusion that can be made from this example is that for markets of a different size, top titles will not provide a good description of how concentrated the market is, despite the fact that hits might have less share of sales.

4.1.3.3 Example 3: Shift in concentration after reaching a certain point in the distribution

It was described above how markets of the similar size could be analyzed using either top titles or percentages, however, there is the possibility that neither of those measures provides with an accurate description of the whole market. The following example will show the possibility of one market going from a more concentrated distribution to a less concentrated distribution. To make

the example clearer, both hypothetical markets were made of equal size, in order to make percentages and top titles follow the same direction.

Table 4.3: *Example 3*

	A	B
Titles Offered	20	20
Top 1	62,11	49
Top 2	82,82	74,73
Top 3	89,03	86,98
10% titles	82,82	74,73
20% titles	91,1	93,1
beta	-1,9	-2,88
R^2	0,94	0,92

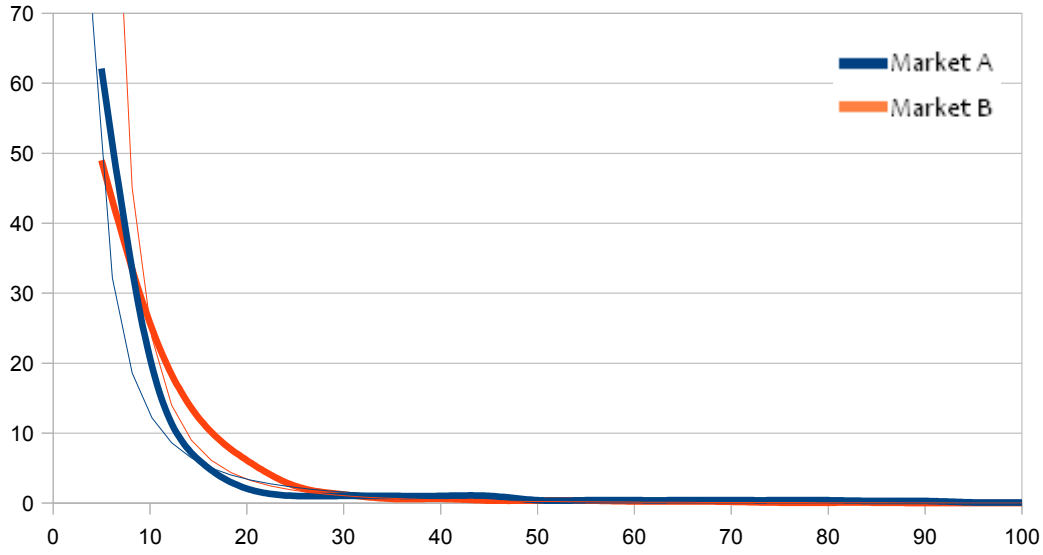


Figure 4.4: *Example 3*

If one had made the cut at the Top 3, the results obtained would have pointed out that Market A is more concentrated at the head of the distribution than Market B. However moving one position further to include the Top 4 title, there is a shift, now Market B appears to be more concentrated towards the top titles than Market A. The same applies in this case for percentages when 20% of the market is reached. This shows the danger of defining a segment of the market to be studied without considering the overall shape of curve.

4.1.3.4 Summary of the examples: precautions when interpreting results and approach to be used in this thesis

Without being exhaustive, the previous examples were meant to show the difficulties of claiming that one market is more concentrated than the other by simply comparing percentages or top titles. From the examples some insights might be given. For markets of a similar size, both percentages and top titles will provide the same information, however, when the difference in size is important using top titles overlooks important information about the overall shape of the curve. Yet, there is a risk in using both percentages and top titles, since the results might shift at a given point of the distribution. Therefore, if the data allows it, a measurement that describes the overall distribution should be used, for example using Lorentz curves and Gini coefficient.

4.2 The difficulties of defining head and tail

Another aspect that present difficulties is how to make the division between what to include in the head and what to include in the tail. There are two main possibilities, to define the head as a percentage of titles in the distribution or to define the head in reference to the physical market. There are difficulties when using both approaches.

In the online world, if a percentage of the titles bought is used to define the head, we might end up with heads that include so many titles that they cannot be used as a source of comparison with the physical world. For example a head including 20 000 items includes not only hits but also niche content, especially if the offline market is taken as a reference, this definition of the head does not have any real purpose: in the offline market, the total of titles might be even less than that.

On the other hand, when the offline market is taken as a reference, the stock of a major brick-and-mortar store is taken as the head. Although this might seem to to provide a good cut-point to measure the impact of internet sales, the problem is that if the stock of the biggest store increases then the head automatically should increase, without any other justification.

Moreover, when Anderson defines head in reference to the biggest record store, he is using the number of CD's carried-out by the store, not the actual number of CD's selling. Head should be measured and defined in terms of titles in the distribution, not in terms of CD's offered, since the long tail plots sales, not titles offered. Moreover, should not the head be defined in terms of the average store? If the 99.9% of stores offer 2000 titles and just the remaining 0,01% offers 10 000, then the market should not be measured in terms of the exceptional case(s).

On this matter there seems to be not just one-size-fits-all answer, both percentages and absolute numbers (top titles) provide different information. Even Anderson uses different rank positions to define the head at different parts of his book: at some points it is the stock offered by

Wallmart, one of the biggest stores in the United States (which he says is around 5000 titles), at other points it is the top 1000 (for example in page 136).

Since this thesis departs mainly from what Anderson has written, the cut-off points that he uses will be used as the starting point for analysis. These points have the usefulness of providing, despite their subjective nature, a reference point with the physical world. Important points for comparison will be the top 1000, the top 2000, the top 5000 and the top 10 000, in this way different head definitions can be tested. Clearly this approach has limitations, therefore different percentages will be used as well, in order to describe more closely the shape of the distribution.

4.2.1 Defining head and tail from a mathematical perspective: a model for analyzing long tails

So far, the boundaries between head and tail have been defined according to somewhat subjective criteria, the size of a brick-and-mortar retailer or a percentage of titles. However, a model for analyzing long tails has been proposed by Kilkki (2007) that might provide the basis for defining head and tail by means of a single mathematical formula:

$$F(x) = \frac{\beta}{\left(\frac{N_{50}}{x}\right)^{\alpha} + 1}$$

Where:

F(x) = the share of total volume covered by objects up to rank x
N_{50} = the number of objects that cover half of the whole volume
α= the factor that defines the form of the function
β= total volume

This equation models cumulative distribution of sales and the distribution is plotted using a logarithmic scale in the x-axis. According to Kilkki, given the similarities that different long tailed distributions have, they can be modeled using this single equation. This model also has the advantage that it allows to approximate the shape of the curve by having a limited number of data points.

According to this author, this formula provides a better approximation for modeling long tails than the power curve (used in the examples above) especially in the case of distributions with a short range of x-values. The main requirement for the data is that "the cumulative distribution should generate a smooth S–shape when the x–axis is logarithmic." (Kilkki, 2007). As an illustration of the model, figure 4.5 shows data from Last.fm artists' rank for one month:

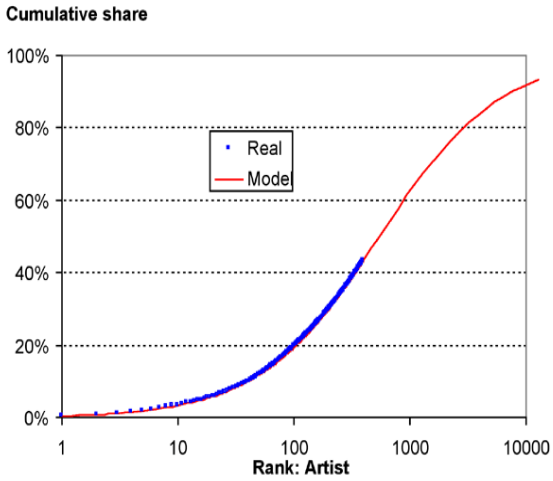


Figure 4.5 Top artists for the week ending 3 September 2006 at last.fm

Source: Kilkki.2007 "A practical Model for analyzing long tails"

This curve can be described using two parameters, steepness and the turning point. In this model steepness is defined by α (having the same role as the power exponent in a power law function) and the turning point occurs at N_{50}.

Kilkki defines three parts where the curve can be divided: base of the tail, middle of the tail and the end of the tail. The "base" would be the head according to the terminology used in this thesis and the tail would be divided into middle and end.

These three points are defined in the following formal terms:

The boundary between base and middle:

$$x_{bm} = N_{50}^{2/3}$$

The boundary between middle and tail:

$$x_{me} = x_{bm}^{2} = N_{50}^{4/3}$$

Therefore, if the 50% of sales occurred at rank position 1000 then the head would be formed from titles 1-100, middle from 101 to 10000 and the end from 10001 onwards. These points separate the logarithmic graph in three similar parts.

Although Kilkki does not mention these points as a way to compare different distributions, it could be used as a way to relate different markets. By using the 50% of sales as a common point, different markets can be compared. Two different measurements can be used: the percentage of titles that falls in each of the three different parts (head, middle, end) and the market share that falls into each part of the curve. Also the changes within one single market

could be described using this methodology, the number of titles that fall under head and tail could be used as a measurement of the shift from hits to niches.

This definition of head and tail has the advantage that follows the logic of the curve since N_{50} represents the turning point of the curve and it has an equation behind it that can be used to model the whole distribution. On the other hand, it has the disadvantage that might provide results that are difficult to relate with the common usage of the words "hits" and "head", for example a head or base consisting of 20 titles only. Nevertheless, these results will be consistent with the shape of the curve and will provide an objective way to define head and tail.

Kilkki also comments on the importance of the α value from a business perspective, according to the author, if α is small (<0,6) the market could be divided into two distinct areas, a limited number of superstars and a very large number of niches, while with a larger $\alpha\,(>0{,}8)$ "the clear majority of the overall business is in the middle of the tail, while the ends of the tail do not provide clear separate business potential" (Kilkki, 2007).

This model is a first attempt to provide with a practical tool for analyzing long tailed distributions. To test the limitations of this model escapes from the scope of this thesis, however, this section aimed to provide an alternative view for how head and tail could be defined without relying on subjective criteria or reference to the physical market. Furthermore, more attention should be devoted to models such as the one presented in this section, which might provide with a practical tool to make estimations and even predictions (see for example the use of β as an indicator of latent demand in Kilkki's article).

4.3 Possible Research Designs

As it was described before, three types of outcomes can be considered when defining what is a "tail-effect": more variety of titles consumed, a shift in the distribution of sales between head and tail, and different titles being sold online-offline. The first two effects can be measured by studying sales' concentration, the last one calls for a study of the composition of both markets in terms of head and tail titles and the respective differences between them.

4.3.1 Measurement of sales' concentration

As it was described before, there are differences between hybrid retailers and pure digital retailers. These differences give the possibility to measure increases in stock variety for on one hand *digital products,* namely mp3 files, the most popular digital audio format in the market, and on the other hand *digital distribution of physical CD's.* For this reason the current market allows research to be done on two main types of retailers: hybrid and digital, and on different aggregation levels: overall

industry and individual retailers. Different research methods can be followed by combining these types of retailers and the level of aggregation.

There are three main research designs possible for measuring sales' concentration:

- Evolution of sales' concentration through time for a specific market.
- Comparison between different distribution channels or different markets at a point in time
- A comparison of different distribution channels/markets at different points in time.

4.3.1.1 Evolution of sales' concentration through time for a specific market:

In this case, the evolution towards a more/less concentrated distribution could be studied. As mentioned before, this can be done at different levels.

a) Evolution of the concentration in sales in the overall industry: Since the number of retailers selling CD's online has increased during the last years, it could be studied if the concentration levels towards hits has decreased through the years. On the other hand, the number of available songs in digital format has increased through the years, and the share of e-tailers has been growing steadily, as more titles become available and more people buy music online, studying the evolution of the digital market could show evidence of changes in concentration through time.

b) Evolution of the concentration in sales for one retailer: This could be done for a CD retailer that has recently ventured (as most retailers have) in the selling of CDs via their own websites. The differences in concentration before and after the online distribution in case of being present, would show the effects of digital distribution on the pattern followed by sales. Moreover, the effect that the incremental variety on stock has on the concentration of sales towards either hits or niche content could be studied .

4.3.1.2 Comparison between different distribution channels or different markets at a point in time

In this case, the data will comprise a single period in time, for example a year, and the differences in concentration will be compared. It could be further divided into online/offline distribution of physical content (CD's), then three areas can be compared: offline market, online distribution of physical goods and the market for digital goods.

c) Comparison between overall (national) mp3 sales and sales of physical CDs: According to the Long Tail theory, digital products will have more possibilities of profiting from the tail, due to the low distribution and storage costs. Comparing overall data will provide the clearest picture of whether or not this has happened in the music industry. Depending on the level of aggregation of the data to be used, the study could be further divided into online and offline distribution of CD's.

d) Comparison between one or more online retailers and overall sales in the offline market: By using information from one retailer, results cannot be generalized to the overall market. However, it can provide valuable information as a case study. In order make the best possible analysis, the online retailer to be chosen should have the characteristics that would favor a long tailed effect according to Anderson: a catalog that surpasses the one of an offline retailer and has a good recommendation tool system. This study can be made for an online retailer selling CD's, or a retailer selling mp3 files.

e) Comparison of sales for a retailer operating both online and offline: This approach won't test the hypothesis directly, rather, it would measure the effect of recommendation tools into purchases made by consumers. The long tail hypothesis will not be tested directly because such a retailer usually has the same stock online and offline and operates through physical stores, a limitation to the long tail hypothesis. In case of a retailer that works with different catalogs online and offline, it would be also possible to measure the effect of extended variety.

f) Comparison of two different retailers, one working online and the other offline: This is a similar approach as the one taken by Anderson, when he contrasted Amazon with other brick and mortar retailers. Again the study can be made for mp3 files and also online distribution.

4.3.1.3 A comparison of different distribution channels/markets at different points in time

In this case, the previous comparison methods can be made over a period of time. This will give a complementary view on how both markets have evolved. For example if concentration towards hits has decreased in a period of time for digital products while for physical products hits have remained stable, this will provide with evidence that the trend found in the digital market is characteristic of this market and not of the whole music industry.

4.3.2 Measurement of the differences in "tail" content online-offline

The previous measurements covered whether the distribution of titles has shifted favoring tail-titles, but only partially. It covers whether hits, defined either in absolute numbers (top titles) or percentages, have less market share. However, it does not study in depth the performance of tail titles and how important are they in a digital market. According to the analysis made above where three different ways to define the long tail were described, it would remain to be tested if the composition of sales is different online-offline. This test has to cover differences in hits between both markets, differences in tail content, and to study titles that were sold only digitally.

In case that the music industry in The Netherlands presents a long-tailed effect in sales, this would also affect aspects like market structure and composition of the head. For this reason, another measure that might complement the research of the long tail and will also help to explain how the long tail affects the market structure is by studying the market share captured by the top 4 record companies. Since main record labels have usually focused on "head"content, the degree on which the market share of small independent record companies changes could be a sign that a long tail effect is visible. It would also explain, in case of being indeed present, the effect the long tail could have had on the music market's structure.

Another angle of the long tail concerns the composition of the "head". Anderson says in his book : "Who knows, with good search and recommendations, a bottom 80 percent product could turn into a top 20 percent product" (Anderson, 2006: 132). As an example, he also describes how the book "Touching the void", a book that went unnoticed when released, became a hit a decade later in Amazon, thanks to recommendation systems and online reviews (15-16).

Therefore, another way to complement the study of the long tail is to analyze if hits are different for digital mp3 files than CD's. Moreover, if the internet favors "tail" titles, these titles should a have a stronger performance online, thus, a comparison using tail titles that were sold via both physical and digital formats will also provide information about the effect that the internet has into the music market.

4.4 Evaluation of the potential research designs

Ideally, a full study of the long tail theory would include more than one approach, preferably the overall industry should be included complemented with data from specific retailers as source of more detailed information. One single retailer either hybrid or digital will not be representative of the market and will not allow to make any generalization. Therefore, the overall industry will be the best measure of any change in sales' patterns. The use of a specific retailer will have the function of a "case study". For example two retailers having different sales patterns, one favoring "superstars" and the other "the long tail", can give valuable information about the reasons behind

the appearance of a long-tail or a superstar effect. In this way, certain specific aspects can be studied, aspects such as the role of recommendation tools or different marketing techniques.

Studying evolution through time will only give certain information about the long tail. In case of physical CD's, the information that will be obtained will be the effect of online distribution on sales, being online distribution a new phenomenon, changes in sales' concentration could be related to the growth of online distribution. On the other hand, with mp3 sales, being a single distribution channel, the effects that can be studied will be different. For example if the number of tracks available has increased, the effect of this increased variety on sales could be studied.

When it comes to comparisons, there are limitations in comparing physical sales with digital sales. Ideally, physical CD's should be compared with album downloads and singles with tracks downloads. However, singles have virtually disappeared from the market and it is likely that the remaining singles are only the hits, this makes it impossible to make a valid comparison between the two markets. On the other hand, physical CD's and album downloads can be compared, however, the volume of sales of digital albums is far less than sales from single tracks, making it a less appropriate measure of what happens online, especially in a market where digital albums are only recently being sold.

This would leave only the comparison between CD's and digital tracks, however, it is not possible to compare the sale of CD which comprise in average 10 songs with downloading a single track, not to mention the huge difference in price between them. There are possibilities for making comparisons, however, these biases will not fade away completely. Any comparison between digital and physical has to be complemented with additional measures, such as the ones described before. From these measures comparing albums remains to be the best option to overcome difference of content and price.

4.5 Methods to be used in this thesis

4.5.1 Measurement of differences in market concentration

For the purpose of this thesis, the main approach will be to study the industry as a whole, as it was discussed before, this approach will give an overview of the aggregate effect in the market. According to the methods described in section 4.3, the two main methods to study the industry will be:

a) *Evolution of the concentration in sales in the overall industry (physical)*: Overall sales of the industry will be compared and analyzed at different points in time. The main limitation that this method will face is the lack of extended data series for a long period of time.

In case of digital products there is a limitation to the number of years available for study, for example it was not until 2008 that download albums and mp3 single tracks appear as a separate item in the NVPI market reports. Therefore, this study will be focused on the market for CD's.

c) *Comparison between overall (national) mp3 sales and sales of physical CD's :* Since the digital market is where a long-tailed effect is predicted to be stronger, it is important to focus most of the attention into this market. In order to measure have a reference point, the comparison with the physical world will be made.

Given the difficulties of finding an adequate way to compare single mp3 tracks with physical CD's, a comparison between download albums and physical CD's will be used. Although it was mentioned before that there is an important difference in sales' volume between the two markets, this comparison will be useful to explore differences between the two markets, especially when it comes to differences in composition.

In this thesis, percentages will be used as the best measurement for markets of different size, since they provide with a better picture of the overall distribution of sales across different titles. The Gini coefficient will be used as an indicator of concentration and it will be complemented with a description of what happens at different percentiles as a way of detailing the differences at the top.

4.5.2 The Lorentz Curve and the Gini Coefficient

The Lorentz curve and the Gini coefficient have been widely used in different fields as a way to measure income and wealth distribution. It can also be used to point out the difference in concentration between the CD market and the digital market.

The Gini coefficient is an useful measurement for comparing the digital and physical market because scale and population can be of different sizes, it does not consider the size of the market and it does not matter how large a population is (Debraj, 1998:188). This allows to evaluate markets of different sizes such as the ones being discussed in this thesis. It gives a description of the overall shape of the distribution.

The Lorentz curve denotes the proportion of total income received by the poorest 100x% of the population, where $0 \leq x \leq 1$. It is an increasing function that ranges from 0 to 1. In other words it shows the cumulative percentage of wealth obtained by different percentiles of the population ordered from the poorest to the richest. The percentage of population is plotted on the x-axis and the percentage of income is plotted on the y-axis. In this thesis, the curve will denote the cumulative percentage of sales obtained by a percentage of all titles, arranged from the lowest

ranked to the top 1 selling title.

The Gini coefficient is a measure of the distance between the perfect inequality line and the Lorentz curve. It is obtained by calculating the area between the 45 degree curve and the Lorentz curve:

$$G = 1 - 2 \int_0^1 L(X) dX.$$

It ranges from 0 to 1 where a value of zero would mean a perfectly equal distribution of income in a giving population and one would be a population which has the maximum inequality, for example one case of the population having all the income. Therefore the closest to one the coefficient is, the highest the percentage of sales that falls in the hands of a few titles.

4.5.3 Measurement of the differences in "tail" content online-offline

The market share of record companies will be compared online and offline, for this purpose the Four-firm concentration ratio will be used. A lower concentration ratio in the digital market could be a sign of a bigger market share of tail-products online.

Since Anderson's book begins with an example of a tail-product that thanks to online distribution became a hit, another angle to be explored in this thesis will be the differences in the composition of the "head" online and offline. For this purpose, the top 200 digital and physical titles will be compared. This comparison will be made using the top 200 digital albums.

Finally a comprehensive analysis of differences in the composition of sales will be done for download albums and CD's. This analysis will cover titles sold only in digital format and the performance of "tail" CD's also sold in digital format (CD's selling less than 20 copies).

5

The market for CD's

5.1 Description of the data used

In order to explore the topic at hand, secondary data provided by different institutions was used. Two main sources were used: the NVPI and the GfK. Some additional data from Cosmox, an online retailer was used as a complementary source.

The NVPI is "the Dutch association for importers and importers of image and sound carriers" (www.nvpi.nl). Established in 1973, it represents approximately 85% of recording companies. Among the data provided by them are the overall sales figures of the market and the market share of record companies. Their data is the most comprehensive source in The Netherlands when it comes to the music market, covering most of the market.

The data used by the NVPI comes to a great extent from the GfK Group. The GfK or *Gesellschaft für Konsumforschung* (Society for Consumer Research) established in 1934, is one of the largest market research organizations in the world (http://www.gfk.com). They monitor retail sales for different markets. When it comes to the music market, their database comprises around 80% to 85% of the whole national dutch market and includes the major online retailers such as Bol.com or Itunes.nl.

The GfK provided with data-sets for CD sales and album downloads for the period 2005-2009. This data will be used to examine concentration levels for physical and digital products and the differences in composition between both markets.

The data provided comprised according to the GfK, all albums sales (both physical and digital) for full price and mid-price sales in The Netherlands, this data represents around 85% of their whole collected data. The only missing data is budget category, CD's being sold at a low price which represents the smallest share of the market.

The data used included information about: artist, title, units sold, record company, carrier and average sale price. Download albums were included in the lists and were differentiated by specifying the format with a different code. Each data set comprised more than 40 000 different titles, with some years selling more titles than others.

There is some important information missing. First, the specific retailer included in the data was not given, which might produce a biased perception either in favor or against the long tail hypothesis. If the 15% missing corresponds to online stores this could affect the long tail effect considerably.

Additionally, another limitation of the data is that having sales' figures does not provide with information on the total number of CD's available in the market, in other words how many CD's did not sell at all. This piece of information would have been particularly useful when it comes to e-tailers, since to a great extent the long tail hypothesis lies on the assumption that most of the niche content will be sold at least once, in a given period of time, therefore making it profitable.

In order to overcome this problem, it is important to have data on a specific e-tailer, in order to know the percentage of titles from its stock that are in fact selling at least one copy. Unfortunately, this information could not be obtained. A considerable number of e-tailers in The Netherlands were asked for cooperation, however, they refused to provide with this information on the basis of confidentiality reasons.

Nonetheless, some data was provided by Cosmox, an online store focusing on music, video and books. They provided with sales figures for the year 2008, including products and the percentage of revenue they made. This data will be used as an example of the profitability of the long tail hypothesis for e-tailers.

5.2 The CD market: the potential effects of online distribution

Online distribution has been growing steadily since 2001. Prior to 2001, online distribution was not even included into statistics. On 2001 this distribution channel appears for the first time with 2% of sales. It is not until 2005 that online distribution acquires a market share higher than 10%.

These e-tailers represent a hybrid type of retailer, one that can manage to offer more variety due to the lack of shelve spaces that a record stores requires, yet they have a limit to what they can offer, because of selling a physical product that has to be stored somewhere before being sold. Since the market share of these hybrid retailers has increased in the last years, it can be explored if this increase has brought with it, a change in the market share of hits and the overall concentration of the market.

By using the period 2005-2009, some potential signs of a long tail effect can be studied. Because of the limited number of years available for study, it is not possible to rule-out other explanations or to say that the results have statistical significance, for this to be possible, an extended number of observations would have been necessary, one that includes the market before and after online distribution began. Nonetheless, given the recent origin of online distribution in The Netherlands, it is worthwhile exploring if there is evidence of any potential trend in the

market. Moreover, some aspects such as the lengthening of the tail can be explored without the need for longer data points or statistical analysis.

5.2.1 The length of the tail from 2005-2009

The first aspect that can be studied is if the tail has grown longer due to the increase in market share experienced by online stores. Most online stores such as Bol or Cosmox have been in the market for already 5 years, which means that long tailed stocks has been available for consumers for already the period to be studied. Therefore an increase in the number of titles consumed would be a sign of a positive tail effect produced by more consumers buying online.

During the period analyzed there was not a sustained increase in the number of individual titles consumed, in fact 2006 had the highest number of different titles. In 2005 around 47000 different titles were sold, the number of titles increased significantly in 2006, that year there were more than 65000 titles in the list. However, the number of titles decreased again in 2007 to slightly more than 41000. In 2008 and 2009 the number of titles remained below the number of titles sold in 2005.

Table 5.1: *Number of unique titles sold 2005-2009*

	2005	2006	2007	2008	2009
# of unique titles sold	47007	aprox. 65000	41075	42020	41692

This data shows that there has not been an increase in the length of the tail during the last years in terms of more titles being consumed, at least within the data that was made available. Online sales have increased, however, this has not been reflected into more titles being bought.

Since a store such as bol.com existed before the period that has been chosen (it started in 2001) , it becomes an important question for future research to explore how did the appearance of such stores changed the market. To answer this question, data for the years before the appearance of these stores should be contrasted with the data presented here.

It can be concluded that considering the size of the stock of stores like Bol or Cosmox, there has been an important percentage of titles not being sold. Today, Bol offers 454646 titles in the music category. Although the number of titles available might have increased during the last year, it is very likely that their stock had more than 200000 items already in 2009. During 2009 there were around 40000 different titles reported in the GfK sales' data, this figure represent a 20% of all the titles available at Bol.com, this estimation could be even lower if their stock had more than 200000 titles already.

5.2.2. Market concentration in the CD's market from 2005-2009

A first look at the CD market shows no sign of the tail growing longer. Another way to explore if there is a long tail effect in this market is to see if sales' concentration has decreased in the last five years. A decrease in the market share of hits could be explained through the increase of CD sales made via online stores, this would mean a change in the shape of the distribution, since there is no evident increase in the length of the tail.

Table 5.2 illustrates the cumulative percentage of sales obtained at different percentiles:

Table 5.2: *Cumulative sales at different percentiles CD's 2005-2009*

% of titles	% of units sold 2005	% of units sold 2006	% of units sold 2007	% of units sold 2008	% of units sold 2009
1,00%	69,1826	60,7088	65,9189	65,6579	64,9000
5,00%	85,0307	82,5546	87,4341	86,8563	86,3608
10,00%	91,4300	89,1397	92,7738	92,5151	92,2557
20,00%	95,4800	94,3726	96,4353	96,3751	96,3238
25,00%	96,5588	95,7178	97,3145	97,2813	97,2677
50,00%	98,8393	99,8488	99,2326	99,2045	99,2208

When looking at the market sales' concentration, the first result that comes to the surface is how concentrated the market for CD's is. For all five years, a 10% of titles fetched around 90% of all units sold. That is a market with higher concentration than what a market with a Pareto distribution would have, in this case the market shows a 90/10 distribution, or a 95/20 distribution if the 20% of titles is used as it is used in the Pareto distribution.

Even a more striking result is that half of all titles accounted in average for only 1% of sales, which is a sign of a very long and extremely flat tail. Moreover, there has been no sign of change during the last five years, although online sales of CD's have been gaining importance as a distribution channel. Figure 5.1 shows the cumulative percentage of sales obtained by a specific percentage of titles, it also shows the best selling 8% of titles, where most sales occur.

As it can be observed in the graph, 2005 and 2006 occupy the extremes with respect to concentration, from 2005 to 2006 there was a marked decrease in concentration in the market, however, the market became again more concentrated in 2007 and has remained at a similar level since then.

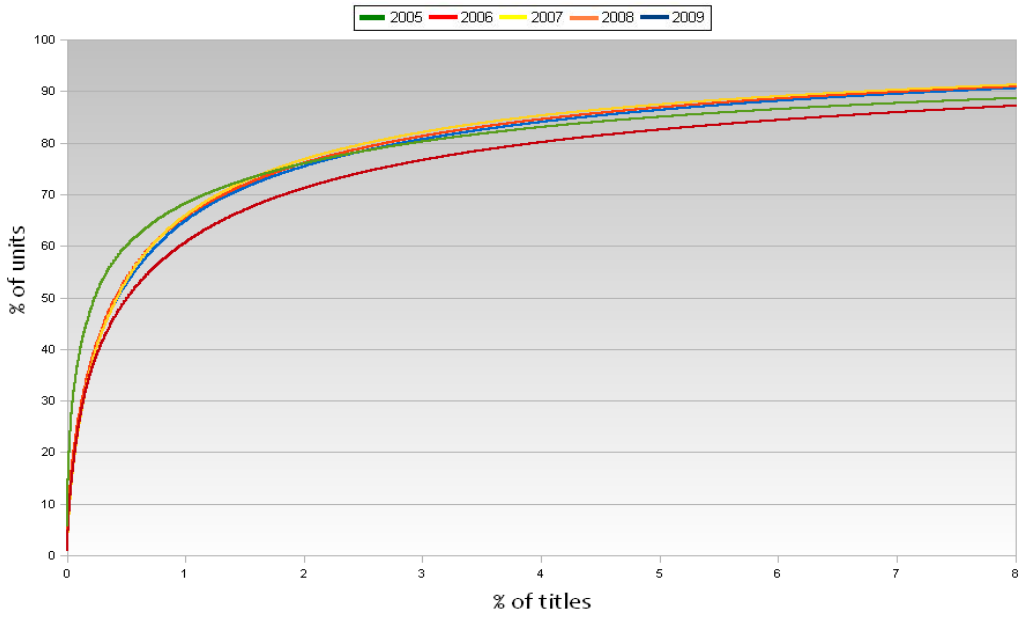


Figure 5.1: *Cumulative sales per percentage of units*

Table 5.3: Cumulative sales at different ranking levels

Rank	Sales 2005	Sales 2006	Sales 2007	Sales 2008	Sales 2009
1	5,74	1,02	1,62	1,76	1,58
10	20,93	7,24	9,61	12,09	10,74
20	28,45	12,21	16,05	18,22	16,86
50	39,1	21,48	27,65	29,2	28,1
100	48,45	31,38	39,75	40,1	39,49
200	58,04	41,8	52,73	52,74	51,93
500	68,95	56,4	69,18	68,35	67,88
1000	76,66	67,25	79,42	78,4	77,86
2000	83,61	76,88	87,19	86,37	85,94
5000	91,95	86,74	93,97	93,64	93,48

The level of concentration can also be observed in table 5.3 when the top titles are used, with the exception of year 2006, the top 200 accounted for more than half of all sales. When the 5000 position is reached, a 90% of all sales have been made and the remaining titles add only 10% of sales. The data sets contained approximately 40 000 different titles, which means that more than 30 000 titles are adding only 10% of all sales. The tail is therefore very long and extremely flat. The results are consistent with the results observed by using percentages, the same levels of concentration occur with 2005 being the most concentrated year and 2006 the least concentrated year. The number of titles sold in 2006 clearly affected these results, however, the overall picture shows no visible trend towards less concentration.

Anderson uses the sales that come outside the stock of a brick-and-mortar record store as a way to support the profitability of the long tail. He defines this threshold at 5000 CD's. If this approach is applied to the dutch market, during the last three years approximately 7% of sales come from items only available online.

Despite the presence of online stores offering more variety than before, consumers seem not to be fattening the tail. Again, no visible trend appears when we observe the graph with the percentage of sales reached by the top 500 titles, there seems to be no tendency towards either more concentration or less concentration. In figure 5.2 the percentage of sales added by each additional offering is shown, in other words, the tail of the market.

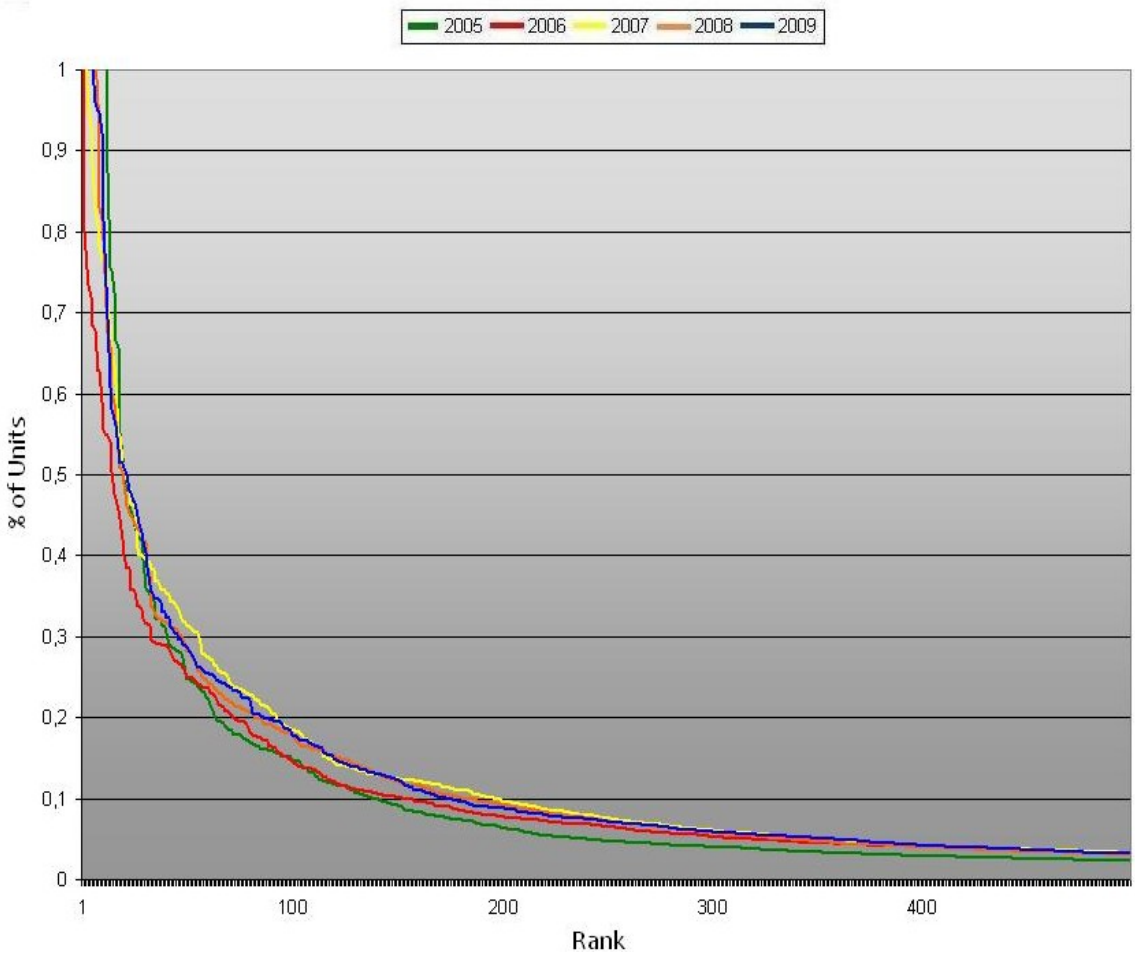


Figure 5.2: *CD Sales per title - top 500*

Either using percentages or using top titles, there is no evidence of a decrease in concentration in the market, concentration changes without a pattern from one year to the next. The results are consistent using both methods which shows that in this case top titles are also a good measurement of the market. Therefore, it can be said that the increase of CD's online distribution *has not had any visible effect on the market for CD's during the last five years*.

Nonetheless it is important to notice that most brick-and-mortar stores carry out less than 5000 different titles. The fact that all five data sets contain around 40 000 different titles is a potential sign of the increase in variety that online distribution has brought. Although the long tail has not had a strong impact in terms of market share in the last 5 years, it is clearly visible in the data. What can be seen in the data is that there has not been a significant decrease in concentration in the market, there is a long tail in the market, but it has not decreased the market share of hits, either in absolute numbers or in percentages. Moreover, the length of the tail has not increased during the last five years. From this analysis no evidence of a growing tail can be found.

5.3 Case study: Cosmox

One of the limitations of the data used to analyze the CD market is that it included both brick-and-mortar stores and e-tailers into a single list. Therefore, it is not possible to compare the differences between them. Contrasting sales figures from one e-tailer with the data of a brick-and-mortar retailer or with industry as a whole would have been a way to overcome this limitation, despite the fact that the findings of one retailer cannot be generalized.

Cosmox, an online store selling CD's, provided with some data that could evidence differences between them and the rest of the industry. The data provided by this company differed from the data used from the GfK. The data they provided does not include the number of units sold by each title but the percentage of revenue made by each title. For this reason the data is not directly comparable with the one given by the GfK. Nonetheless it provides some complementary information about one aspect that has not been discussed in the previous section: *how might an e-tailer profit from the tail.*

5.3.1 About Cosmox

Cosmox started to operate in 2006 selling music, books and videos online. Table 5.4 shows the total number of titles available in June 2010, according to the information available through their website:

Table 5.4: *Cosmox inventory in 2010*

Categories	# of titles	% of titles
Rock & Metal	107027	26,8389%
Pop	92744	23,2572%
Jazz, Blues, Country & Folk	84496	21,1888%
klassiek Cd's	80294	20,1350%
Dance, House, Hip Hop & Rap	51129	12,8215%
Wereldmuziek	42949	10,7702%
Musical, Soundtrack & Cabaret Cd's	10588	2,6551%
R&b & Soul Cd's	7372	1,8487%
Amusement Cd's	2255	0,5655%
verzamel Cd's	216	0,0542%
Total CD's	**398776**	100%

In total, almost 400 000 different titles are available today. Rock and Metal music are the category with more titles available, followed by Pop music. In general terms, the distribution between categories is relatively equal for the first 4 categories. What makes this catalog different from a regular brick-and-mortar store, besides the magnitude of its offering, is the importance of genres such as blues, classical music and world music, genres that in previous times were left to specialized stores.

The website has a recommendation system that shows information about other titles in the same genre, what other customers bought, reviews from customers and other titles from the same artist. Since the system works based on sales' information, popular items have more reviews and suggestions than obscure items, therefore, their system might also reinforce hits' consumption. Nonetheless surfing through the website can also lead to niche and obscure titles.

5.3.2 About the data used

The data used in this thesis comes from the year 2008, therefore the current amount of titles offered cannot be assumed to have existed at that point. The total number of available offerings for that year was not provided, only titles that sold at least one copy appear on the data. Therefore the percentage of titles that did not sell a copy is unknown. It can be assumed without compromising the results that the catalog was smaller in 2008.

Nonetheless, the data from 2008 contains slightly more that 28000 unique titles, which is still more of what a regular store would have been able to keep in stock. The data provided the following information: rank, rank as a percentage of total offerings, percentage of revenue per title and cumulative revenue per rank position.

Because of the information available, the long tail of Cosmox cannot be calculated as it was calculated for the whole industry. What the data will provide is some evidence of *how profitable* the long tail has been for this e-tailer.

5.3.3 Results

Figure 15 shows Cosmox's revenues for 2008, both titles and revenues are displayed as percentages. The cumulative percentage of titles is plotted on the x-axis and the cumulative percentage of revenues is plotted on the y-axis.

The first important result that can be drawn from the graph is that the Pareto principle do not hold for Cosmox's case: 20% of products do not earn 80%. In Cosmox's case, 20% of products accounted for 52,29%, this occurs at ranking position 5669, 80% of revenues was made by 52,63% of titles (position 14921). The Pareto rule would in this case shift to a 80/52 relationship.

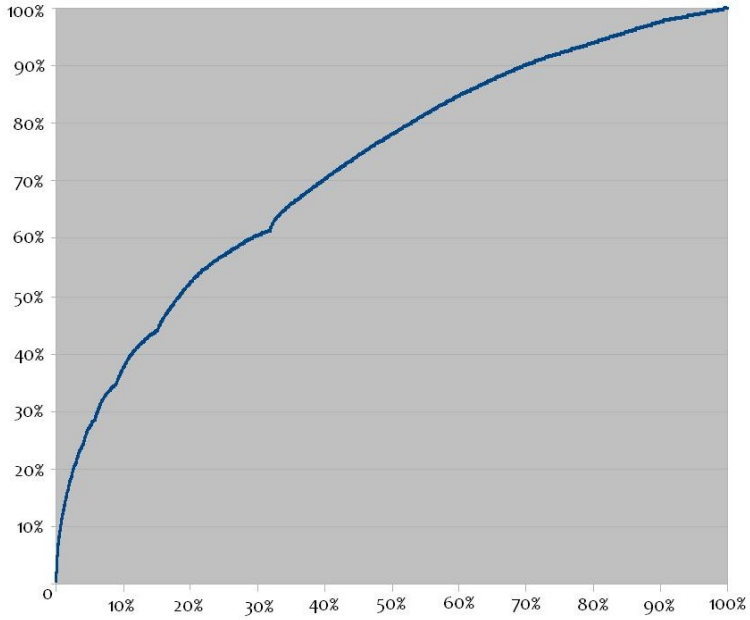


Figure 5.3: *Cosmox's revenues 2008*

Another way to analyze this data is to see how much percentage of revenue comes from items that would possibly be outside the reach of a brick-and-mortar store. Anderson sets that limit in his book at 5000 unique offerings. If this is the case, 51,07% of revenue would come from content that is only available at Cosmox. The following table provides with more details about the cumulative percentage of sales at different ranking levels:

Table 5.5: *Cumulative revenues obtained at different ranking levels*

Rank	% of total items	Cumulative % of revenue
1000	3,5278%	23,1464%
2000	7,0557%	32,1532%
5000	17,6392%	48,9282%
10000	35,2783%	66,2476%
15000	52,9175%	80,1903%
20000	70,5567%	90,4078%

It can be seen that at rank position 10 000 there is still an important fraction of revenue to be gained, items after this position accounted for 33,75% of sales. The same could be said if we continue further down the rank, at rank position 20 000, there is a 10% of revenues that were made by items below that level. Revenues decrease but still tail titles keep adding an important part of total revenue.

In order to see if this case is different from what occurred in the rest of the industry, a comparison can be made with the overall data of sales for this year. As mentioned before, the data provided by the GfK did not provide revenue made per title, however it did provide with average

sale price for each item and the number of units sold, based on these data, an average revenue per title could be calculated.

There are many limitations that have to be considered when looking at the results coming out of this comparison. Since there is no information about how this average price was calculated, it is not possible to provide any further proof of its validity. It has to be mentioned that average prices ranged from 7 euros to prices over 200 euros. Since not only individual CD's are included in the list but also collections, box-sets and the like, prices that appear to be too high were still included in the calculation, without any additional information there is no criteria to set an upper limit (in total 40 titles had an average price above 100 euro). Moreover, there were 77 titles without any price indication and they were not used in the calculation, all of them coming from the lowest selling titles.

Furthermore, Cosmox is also included in the overall statistics, so there is an issue of double counting. In addition, as it was described above, the data includes other stores such as Bol.com, making it not a real comparison between an online retailer and brick-and-mortar stores. As mentioned in chapter 3, the market share of these stores was of 17% in 2008.

Since the data set for CD sales in 2008 includes over 40 000 items, there is an important size difference between the two data sets. However, at ranking position 28 346, point where the sales of Cosmox stop, the cumulative average revenue of the CD market is of 99,6285% , which leaves only 0,3715% of additional revenue for the remaining of titles. Therefore, the analysis could be made focusing on the first 28 346 titles, making both data sets of an equal size.The comparison was made using the cumulative revenue at each ranking point from the best selling title downwards in the x-axis.

Figure 5.4 shows that there is a clear difference in how much revenue is made at each rank level. The most dramatic difference can be seen at the top of the rank, already in the range from top 1 to top 1000. The top 1000 accounted for 78,08% while in the Cosmox's case the same number of titles accounted for only a 23,15%. In the overall CD market, already at top 4000, 90% of revenue has been made. If the average price used reflects the situation of the market, the tail in the overall CD market, is as concentrated in terms of revenues as it is in terms of units sold (the top 4000 accounted for 92,17% of units sold in 2008).

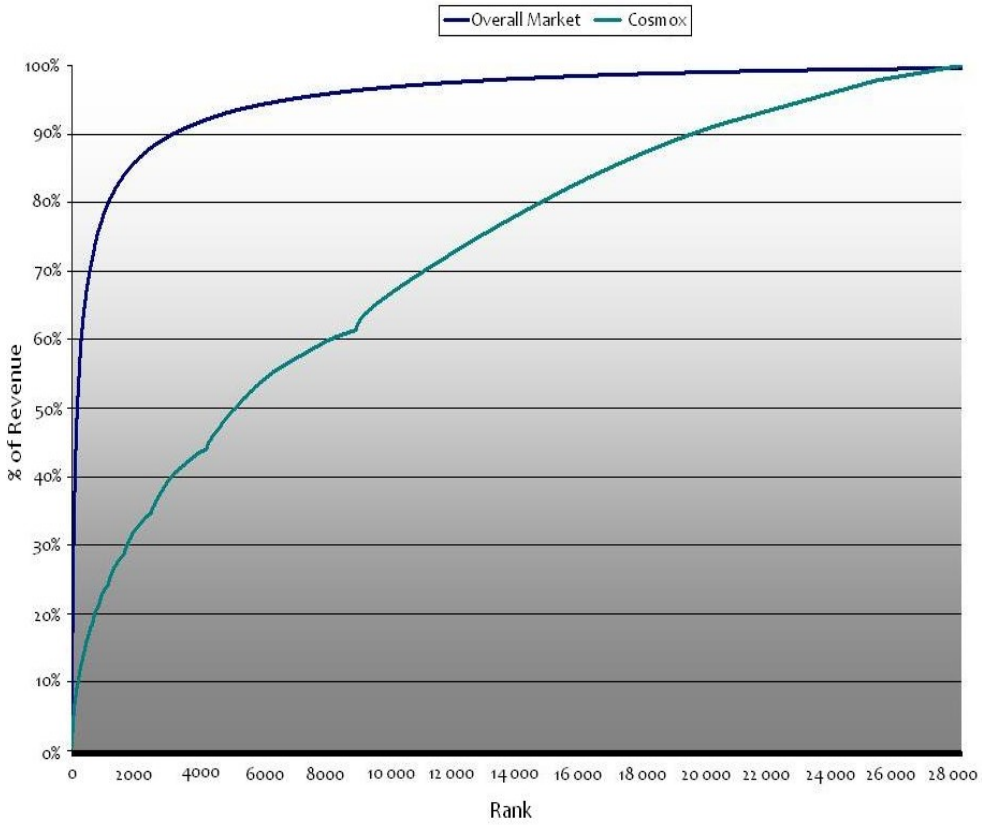


Figure 5.4: *Comparison of cumulative revenues per titles Cosmox and the Overall CD Market*

Table 5.6: *Comparison of cumulative revenues per titles Cosmox and the Overall CD Market*

Rank	Overall Market	Cosmox
Top 1000	78,08%	23,15%
Top 2000	85,85%	32,15%
Top 4000	91,63%	43,31%
Top 6000	94,28%	53,71%
Top 8000	95,79%	59,49%
Top 10 000	96,79%	66,19%

In Cosmox's case, having less concentrated revenues does not provide evidence that sales in units are less concentrated as well. It could also be a sign that titles outside the head are adding a significant amount of revenues despite the fact that they are selling less units. Prices are not necessarily equal for all items, this can be seen if a graph is plotted using revenues on the y-axis instead of units.

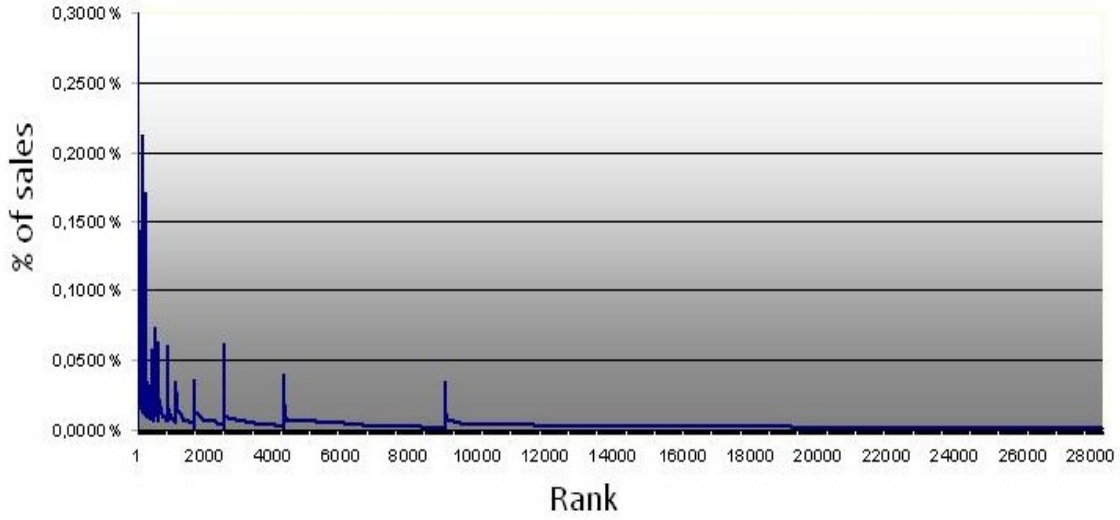


Figure 5.5: *Cosmox's revenue per title*

The curve does not have a uniform shape, it has peaks at different points, these peaks show items that earned more revenue than other items ranked higher. Most of the oscillation occurs at the high end of the curve, within the first thousand items. However, there are various prominent peaks further down into the distribution. The importance of these peaks can be observed if we zoom-in into different parts of the curve.

Figure 5.6 shows the the range from the top 2000 to the top 4000. What is interesting about it is that revenues for an item ranked in the position 3000 are higher than revenues for items ranked between 2000 and 2500. There is also evident a peak near position 2500, this title earned more revenue than most of all other titles in the ranking.

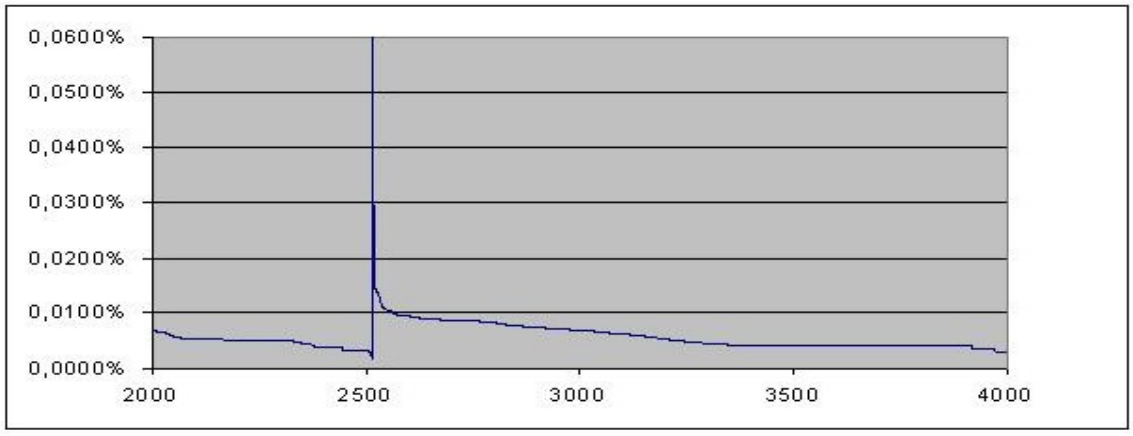


Figure 5.6: *Items ranked from position 2000 to position 4000*

Figure 5.7 plots titles ranked between positions 5000 and 10 000. A look further down into the ranking shows that revenues remain at a similar level even after titles ranked at position 10000. If we consider items beyond the top 5000 to be niche or tail albums, then for Cosmox, it was as profitable to have in stock tail albums as hit albums.

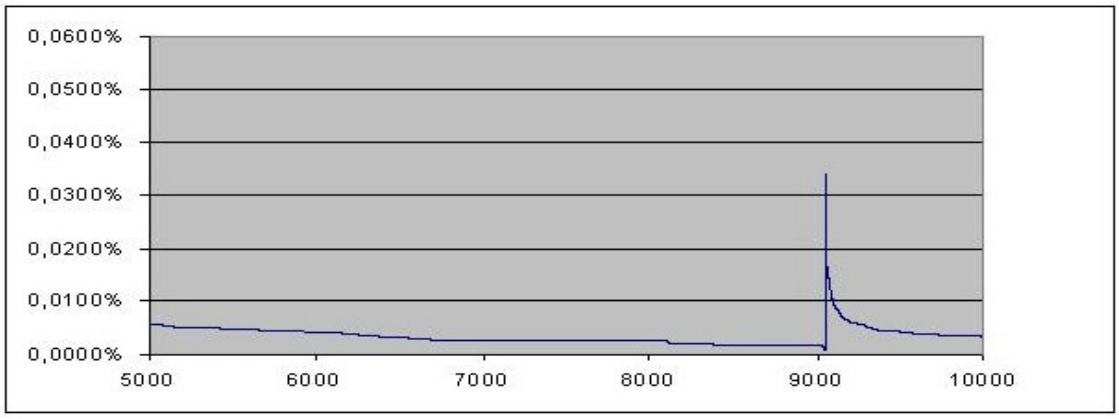


Figure 5.7: Items ranked from position 5000 to position 10 000

If a brick-and-mortar store is taken as a reference point to define the tail (it is assumed that such store carries 5000 different titles), then more than 51% of revenue for Cosmox came from tail albums. Still if the head is moved to include items up to the top 10000, revenues from the additional titles available at Cosmox accounted for 33,76% of the total revenue.

5.3.4 A pricing strategy of e-tailers?

A possible explanation for these results is that some consumers might be willing to pay a higher price for an item that suits their taste. An item ranked beyond a certain position is likely to appeal to a niche audience, therefore a peak point as the one visible around position 9000 shows that although the album sold substantially less copies than the top titles, it appealed enough to a group of consumers in order to make them pay a higher price than the price charged for most CD's.

Therefore, a strategy for e-tailers could be to charge higher prices at the end of the tail, since consumers have a higher willingness to pay for the product. This is the opposite to what Anderson propose in his book, he describes music as a *want* market where lower prices might encourage consumers to try new artists. In contrast these results suggest that charging higher prices in the tail might be more profitable, since many consumers looking for albums know what they want and they are willing to pay for it.

It could be argued that customer's buying tail albums from Cosmox were not driven by recommendation tools, but they were already acquainted with the titles they bought. Given the higher risk posed by unknown/unpopular artists, it is very unlikely that consumers were willing to

pay more for them. Then the question becomes how to drive demand down the tail without discouraging consumers with higher prices. It seems that there would be a trade-off between attracting new consumers and losing profits from consumers willing to buy at a higher price.

It would important to make further distinctions between new niche artists and back catalog from popular or niche artists. Back catalog could be charged higher than albums by completely new artists, since it could be assumed that these titles would already have a fan base. This topic raises interesting questions for further research.

5.4 Can it be spoken of brick-and-mortar retailers in 2010?

When Anderson wrote "The Long Tail" he described a very clear cut division of the type of stores operating in the music market: brick-and-mortar, hybrid stores, and purely digital stores. Brick-and-mortar stores sold CD's in physical stores, hybrid stores sold physical items via online websites and digital stores sold digital products such as mp3 files. Four years have gone by and the music market in the Netherlands does not seem to fit into this division anymore.

Today, it is virtually impossible to speak about offline sales since the most important CD stores of the Netherlands have adopted an intermediate model of business that is between brick-and-mortar and e-tailer. These stores still operate and obtain a big share of their sales via traditional stores, however, they have a much bigger assortment of products that they make available through their website.

For example Van Leest, a record shop with more than 20 stores in The Netherlands displays 249 191 titles in their website. These records could be shipped to the customer's address or can be picked-up at one of the physical stores without having to pay sending costs.

Another example of a record store venturing into online sales is Free Record Shop. This chain of CD stores has 183 stores across The Netherlands. In their website there are currently 237 553 titles available. Just like in the case of Van Leest, the CD can be shipped to the customer's address or picked-up at one of the chain's stores. These two stocks still sell approximately half of the assortment shown by Cosmox or Bol, however, their stock is clearly many times bigger than the sizes used by Anderson.

With the main stores becoming a mix of physical store and web store, the black and white division made by Anderson might be out of context in the near future, maybe even today. As the share of online sales increases and more stores adopt this mixed format, measuring the difference between brick-and-mortar and online stores will lose its meaning. The difference between both distribution channels could be still measured, but the long tail theory seems to have been adopted retailers in The Netherlands, which suggests that for e-tailers such as Cosmox or Bol, it has proved profitable.

5.4 Summary of the results

CD sales via online channels has grown up to a 20% of total sales in the last years and promise to grow more in the coming years. However, the tail has not been growing during the last five years and the market continues to be highly concentrated towards hits. Contrary to what the long tail hypothesis would claim, the middle of the demand curve has not gained market share. Increased variety has not decreased the importance of hits during these 5 years. Moreover, there seems to be an important number of titles not selling at all, which contradicts the idea that consumers will be matched with new albums. Regarding the length of the tail and the concentration in the market, there was no evidence found that online distribution has had an effect in the market during the last five years.

However, the long tail hypothesis also suggest that e-tailers can obtain a substantial part of their revenues from the tail, in the case of Cosmox, this outcome was observed in their sales from 2008. It seems that at least for specific retailers, the tail could be an important source of revenues. It also suggests that it could be possible to obtain profit from the tail by charging even higher prices for this titles than for the best-sellers.

6

Differences between the digital and the physical market in 2009

6.1 Description of data chosen for comparison

After having studied the effect of online distribution on the CD market, the following step in order to explore the market is to compare a purely digital market with the market for CD's. For this comparison, the year 2009 was used, this year had the largest number of digital albums sold. In fact, before 2008, there is no information about download albums. The market for digital albums in The Netherlands seems to be of a recent origin, at least that can be inferred from the data given by the GfK. It is only in 2008 when digital albums began to appear in the data (in 2007 there was only 1 album reported). Although the market has grown in the last two years, still, the number of unique titles sold in digital format is by far smaller than the number of unique titles sold via CD. In terms of unique titles digital albums are only a 5% (approximately 2000 titles) of the total of unique titles sold offline.

Although it is very unlikely that there was not a single digital album sold before 2008, since the data includes, according to the GfK, the most important retailers in The Netherlands, it could be assumed that the number of titles sold was, if existent, very small. Either way, according to the year reports of the NVPI download albums' sales have increased year by year, which makes 2009 the best example of the market yet, since it has the highest number of sales.

6.1.1 Limitations of comparing downloaded albums and CD sales

The fact that the size between both markets is so different, sets a limit to the validity of the conclusions that can be drawn for the online market. As the online market grows, the tail might become longer, or the concentration towards hits might change. However, the market can only be judged for what it is today: an emerging market that can grow into any direction.

It is important to underline that download albums are still a fraction of the whole digital market, therefore, the real shape of the digital market can only be seen if both albums and tracks are included. Yet, download albums can be taken as representation of the trend in the overall market, as long as it is taken into consideration that the market for mp3 single tracks and the market for download albums can have differences in terms of variety available and in terms of

consumer preferences.

Furthermore, the data used in case of the CD market includes the units sold via online stores. This means that the comparison being made includes already a physical market that might present a long tail effect due to these online sales. The fact that the list contained nearly 40 000 different titles already shows a market that includes more titles than a typical brick-and-mortar store will have.

6.1.2 Is there a long tail for the digital albums market?

The first impression that comes from the data, is that for download albums there is no visible long tail since the size of the market is only a fraction of the CD market. Since it is online where variety can be increased with virtually no limit, this shorter data set for download albums seems to be in contradiction with the hypothesis.

At this point in time, it cannot be said that *thousands* of niche albums in the tail can be aggregated to become a contender for hits and a profitable model for e-tailers. The main hypothesis used by Anderson to describe the relevance of the long tail has *not materialized in the market for digital albums yet*. The market for download albums is still too small to fulfill that prediction. The reduced size of the data shows that although variety of offerings is assumed to be larger online, the majority of those albums has not been sold. With the limited number of consumers buying digital albums at the moment, a big portion of digital albums are not being matched with consumers.

Nonetheless, the array of products available online might also have the effect of weakening the importance of hits and the effect of increasing the possibility of niche content to become sold. As it was described in the previous chapter the definition of long tail *should go* beyond market size. Long tail effect should be also understood in terms of the importance of niche content, differences in artists being sold and concentration of the market.

6.2 Market concentration: Download albums and CD's in 2009

6.2.1 Distribution of sales across the curve

As it was explained in chapter 4, comparing markets of different size is more difficult than comparing one market over time, the use of percentages or top titles might provide a different result. With markets that are so different in size, and particularly with one being so small, the use of hits does not provide the best description of how concentrated sales really are along the *whole* distribution. What is important to describe is the concentration of the market in a way that normalizes the differences in size and volume between them.

Both markets' distributions are plotted in figure 6.1. It can be seen how not only the top ranked titles add a bigger share of the market, but overall across the curve, titles added more sales in digital format than in the CD sales' curve (the curves cross only after approximately rank position 1400.) In other words, both head and tail are adding more across the market.

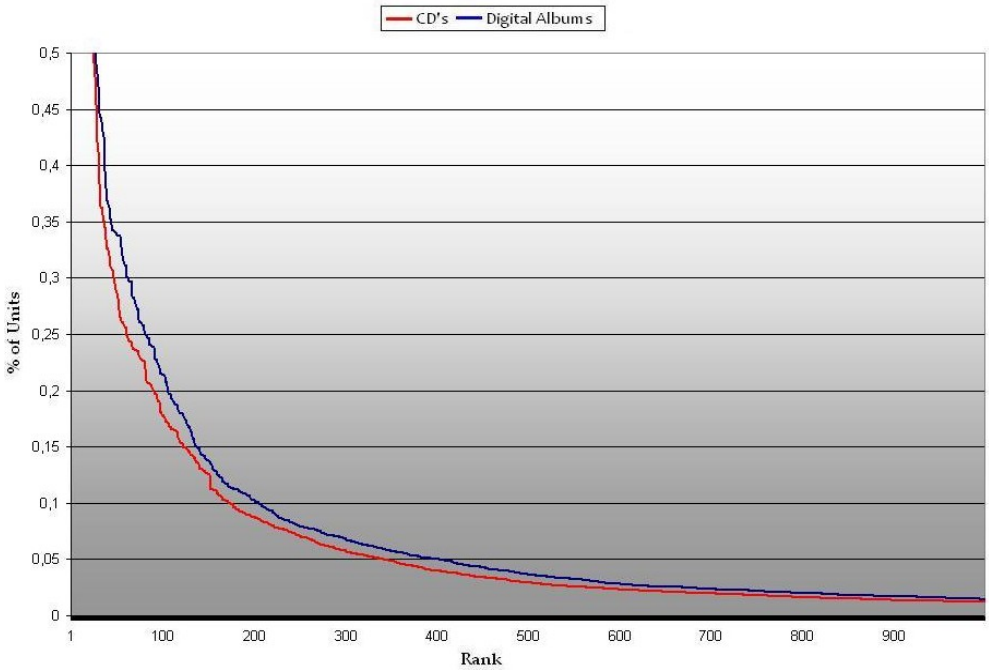


Figure 6.1: Sales percentage per title CD's and Digital Albums

In order to get a more detailed picture of the distribution of sales across the whole data, it is useful to describe the units sold as a percentage of total offerings. Table 6.1 shows the cumulative percentage of sales at different percentage of titles:

Table 6.1: *Cumulative percentage of sales per percentile of items*

% of items	Physical 2009	Digital 2009
1,00%	64,9000	20,9909
5,00%	86,3608	47,2851
10,00%	92,2557	61,6855
20,00%	96,3238	76,0308
50,00%	99,2208	92,6179
80,00%	99,8364	98,8379

The fact that the market for CD's presents a more skewed curve can be seen in the difference that 1% of titles accounted for. While in the digital market 1% of titles represented around 21% of sales, in the case of CD's, 1% of titles represented almost a 65% of sales. If the 20% of items is observed, it shows that there is only less than a 5% of sales done by 80% of titles

in case of CD's, while in case of download albums this percentage is 25%.

Figure 6.2 and table 6.2 show the sales' percentage obtained by different percentiles in both markets:

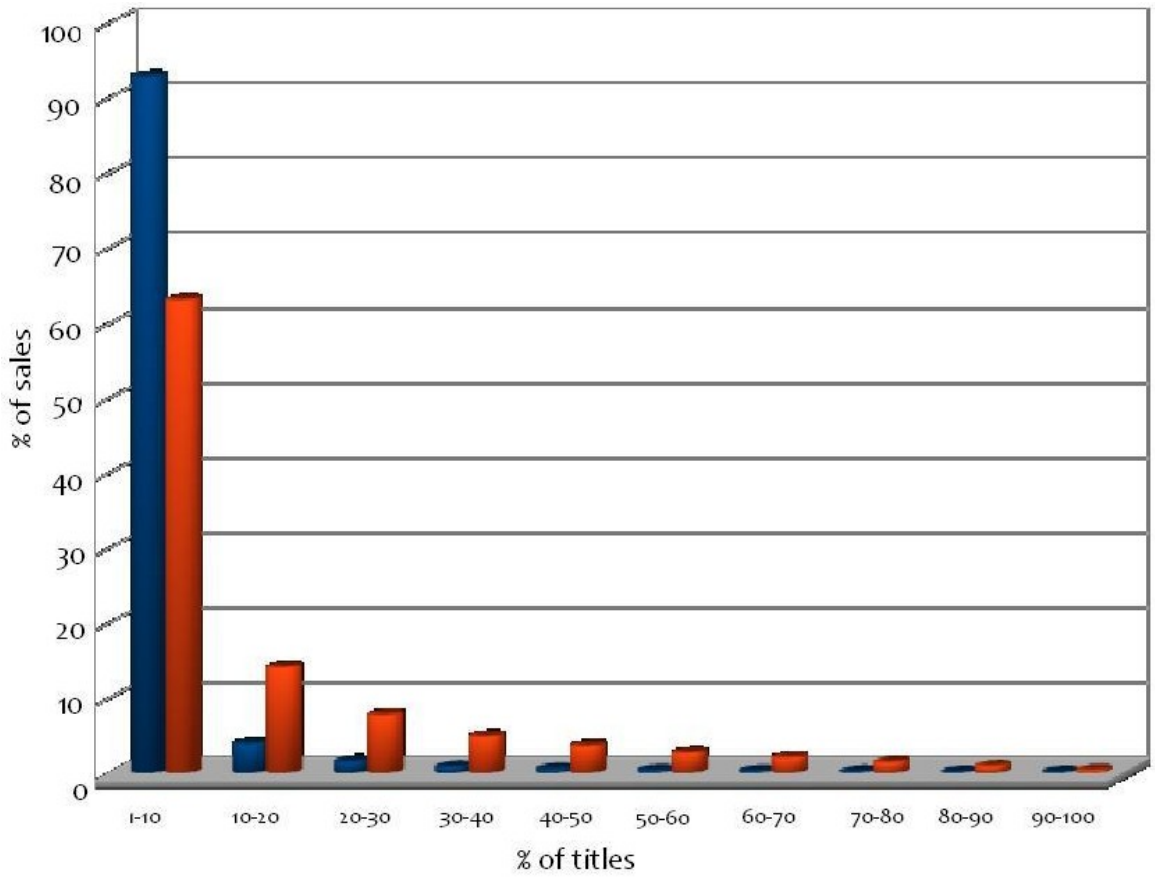


Figure 6.2: *Sales obtained at different percentiles*

Table 6.2: *Sales obtained at different percenties*

% of items	% Physical 2009	% Digital 2009
0-10%	92,7980	62,9722
10%-20%	3,7894	14,0149
20%-30%	1,4723	7,6462
30%-40%	0,7562	4,7735
40%-50%	0,4455	3,4974
50%-60%	0,2774	2,5971
60%-70%	0,1814	1,9739
70%-80%	0,1217	1,3699
80%-90%	0,0790	0,8575
90%-100%	0,0790	0,2973

The first decile accounted for 92,79% in case of CD sales, while for digital albums the first decile accounted for 62,97%, at this percentile it is where the difference in concentration is more pronounced, after that point every decile in the download album data accounted for a bigger market share, as it can be observed in the previous table.

6.2.2 A definition of head and tail for the digital and CD's market in 2009

The definition of head and tail presented in section 4.2.1 will be applied here to compare both markets. In the model proposed by Kilkki (2007), the cumulative distribution of sales is plotted using a logarithmic scale in the x-axis, resulting in an S-shape of the curve, the N_{50} point marks the point where 50% of sales has been reached and it also marks the turning point in the curve.

In order to divide the curve, two points are used: Xbm marks the limit between the base and the middle of the curve, it is found at $N_{50}^{2/3}$ .The second point used, $X_{me,}$ marks the limit between the middle and the end of the curve and it is found at $N_{50}^{4/3}$. On a logarithmic scale these two points divide the curve into three parts of similar size, not based on the number of titles but according to the percentage of sales. The usefulness of including a middle section is that it provides with a way to describe if demand is shifting from hits to the tail instead of only being distributed across many more titles.

Table 6.3 shows the points where the markets are divided for both markets as well as the point where 50% of sales occurred.

Table 6.3: Division points

	Digital Albums	CD's
N_{50}	103	149
X_{bm}	22	28
X_{me}	483	790

In order to make a comparison, not only the percentage of sales in each part should be used but also the percentage of titles included on each section should be used in order to reach conclusions. The percentages obtained are showed in table 6.4. If only sales are taken into account, the base/head would have the same importance in both markets. However, in the case of the CD market, titles in the head represent just a 0.07% of all titles, while in the digital market the head consists of 1,14% of titles. In other words, in the physical market a few titles are dominating the market, the head is reduced to a few superstars.

Table 6.4: A division of head and tail for the music market 2009

	Digital Albums		CD's	
	% of sales	% of titles	% of sales	% of titles
Base (Head)	23,9	1,14	23,61	0,07
Middle	57,55	24	53,29	1,9
End	18,55	74,8	23,1	98

The end-tail seems to be more important for the CD market: 23% of sales in the CD market vs 18,5% of sales in the digital market. However, again a look at the percentage of titles included in the end shows that in case of the CD market, 98% of titles would fall into the tail category based on this definition. This confirms the high level of concentration of the market.

Again, the difference in size between markets makes this comparison of head and tail deceiving, especially because the digital market is so reduced that there is not really what can be called to be a long tail: thousands of titles selling a few copies. The model presented in section 4.2.1 would be useful in this case for modeling the curve for the digital albums' market. Modeling the digital market would allow to predict what would happen if the market grew following the current pattern until reaching the dimensions of the market for CD's. Without such predictions, comparing head and tail won't provide the information that it is being looked for, in this case, the effect of more titles available into the online market.

As it was observed by the cumulative sales analysis, sales are less concentrated online, From the head and tail analysis what could be said is that this lower concentration online is reflected in more titles being in the head and more titles falling into the middle of the distribution.

6.2.3 Gini coefficient and Lorentz curves for the music market in 2009

In order to quantify the magnitude of the difference in concentration another measurement is necessary, one that summarizes the whole trend observable in the data and is not sensitive to size of the data. For this reason the Gini coefficient was chosen. Using the software *Matlab* and an inequality package develop by F. Pozzi, the Gini coefficient for both digital albums and the CD's was calculated. The results obtained are plotted in *figures 6.3 and 6.4*.

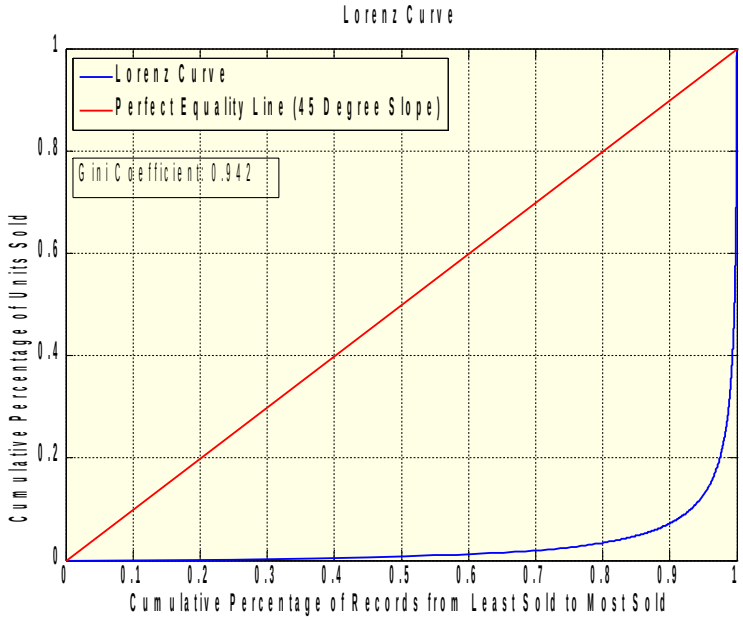


Figure 6.3: *Lorentz curves and Gini coefficient for the CD market*

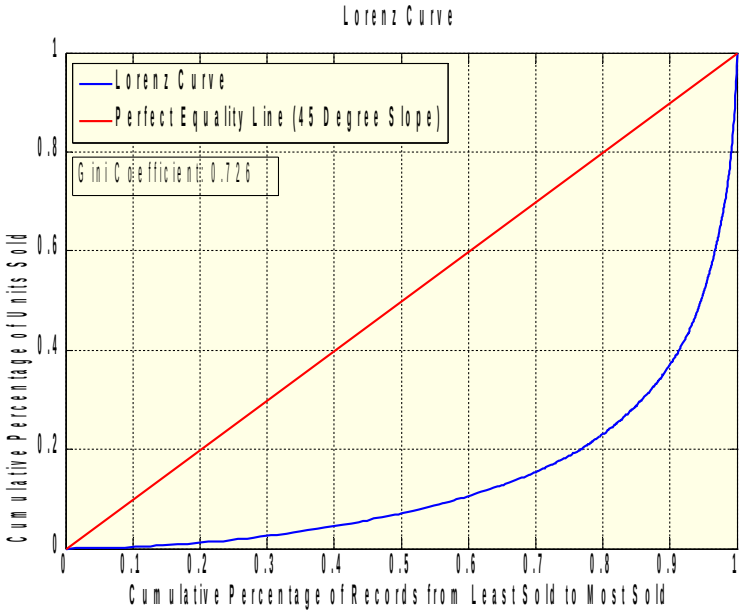


Figure 6.4: *Lorentz curves and Gini coefficient for the digital album's market*

It can be observed how the Lorentz curve for the CD market is further away from the perfect equality line than in the digital market. This is reflected in the higher Gini coefficient of this market. While the Gini coefficient for download albums has a value of 0,726, the Gini coefficient for the CD market is of 0, 942. The Gini coefficient of the CD market is very close to one, which reflects the huge share the top titles have and how flat the long tail is. This has already been observed in tables 17 and 18, where sales obtained at different percentiles were shown.

For the year 2009 it is clear that sales are more equally distributed across all titles in the digital market, the Gini coefficient showed that in the online market sales are distributed more

equally across different titles.

The limited size of the digital market might lead one to question if what has been observed in the download albums curve is nothing more than just the head of the CD market. In order to answer this question, not only concentration but also the composition of the list has to be explored. For this purpose, a comparison between both lists looking in detail at certain aspects could provide more evidence of the presence of a long tail effect, in the broader sense defined in chapter 4.

6.3 Analysis of the composition of sales in digital and physical format

6.3.1 The market share of major record companies in physical and digital format

Traditionally, major record companies have followed a hit-driven strategy, their aim has been to support and record artists who could become superstars. On the other hand, indies or small record companies, have focused on what can be called niche artists, many of them specializing in certain genres and aiming to serve a more focused market.

The music market in The Netherlands is dominated by the same four big record companies that are also present other parts of the world: Universal, EMI, SONY/BMG, and Warner. The remaining market is divided among a variable group of small record companies.

Data was available only from the year 2007 onwards, any result will be limited by the lack of a longer time-series. Nonetheless, these results will show the situation of the market in the last years and these could be taken as a good example of the market, given the recent origin of the digital market in The Netherlands and given the recent increasing tendency it has shown.

As it is expected, the CD market is highly dominated by the major record companies. During the period 2007-2009 these four major companies obtained more than 70% of sales. In contrast, the market for downloads is less concentrated, here the major record companies obtained around 54%. This is sign that the digital market is to a certain extent (around 15%) less concentrated and less dominated by the major record companies. There is also a decrease in the market share of the "big players" during these three years in the CD market, with the consequence of an increase in the market share of all the rest. From a 74,3% in 2007 their market share fell to a 71,3% in 2009.

Table 6.5: Market share of the four major record companies online-offline

	2007		2008		2009	
Record company	Album	Download	Album	Download	Album	Download
Universal Music Nederland	32,7	27,4	35,3	26	28,3	23,8
EMI Music Holland B.V.	18,6	9,6	16,9	9,6	14	9,3
SONY BMG Music Entertainment	14,1	11,4	13,8	11,2	22,2	16,3
Warner Music Benelux B.V.	8,9	5,6	7,2	7,8	6,8	5,5
All others	25,7	46	26,8	45,4	28,7	45,1
Four-Firm Concentration Ratio	**74,3**	54	**73,2**	54,6	**71,3**	54,9

When we analyze the digital market, major record companies appear to be gaining ground. In the digital market, the majors have grown from 54% in 2007 to a 54,9% in 2009. However, a closer look also reveals that the fragmentation is higher towards small record companies. Although major record companies have gained around 1% of market share, the smallest record companies, those who's market share is not even shown in the table because of its small size, have grown almost 10% in those same three years (see the "overig" category). Overall fragmentation has increased during the last three years.

Table 6.6: Market share of record companies online-offline

	2007		2008		2009	
Company	Album	Download	Album	Download	Album	Download
Universal Music Nederland	32,7	27,4	35,3	26	28,3	23,8
EMI Music Holland B.V.	18,6	9,6	16,9	9,6	14	9,3
SONY BMG Music Entertainment	14,1	11,4	13,8	11,2	22,2	16,3
Warner Music Benelux B.V.	8,9	5,6	7,2	7,8	6,8	5,5
Artist & Company	4,5		7		4,6	
Rough Trade	3,4					
PIAS	2,1		2,9		3	
CNR	2,1					
V2	1,8		2,6		4,4	
N.E.W.S.		3		3,2		3,3
8ball		2,5	2,4	3,1		1,6
Spinnin'		9,5		4,9		
Digidance		2,7				
CNR			1,9		2,1	
Marista						2,3
Berk				2,7		2,3
White Villa				3,5		
NRGY		4,6				3,3
Studio 100					2,5	
Overig	11,8	23,7	10	28	12,1	32,3

Another important feature of the digital market is that some small record companies are only featured in the download list, meaning that outside the digital market their market share is either too small or that they are not even operating in the physical market. For instance, Spinnin' appeared in the list in 2007 and 2008 in the download list only, in 2007 having even more market share than Warner and having an equal share than EMI. Surprisingly, it did not appeared in the list in 2009.

In fact, there are only few indies appearing the three years in the download chart, which also shows how volatile and turbulent this market is. The fact that the "overig" category is

increasing shows that it is possible that small record companies such as Digidance, that appeared in the list in 2007, are now together with other companies in this overig category. It can also be seen how the market share for any specific minor has decreased, while in 2007 Spinnin' had 9,5% of the market, in 2008 this had decreased to a 4,9% and by 2009 no independent record company reached more than 3,3% (in this case N.E.W.S and NRGY).

The comparison for the market share of record companies in the digital and physical markets can be summarized in two results. First, the market share of the "big four" in the digital market has been lower during this period of three years, and the difference with the CD market has remained close to 20%. This shows that online, small companies carrying niche content have more presence and might even compete with majors. Second, fragmentation of the market has increased, with almost a 10% increase of the "overig" category, that is, the category grouping all the minors with a very small market share. However, this fragmentation has not touched the majors, which have remained stable during those three years. Competition seems to be occurring more at the "indie" level in the digital market.

Overall, the lower concentration for download products is a positive outcome for the long tail hypothesis, however, the fact that major record companies have not lost market share online during these last years remains a puzzling result. If minor record companies represent mostly tail content, then the tail is bigger online, but it has not grown in the last years. As the market matures, it will be important for further research to broaden this results with more points in time that verify the validity of these results.

6.3.2 Top 200 differences CD's -digital albums

After having studied the concentration of the market for download albums and CD's there remains the question of how different hits are in both markets. Hits will be defined for this purpose as the top 200 titles, this quantity is the usual limit used to define hit lists in many countries, the best example being the Billboard charts.

When comparing the two lists, the first striking result is that 99 titles are different in both lists, in other words, half of the titles ranked in the best-selling 200 download albums are different from the best-selling 200 CD's. In order to evaluate if this differences can be considered to be important, it is necessary to analyze in which position were these different titles ranked. If most of these different titles were ranked in a position close to the top 200, then the differences between lists would be hardly of any significance in terms of a real difference between these two markets. Table 6.7 shows the position at which these digital albums ranked in CD format:

Table 6.7: *Rank from titles outside the CD top 200*

Rank	Titles	%
Not sold on CD format	**15**	**15,15%**
200-299	26	26,26%
300-399	17	17,17%
400-499	12	12,12%
500-599	5	5,05%
600-699	5	5,05%
700-1000	7	7,07%
<1000	12	12,12%

A considerable number of these titles ranked between position 200 and 1000. It is difficult to define a point where the ranking difference becomes important in terms of a real market difference. However, items ranked below the top 1000 position can be considered to be tail albums if we take into account the number of titles carried out by a small record store, this number is also mentioned by Anderson in his book to define tail content for music albums (Anderson, 2006: 136). Taking the top 1000 as a limit between head and tail, it could be said that 12% of music titles rank in the top digital top 200 was tail content in the CD market.

Perhaps the most important difference between the two top 200 lists are the titles ranked in the digital top 200 that were not sold during 2009 in CD format. The digital top 200 albums not sold in CD format shows a mix of hit artists, back catalog and in a lesser degree, niche artists. What can be observed is that many of these albums are different editions of hit albums by superstars. For example we can find two *deluxe edition* Madonna's albums. In many cases these albums are extended versions of the regular album.

There is also back catalog albums from artists like Pink Floyd and live performances of superstars like U2. In its majority, these titles are additional albums from hit artists. Few of these albums are from artists that can be called niche artists.

However, it is still important to mention that there are some niche artists reaching the top 200, for instance Lele, a dutch group that did not sell a CD and is signed by Magnetron Music, an independent dutch record company. Additionally, there are also albums that despite being from a top hit artists are only available online.

Summarizing, the differences observed in titles ranked in the top 200 for download albums can be explained through different causes: different versions of hit titles, back-catalog and "real" tail content, in the form of albums by lesser known artist. Although hit albums in different versions are found in many cases, there is some evidence of niche artists becoming hits. Overall, the difference show a digital top 200 that has given more space for older and niche content.

This analysis of the top 200 ranked items opened two ways to compare the digital and physical markets: content consumed online-only and the performance of tail CD's in the digital market. The following sections will explore these two angles in a more detailed way.

6.3.3 Digital Albums not sold in CD format:

Albums selling in digital format only, are a significant part of the digital market. In 2009, 374 albums sold digitally were not sold in CD format, representing a 19,5% of all digital albums being bought during that year. These albums accounted for nearly 10% of the total units sold. Major record companies also benefit from these albums since 54,5% of them belong to one of the major record companies. The titles not sold on CD format are distributed across the whole sales' list. The following table summarizes the distribution of these titles in the *digital* ranking:

Table 6.8: *Rank of digital albums not sold in CD format*

Rank range	# of different albums	%
1 to 500	57	15,24%
501-1000	93	24,87%
<1000	224	59,89%
Total	374	100,00%

As it can be observed in the table, the albums sold in digital format only populate the whole list, and almost half of them ranked in the upper end of the list. Therefore, it can be said that the these albums represent an integral part of the market, since not only a considerable amount ranked in the higher half of the sales' list, but a 15% ranked higher than the top 500 position, which could be considered as the head of the distribution.

These album differences are a sign that despite the differences in size between the two markets, there are important differences in the titles being sold. The digital market is more than just the head of the CD market, it is also more than just a selection of what is available on CD. It is a market were almost 20% of sales come from content that either is only available online or is not widely available on CD.

6.3.3.1 Albums from artists who had at least one other title sold on CD

These 374 albums become reduced to 331 albums when only different artist are considered. Out of these 331 different artists, 65% were present in the CD list with a different album or in collaboration with another artist.

Since these artists were ranked with a different album in the physical market, in order to have a reference point that tell us if the artists could be considered hit or niche, the highest ranking CD of these artists is the best measurement of their importance offline. Figure 6.6 shows how these artists ranked in the CD sales' list, based on their highest ranked title:

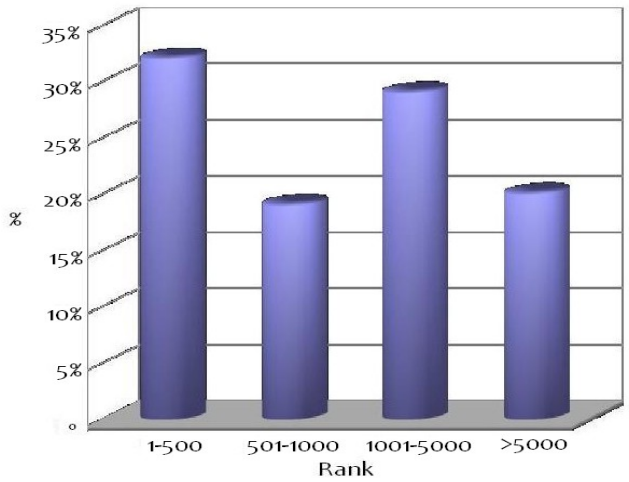


Figure 6.6: *Highest ranked CD for artists with a different album in the digital market*

A 32% of artists with an album on the digital albums sales' list, ranked between 1 and 500 in the CD sales' list with a different album. Therefore the remaining 68% of titles comes from artists that ranked below the top 500. A 29% ranked between position 1000 and 5000 and a 20% ranked below position 5000.

If we consider artists ranked below the top 5000 as "tail" artists, then a 20% of the albums that were not sold in physical format comes from one of these artists. If we consider that artists ranked below position 1000 are outside the head of the distribution, then almost half of the albums not sold physically belongs to an artists that could be considered tail artists, this could be justified if we consider that titles beyond the top 1000 sold less than 1% of the number of units sold by the top 1 title.

6.3.3.2 Albums from artists who did not appear in the CD sales' list

The remaining 35% of artists who's albums were not sold on CD format did not appear either with a different album or in collaboration with another artist in the CD sales' list. These albums may fall directly into the *tail content* category, since not only the album was not sold offline, the corresponding artist was also not sold offline. Among these albums we can find compilations, soundtracks, back-catalog and independent artists unknown to the physical market. Not only hit artists are absent, also niche artists of the physical world are absent. Therefore, these 116 albums bought only in digital format, represent an increase in variety consumed online not only in terms of titles but also in terms of artists.

6.3.4 Performance of "Tail" CD's also sold in digital format:

So far, the differences in the composition of both markets have been explored, yet, there remains an important element to be analyzed, namely, the performance of titles sold both ways: CD's and digital. One of the outcomes that could be taken as a sign of the long tail is a better performance online from albums classified as niche or tail content in the physical market.

6.3.4.1 How to compare performance in digital vs physical

When comparing markets of a different size, any definition of tail based on rank would lose some of its meaning, the differences in sales' volume and the differences in the size between both markets make it difficult to compare them. For example, the top 1 title, sold 12 times more in the CD market than in the digital market. Moreover, the CD market has almost 21 times more titles that the digital market.

Moreover, if we look at the lowest part of both data-sets, there are several titles selling the exact same quantity, and having because of alphabetical reasons, a higher or lower ranking. For instance, from rank position 30059 onwards, all CD's sold exactly one copy. Therefore, it is pointless to speak of titles 35000 and 30000 since both sold actually the same. Therefore, the approach chosen was to compare units sold between both markets. The worst selling CD's appearing also in the downloads list can be compared in terms of units sold, performance is measure in units sold, not in ranking level.

Two points were used for comparison: titles selling from 1 to 10 copies and titles selling from 11 to 20 copies. Putting into perspective what these numbers mean in terms of ranking, titles ranked below position 13 197 sold 10 CD's or less and titles and titles ranked below position 9273 sold 20 copies or less. Although the tail could be defined from a higher ranking level, for example position 5000, by using the chosen levels problems of setting the border between head and tail at a doubtful position are avoided.

6.3.4.2 Performance of CD's selling between 1 and 10 copies

In total there were 104 titles selling between one and ten CD's that were also sold in digital format. Figure 6.7 shows how download versions unweighted CD sales at every sales level from one to ten units per titles.

As its shown on the chart, the differences are significant in a majority of cases, particularly for albums selling from 1 to 6 copies. In all cases titles sold better digitally than on CD format. Out of these 104 albums, 68 albums sold more units in digital format, which represents a 65,38% of tail titles performing better on digital format. These titles combined sold more than 5 times the

units sold in CD format. Although there are 35% titles that sold more in CD format, when all 104 titles are taken into account, still the download albums sold 3,7 times what the same titles sold in CD format.

This positive balance can be explained through the performance of some of these albums online. For instance, the most pronounced difference occurs with a live album by B. Springsteen, this album sold only 3 CD's, while online 217 copies were downloaded. Other 5 albums sold 50 units more on digital format than on CD and another 28 albums sold between 10 and 50 albums more.

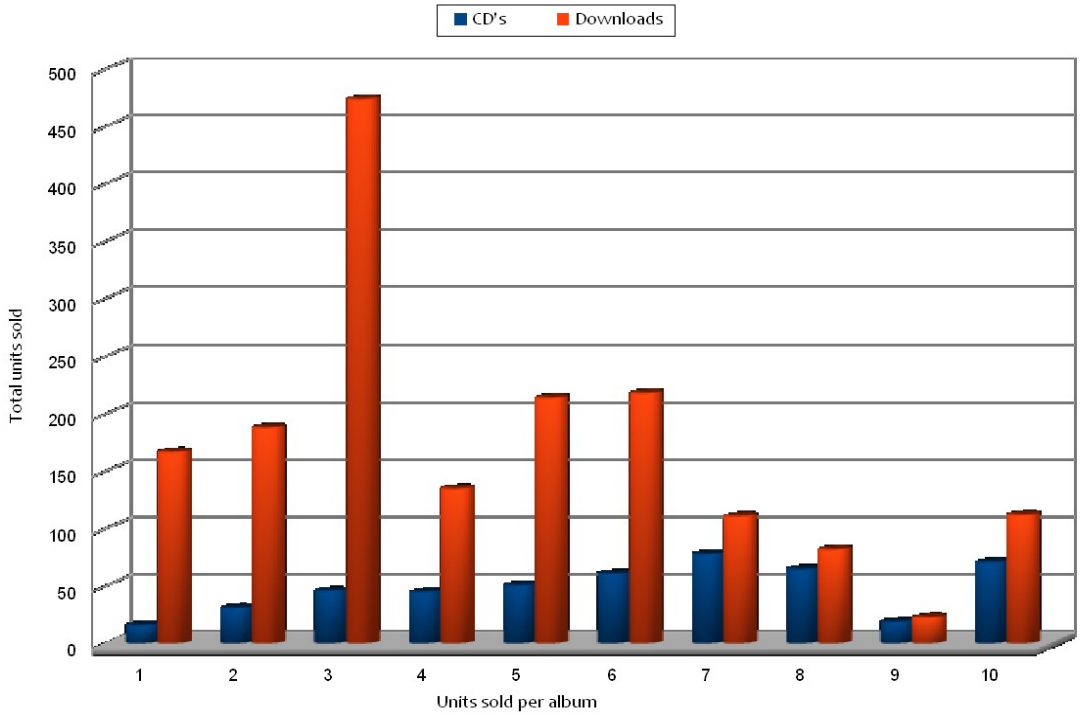


Figure 6.7: *Performance of CD's selling between 1 and 10 copies*

6.3.4.3 Performance of CD's selling between 11 and 20 copies

It is possible to move forward up through the tail and perform the same analysis for titles selling between 11 and 20 copies, these are titles ranked between position 9273 and position 13 197 in the CD sales' list. These titles are showed in figure 6.8.

There were 36 albums selling within this range that were also sold on digital format. Within this sales' range, still digital formats surpass physical sales. Although distances are shorter than in the case of titles selling less than 10 copies, titles sold in both formats performed better on digital format. In the case of titles selling 13 copies and 16 copies, CD's were sold more, however this difference is more than balanced by the total units sold by albums selling 14 copies.

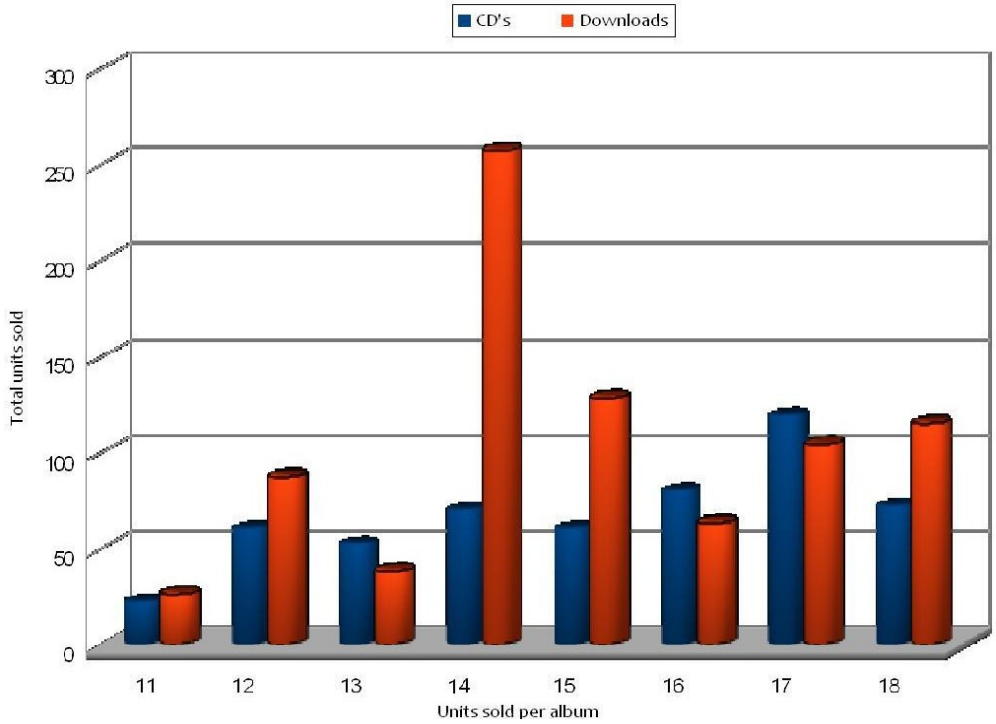


Figure 6.8: *Performance of CD's selling between 11 and 20 copies*

From this data, it becomes evident that tail albums that sold less than 10 CD's were much more important in the digital market. The given example of the album by B. Springsteen shows how an album ranked after position 20000 in physical format ranked close to the top 200 online. Other examples are less dramatic, but nonetheless still significant.

It is important to remember that the total number of different albums downloaded is only a 5% of the titles sold on CD format, therefore, the simple presence of these tail titles on the sales is a sign that tail content is more important online. The market online is not including only the head, despite its limited size. However, the real sign indicating that there is a difference between both markets is the fact that titles that hardly sold copies on CD format performed much better on the digital world.

The popularity of tail titles online might give the impression that independent artists are the most benefited from this popularity. However, although some independent artists are enjoying more popularity thanks to digitalization, it is important to bear in mind that, out of the tail albums selling between 1 and 10 copies, 55,57% belongs to a major record company. When all titles selling between 1 and 20 copies are counted, this percentage increases to a 58,57%. This suggests that not only minor record companies and niche artists are benefiting from sales on digital format, major record companies can also increase their profit with back-catalog and some of their least sold artists.

6.4 Summary of the results

The long tail hypothesis claims that online, thousands of niche albums selling a few copies will grow to be a contender for hits. In the case of The Netherlands, and the digital albums' market, this is far from being the current state of affairs. The market for digital albums is a small market, where thousands of titles available are not being bought. On this respect the hypothesis has to be rejected.

Nonetheless, if size difference is left aside and both markets are compared, the online market shows a less concentrated distribution of sales. The tail is much shorter but the overall distribution is more evenly distributed.

Besides concentration, it has been argued here that the long tail should be defined in broader terms to include the relative importance of niche content and differences in composition. When compared on these other aspects, the digital market showed important differences with respect to the physical market. On the digital market, niche content had a better performance than in the physical market. This is supported by the lower market share of major record companies online. Moreover, hits are different in both markets and titles available online only play an important role in the digital market. If these aspects are taken into account, there is evidence of a long tail effect.

7

Conclusions

The main goal of this thesis was to find evidence of the existence of *long tail* phenomenon in the Dutch music market. According to the long tail hypothesis, in the online music market, hits will lose market share due to the availability of more titles and the presence of recommendation tools that allow consumers to find them, by adding-up many niche titles these can become profitable. The main goal would have been easily accomplished by measuring the market share of hits. Nevertheless, a more careful reading of the long tail hypothesis revealed that the scenario was more complex than it seemed.

The long tail hypothesis is more than a business model based on profiting from niches, it touches many other issues that are not directly related with quantity or revenue. Anderson's book "The long tail: why the future of business is selling less of more", makes more than just one simple prediction about the market share of hits; it describes a future for the music industry that challenges other theories like the superstar theory.

What makes the long tail so difficult to investigate is that it is described using so many different outcomes, that it is hard to test them all. This multiplicity of results allows to measure the long tail idea in at least three different ways: the length of the tail, the tail's importance as an economic source of income and differences in composition of the market thanks to recommendation tools and online access. However, the focus has been usually placed on the economic potential of the tail.

In this thesis, it was attempted to test all these possibilities in order to give a more comprehensive view at the online market and in order to test statements made by Anderson that cannot be answered by only measuring market share of the top titles.

7.1 Limitations encountered while making this research

When it comes to measuring market concentration there are multiple issues that make the task complex. First, there are concepts that are loosely defined, making it difficult to know how to measure and test the hypothesis. An example of this lack of definition is the multiple ways on which terms like "tail" or "niche" content are defined. Sometimes, tail is what goes beyond the stock of a brick-and-mortar store, sometimes it is defined after a certain Top X rank or percentage.

Secondly, measuring concentration in one market over time and measuring it on two different markets simultaneously poses different challenges. Some of the difficulties of measuring differences in concentration between two different markets were explored in the fourth chapter of this thesis, as a way to show that measuring concentration is not so clear cut as it might seem at a first glance.

Another aspect that makes this topic hard to research are the difficulties involved in obtaining adequate secondary data. There are different sources of data that could be used: individual retailers' data or overall national level data . At a retailer level, it was extremely difficult to obtain data. Without quality data from retailers the topic will cannot be adequately researched, this is probably one of the reasons why this topic has not been studied more.

On the other hand, statistics of the whole market present many limitations. Although both the NVPI and the GfK were extremely kind in providing with information, the data needed to test some hypotheses was not available. For example, the information about the specific retailers included in the data was not provided as well as the total number of titles available in the market. Moreover, some data useful for this topic is of a very recent origin, as it was the case of the online market share of record companies which was only available from 2007 onwards.

7.2 Methods and findings

This thesis intended to answer one main research question: *Is there evidence of a long tail effect in music market in the Netherlands?* The results obtained give supporting evidence for some of the outcomes predicted by the long tail hypothesis. However, in order to consider some of these results as evidence of the long tail in the market, the long tail hypothesis had to be defined in a broader way than it has been defined by other research done before. If the long tail is defined mainly by the quantity of niche titles being consumed, it is hardly possible to speak of a long tail in the Dutch digital albums' market. If the relative importance of niche content is taken into account, then, the ideas behind the long tail are supported by the data used in this thesis.

One of the main insights gained from this research is that the long tail hypothesis as described by Anderson describes many outcomes that are not necessarily dependent on one another, however they are described as a single phenomenon. For this reason, some results seem to be contradictory if the long tail idea is approached searching for a simple numeric result. The main findings of this thesis are summarized in the following paragraphs.

7.2.1 Evidence against the long tail hypothesis

The most important argument against the long tail hypothesis as described by Anderson, is that a great number of titles available online is not being consumed. Anderson describes the long tail as a "paradise of choice" where every item finds a customer. However, as the results show, both in the case of CD's as in case of digital albums, only a fraction of the content available is being bought. This is particularly evident in the digital market. In the case of the CD market, despite the increase in market share from e-tailers, the tail has not increased in terms of more titles being consumed during the last five years.

On the other hand ,by simply looking at the data from digital albums, a conclusion can be made: at least as a business model -where thousands of niche items translate in to revenues for e-tailers- the hypothesis has to be rejected. It remain to be studied if the same applies for the digital singles' market, where the majority of sales take place.

In terms of concentration, in case of the CD's, five years of increased content available online have not had a significant impact on the market. In terms of concentration, there has not been a visible change in the market. The CD market is highly concentrated towards the top titles, more than what the Pareto principle describes, in this market the market is closer to a 90/10 distribution.

7.2.2 Evidence supporting the long tail hypothesis

In the CD market there is a long tail, that despite not having increased, it would not have being possible before e-tailers' appearance in the market. Moreover, Cosmox's data showed an example of an e-tailer making revenues from the tail. In Cosmox's case, titles in the tail were as profitable as hits. Cosmox's data suggests that a pricing strategy that charges higher prices in the tail could important as a source of profits.

Market concentration was measured on different ways. The digital market showed lower concentration levels, as showed by means of using the Gini coefficient. This result can be supported with the fact that online, the four major record companies have a lower market share.

In the digital market, there are important differences between the albums being bought physically and digitally. A detailed analysis of the composition of sales in both markets showed significant differences between them. First, the top 200 titles showed differences in composition with half of titles being different between markets, 15% of titles in the digital top 200 are albums not sold at all in the CD market. Titles not sold on CD are an important part of the digital market, in fact 35% of these albums are from artists who did not have any title being sold in CD format that same year.

One important result in favor of the long tail hypothesis is that a majority of albums selling

less than 10 or 20 copies in the CD market, sold more units in digital format. Taking into consideration the limited number of different titles sold in the digital market, this result becomes even more striking.

Overall, despite being still a very small market, the market for digital albums does show important differences from the physical market. If the long tail is defined in a broad sense as it has been done in this thesis, the effects of increased variety can be confirmed in the digital market. If we consider the market for digital albums as an emerging market that should grow much more in the following years, the results obtained in this thesis should encourage retailers to make niche content available.

7.3 Answer to the research question

After presenting these results a final balance has to be made in order to answer the main research question presented at the beginning of this thesis:

Is there evidence of a long tail effect in the music market in the Netherlands?

It is not possible to provide with a definitive answer to the question. The overall balance provides evidence that online sales and digitalization have had an impact on the market in many aspects. However, the long tail as a profitable business model could not be observed in the digital market.

The sub-questions formulated in the introduction attempted to cover elements of the long tail hypothesis not covered in previous research:

Is the market share of hits lower for digital products?

For this question the answer is no. However, it was shown that when comparing different size markets, hits defined in terms of the top x titles might have more market share and still not provide with an accurate description of how concentrated the market really is.

Is the overall demand for records more equally distributed in digital markets?

Yes, using Gini coefficient and exploring the distribution of sales across the data, the digital market turn out to be less concentrated.

Are the best-selling records different in digital format than in physical format?

Almost half of the top 200 is different between markets, within this different online titles we can find tail titles, niche artists and content sold only online. The answer in this case is yes, an important part of the hit titles is different online.

Does niche content sell better in digital format?

Yes, titles that sold less than 10 copies on CD format sold much more on digital format. This is a clear result that online niche content can perform better than offline. Also niche artists were able to reach the top positions in the digital market.

When faced with these results, the effect of digital sales can be clearly observed, particularly in terms of niche content performing better online. It also justifies the importance of having a wider view on this topic, since otherwise these results would have been overlooked.

7.4 Recommendations for future research

Further research should be done on the differences between the digital and the physical music market that were studied in this thesis. Not only concentration should be studied, but also the differences between titles consumed and market structure demand more attention. The results that were obtained here, suggest that this might be a fruitful avenue for research and a valuable addition into our understanding of how technological change is affecting the market for music products.

Furthermore, the study of Cosmox's sales raised interesting questions about pricing structure and ultimately about who is consuming niche content. Therefore, it could be an interesting area for research to study and track down who is consuming tail titles. For example, are the consumers buying niche titles buying also best-sellers? Is niche content being bought by heavy (frequent) consumers or by light consumers? Also it is important to test if this pattern is observed in other e-tailers or if was only the specific case of Cosmox.

This thesis described how the divisions of retailers made by Anderson no longer represents the market, therefore future research should find ways to study the long tail in a way that includes that changes that have occurred in the market during the last years. Therefore, future research should be done with a different reference point when it comes to the physical market.

Finally, this thesis aimed to provide in section 4.2.1 with a more objective definition of head and tail, one that does not rely on the stock of a brick-and-mortar store-which as mentioned above-might not exist in the future, therefore losing its meaning as a point of reference. More research should be done in this area, since it would provide with a tool to explore this topic.

7.5 Towards a more comprehensive definition and measurement of the long tail

The long tail hypothesis implies, besides measuring the market share of hits, topics like differences in the titles being consumed offline-online and the importance of niche content online. Anderson himself seems to underestimate these aspects, aspects that nonetheless could have great importance in current music market.

Instead, his main thesis focuses on predicting "the fall of the hit" and on promoting a new business model based on niche content. It is precisely at these aspects where the long tail hypothesis has fallen short, since the changes brought by online distribution and digitalization have turned out to be less dramatic than what his writings have predicted.

The emphasis Anderson puts into this *new business model* also focuses the attention of research into simple measurements of whether "the top X accounted for x percent of sales". Therefore losing perspective of important changes that are occurring in the music market.

This thesis aimed to cover these changes, the long tail was looked from a broader perspective and under this perspective, there is evidence that the long tail has indeed brought significant changes into the music market.

It is also because of this broader approach, that the results found cannot be conclusive and support or discard the hypothesis on its entirety. What these results show is that the long tail hypothesis was defined in very broad terms and that it includes many different areas, but it has been measured in a *narrow way*. It is necessary to re-evaluate if defining the long tail only as a business model is the right way to address a market as complex as the digital market.

The business model proposed by Anderson seems to be far from becoming the norm, superstars are still the dominant figures in the music market. From this point of view the long tail hypothesis has to be rejected. In the CD market what can be seen is a market dominated by superstars with an extremely flat and long tail. In the digital market, the tail has more share but the limited size of the market does not support Anderson's hypothesis.

Nonetheless, it is important not to dismiss the changes the whole digitalization process has brought. Even if as a business model the long tail did not appear to be strong in the digital market, other outcomes mentioned by Anderson were indeed observed in the data. There are important differences between the physical and the digital market in terms of hits, overall composition and popularity of niche content, *all these aspects are part of what the long tail idea is all about*.

The long tail is in the end more than just a business model that fails if not enough profit can be made from the tail. It is a concept that encompasses the role that technology has into our life's and the effect more choice has for us as consumers.

References:

Books and Articles:

Anderson, C. 2004. *The long tail*. Wired Magazine 12(10) 170-177.

Anderson, C. 2006. *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.

Benghozi, P.J. & Benhamou, F. 2008. *The long tail: myth or reality?* Digital Music Report 2009: New Business Models for a Changing Environment. Published by IFPI. January 2009.

Brynjolfsson E., Hu, Y. & Smith, D. 2007. *Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales.* Social Science Research Network. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=953587

Debraj, R. 1998. *Development Economics*. Princeton University Press

Elberse, A. 2008. *Should you invest in the long tail?* Harvard Business Review (July-August) 88-96.

Elberse, A., F. Oberholzer-Gee. 2008. *Superstars and underdogs: An examination of the long tail phenomenon in video sales*. Working Paper.

Frank, R. and Cook, P. (1995), *The Winner-Take-All Society*, New York: The Free Press.

Hirshleifer, David A. *The Blind Leading the Blind: Social Influence, Fads and Informational Cascades*. The New Economics of Human Behaviour, Ierulli, K. and Tommasi, M., eds., Ch.12, pp. 188-215, Cambridge University Press, 1995; Anderson Graduate School of Management, UCLA, Working Paper No. 24-93

Kilkki, K. *A practical model for analyzing long tails*. First Monday, Volume 12, No.5, May 2007 http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/1832/1716 Retrieved on June 15, 2010

Kretschmer, M., G.M. Klimis and J.C. Choi. 1999. *Increasing Returns and Social Contagion in Cultural Industrie*s, *British Journal of Management* 10; 61–72.

Liebowitz, S. and S. Margolis. 1998. *Network Effects and Externalities,* in *The New Palgrave Dictionary of Economics and the Law* Vol. 2, MacMillan; 671-675.

Nelson, P. 1970. *Information and Consumer Behaviour. Journal of Political Economy* 78(2); 311-329. *smu.edu/meadows/.../Benghozi%20Benhamou-223.pdf*. Retrieved 5 October 2009.

Peitz, M. and Waelbroeck, P. 2005. A*n Economist's Guide to Digital Music*. CESifo Economic Studies, Vol. 51. 2-3/2005, 359–428

Rosen, Sherwin 1981, *The Economics of Superstars*. *The American Economic Review*, 71 (5), 845-858.

Tan T. & Netessine, S. 2009. *Is Tom Cruise Threatened? Using Netflix Prize Data to Examine the Long Tail of Electronic Commerce*. *knowledge.wharton.upenn.edu/papers/1361.pdf.* Retrieved 5 October 2009

Towse, R. (ed) 2003 *A Handbook of Cultural Economics.* Edward Elgar Publishing.

Tucker, C. & Zhang, J. 2007. *Long Tail or Steep Tail? A Field Investigation into How Online Popularity Information Affects the Distribution of Customer Choices*. Massachusetts Institute of Technology (MIT), Sloan School of Management.

Tucker, C., Zhang, J. 2009. *How does popularity information affect choices? A field experiment*. MIT Sloan Working Paper.

Sales' data over the Dutch market:

Provided after request specifically for this thesis by the GfK and the NVPI.

Cosmox's Sales data 2008:

Provided by Cosmox for this thesis after request.

Online Sources:

http://www.longtail.com

Times Online:
http://entertainment.timesonline.co.uk/tol/arts_and_entertainment/music/article5380304.ece

Wolfram demonstrations Project:
http://demonstrations.wolfram.com/TheLongTail/

GfK:
http://www.gfk.com

NVPI:
www.nvpi.nl

International Federation of the Phonographic Industry:
http://www.ifpi.org/

Hoofdbedrijfschap Detailhandel:
www.hbd.nl

Entertainment Retailers Association
http://www.eraltd.org
http://www.eraltd.org/_attachments/Resources/yearbook.pdf

Billboard Magazine:
http://www.billboard.biz/bbbiz/content_display/magazine/features/e3i35ed869fbd929ccdcca52ed7fd9262d3

De Vereniging van Organisaties die Intellectueel eigendom Collectief Exploiteren (Voice)
http://www.voice-info.nl
http://www.voice-info.nl/assets/voice/het%20digitale%20aanbod%20van%20beeld%20en%20muziek%201%20oktober%202009%20_2_.pdf
Retrieved: 10 May 2010

Record stores cited:

Free Record Shop:
www.freerecordshop.nl

Van Leest:
www.vanLeest.nl

Cosmox:
www.Cosmox.nl