The Box Office and the Long Tail: An Examination of the Effects of Streaming on the Distribution of Box Office Revenue

Tristan Drost

Student number: 579524

Supervisor: Dr. C.W. Handke

MA Cultural Economics and Entrepreneurship

Erasmus School of History, Culture and Communication

Erasmus University Rotterdam

Master Thesis

20-06-2021

The Box Office and the Long Tail:

An Examination of the Effects of Streaming on the Distribution of Box Office Revenue

Abstract

The long tail theory suggests that the development of online distribution channels will increase the demand for niche products and shift the demand towards the tail of the sales distribution. Conversely the body of literature that has analysed the effects of online distribution channels on the film industry has shown that the opposite has happened here. In the film industry sales have become more concentrated on a small number of hits. The existing literature on the long tail theory in the film industry has put little focus towards the effects on the primary source of income in the film industry: the box office revenue. Moreover in recent years streaming has become an increasingly popular form of online distribution in the film industry. The literature regarding the effects of streaming is currently relatively limited. This research analyses the effects of streaming as an online channel on the distribution of box office revenue. The research uses box office data from all movies that were screened in cinemas in North America in the period from 1996 until 2019. A time series analysis is used to determine whether the distribution of revenue has become more concentrated after the introduction of streaming. For this research the launch of the streaming platform of Netflix in North America has been considered as the introduction of streaming. The results of this research show that the distribution of revenue has become more concentrated on the top films since the introduction of streaming. Furthermore the results show that the combined market share of the six major film distributors has not increased after the introduction of streaming and only one major distributors has been able to increase its market share in this period. Thus this research provides additional evidence that revenue in the film industry has become more concentrated and supports the claims from previous research that the long tail effect is not present in this particular industry.

Keywords: long tail theory, streaming, box office, concentration of revenue, superstars

Photo credits title page: Jay L. Clendenin/Los Angeles Times via Getty Images (2021)

Table of Content

I. Introduction

As part of the evolving role the internet plays in the world in the last decades, the introduction of ecommerce has had a strong influence on the way people consume products. Online distribution channels have facilitated the purchasing process of consumers and have allowed suppliers to distribute a wider range of products and services. Academic researchers also agree that the emergence of ecommerce and online distribution channels have changed the type of products and the variety of different products that is consumed. Yet no universal view exists among researchers on the type of change that is brought by these new distribution channels. One theory suggests that ecommerce will lead to more heterogenous consumption and a less concentrated distribution of sales. This theory has been called the long tail phenomenon by Anderson (2004). The theory is mainly based on the fact that suppliers are able to provide a greater number of different products online than in traditional brick-and-mortar stores. In certain industries the consumption of niche products has indeed grown, while big hit products have suffered and become less dominant (Brynjolfsson, et al., 2003). However this change is not generalizable for all industries. Especially in the cultural industry the type of change that has occurred seems to be different. Research on the distribution of sales in the music and film industry has shown that the opposite is true for these particular industries (Elbers, 2008; Elberse & Oberholzer-Gee, 2007, Ordanini & Nunes, 2016). While more products may be available, the distribution of sales becomes more concentrated. The causes that are often cited for this phenomenon are based on the superstar theory developed by Rosen (1981). The internet has lowered distribution and transaction costs for producers of music and films and people generally rather watch or listen to a more talented performer. According to this view, now that high quality cultural products have become available nearly everywhere at the same time, consumers should all flock towards these products provided by the most talented producers.

The findings of the current literature regarding the distribution of sales in the film industry have pointed towards a more concentrated market as a result of online distribution channels (Elberse & Oberholzer-Gee, 2007; Tan, et al., 2017). However few papers have addressed the impact of the technology of streaming. Streaming is a specific form of online distribution that has become increasingly popular in the film and music industries in recent years. Instead of purchasing physical products through online channels, streaming allows consumers to watch or listen to cultural products online instantly in real time. In general the impact of streaming services on the film industry have not been fully assessed yet by academic literature, presumably due to the fact that streaming only flourished in the last decade or so. Some researchers agree that streaming services have challenged the traditional market players of the entertainment industries and are the frontrunners in terms of new business practices in this industry (Burroughs, 2019; Cunningham, et al., 2010). Still empirical results are relatively lacking in this area. Furthermore the current literature regarding the long tail in the film industry has mostly used data regarding DVD sales, either offline or online, yet the impact on the main form of income for the film industry i.e. the box office, has not been researched. Whether the long tail phenomenon actually holds for cinematic releases or whether the distribution of revenue earned by cinematic releases has become more concentrated is not clear. Thus there is a gap in the existing literature on the effects that online channels have had on the of box office revenue.

This research will therefore focus on the impact that the introduction of streaming has had on the distribution and concentration of film performance in the form of box office revenue. The main research question of this paper is: *How has the introduction of streaming affected the concentration of box office revenue?* This question will be examined by answering several sub-questions:

1. Has the introduction of streaming led to a more concentrated distribution of revenue earned by individual films?
2. Has the introduction of streaming led to a more concentrated distribution of revenue earned by the film distributors?
3. How has the introduction of streaming affected the overall shape of distribution of revenue earned by individual films?

This research will address these questions by using box office data of movies screened in movie theatres in North America from a sample period of 1996 until 2019. This data is taken from Boxofficemojo (IMDbPro, 2021) and The-Numbers (Nash Information Services, 2021). By using a time-series analysis this research will try to answer whether the introduction of streaming has led to significant trends in the concentration of revenue in the period following the introduction. The introduction of streaming is identified by the launch of the streaming platform of Netflix, which was in 2007. Therefore the data is split into two different periods: the pre-period before the introduction of streaming until 2006 and the post-period after the introduction of streaming in 2007. As this research deals with a limited dataset an intervention analysis for short time-series developed by Tryon (1982) will be used. Moreover an assessment of the Gini coefficient will be used to inspect the overall shape of the revenue distribution. The use of this method will indicate whether trends in the concentration of revenue are consistent with the introduction of streaming, yet full confirmation of a causational relationship cannot be provided.

The results of this research show that the box office revenue has become more concentrated on the top films. A positive trend in the percentage of revenue earned by the top 10 films as well as the top 20 films is found in the period after the introduction of streaming. Moreover the results provide no evidence that the box office revenue has become more concentrated on the six major film distribution companies. There is only one major distributor that has increased its market share in the post-period, namely Walt Disney. This increase in market share might be due to firm-specific characteristics, but does not suggest that the largest distributors have become more dominant since the introduction of streaming. The results of this research are mainly in line with the superstar theory and provide additional arguments against the presence of the long tail phenomenon in the film industry. The results of this research are consistent with previous studies that suggest that the concentration of revenue is getting more concentrated in the film industry. For producers in the film industry the practical implications of this research are that the attention for hit products has become more intense. Where perhaps earlier the strategy to provide more niche products and supply a larger variety of products was encouraged by the long tail theory. The results of this research suggest that producers would be better off to focus their efforts in producing a small amount of absolute hit products instead.

Similar to the abundance of the literature on the long tail this research provides descriptive claims about the changes in the concentration of sales and revenue as a result of online distribution channels. The results of this research provide positive statements regarding the current conditions and developments of the cultural industries. Whether these developments are beneficial for consumers or suppliers is not stated and left as a question for normative studies in the future.

The research paper is structured as follows. In section II the relevant literature regarding the subject of this research are discussed. In section III the data that is used in this research is described. Furthermore this section includes the description of the method that is used in this research. In section IV the results of this research are presented. Finally in section V the results of the research will be discussed, the limitations of the research will be discussed and suggestions for future research are presented.

II. Literature review

Superstars

According to a popular body of research the evolution of online distribution and the digitization of information will mostly be beneficial for the most popular artists and products. The relatively new technologies would facilitate the convergence of demand on a small group of winners even if more options become available. This superstar phenomenon was explained by Rosen (1981) as a result of two factors. The first factor is a hierarchy of talent, which leads to imperfect substitutability of product with different qualities. This means that consumers generally prefer a certain service or product of a high quality over multiple similar products or service of lower quality. Evidently this hierarchy of talent is only relevant for certain types of products and services, such as cultural performances or medical services. Rosen (1981) explains this as follows:

> Hearing a succession of mediocre singers does not add up to a single outstanding performance. If a surgeon is 10 percent more successful in saving lives than his fellows, most people would be willing to pay more than 10 percent premium for his services. (p.846)

The second important factor is the nearly perfect reproducibility of a product or service. As the marginal costs of producing and distributing become lower and approach zero, the highest quality products can become available to all consumers and more talented artists can service the complete market. Any new technology that would lower these costs would therefore increase the advantage that the superstars have in the market. According to Rosen, online distribution and more specifically streaming would therefore lead to a greater convergence of consumption.

There is also a social explanation for the superstar phenomenon. Adler (1985) suggests that superstardom is not necessarily due to talent, but due to the preference of people to consume the same cultural products as others do. Adler also notes the importance of “consumption capital”, which is based on the idea that the enjoyment of art or an artist’s output increases with every consumption of art or production by this artist (Stigler & Becker, 1977). Moreover Adler states that people can increase their consumption capital by discussing art or artists with others. If an artist is more popular, more possible discussants and more media coverage increase the opportunities of such discussions, thus resulting in a convergence of consumption. Frank and Cook (1995) share a similar view about "winners-take-all markets". They state that the enjoyment of consuming a cultural product is not limited to the consumption in itself, discussions about these products offer joy and pleasure to consumers as well. Thus the whole experience of these products becomes more valuable, the more popular they become. On top of that Frank and Cook suggest that a rational consumer will often use popularity as an indicator of quality, which would also result in a vicious circle of increasing popularity for the best-selling hits.

The idea of the superstar effect in the arts, by Rosen (1981), was originally developed to pertain to persons or individual artists. It explains how the income levels between individuals can far exceed the talent levels between these individual artists. However the idea of superstars is nowadays not only limited to individuals. Frey (1998) applies the concept of superstars to museums and suggests that the same demand-side and supply-side drivers that cause the superstar effect in individual artists are behind the success of superstar museums. In recent years museums have been able to use new technologies, such as virtual tours, to reach economies of scale and thus also lower their reproduction costs. So even though the idea of superstar film production companies does not explicitly appear regularly in the academic literature, the idea of superstar organizations in the art industry is not completely new or without reason.

The long tail theory

Contrary to the idea that new forms of ecommerce, such as streaming, would be mostly beneficial to superstars stands the long tail theory proposed by Anderson (2004). According to Anderson, the future of entertainment would lie in the large amount of niche markets that have become available through online distribution instead of the traditional big hits topping the charts. According to Anderson (2004), one of the main reasons to this increased accessibility for less popular and less demanded products is the nearly unlimited shelf space that online retailers possess. Physical stores can generally carry a limited amount of products and will therefore not hold products in store that only amount to a small amount of sales. This constraint of limited shelf space is not a factor for online retailers. Making products accessible to a market consisting of nearly the whole world will lead to demand for even the smallest niches. As Anderson (2004) states: "…almost anything is worth offering on the off chance it will find a buyer."

On the demand side online distribution has significantly decreased information costs and search costs (Bakos, 1997). This decrease of search costs could lead to a lower concentration of sales and relatively less demand for hit products, as consumers are able to find obscure products and products close to their own specific taste easier (Lynch & Ariely, 2000). Evidence for this is also provided by Brynjolfsson, et al. (2003), as they find that the share of niche products that is normally unavailable in most traditional stores accounted for a large share of sales made on the internet. Additionally Peltier and Moreau (2012) provide evidence that the long tail theory holds up for at least some type of products in the cultural industry as well. They show that the use of online information and online distribution has led to a shift from bestseller books to medium-or low-selling books.

Another reason for less convergence of product sales online could be due to the use of recommendation systems by online retailers (Brynjolfsson, et al., 2006; Oestreicher-Singer & Sundararajan, 2006). Personalization and recommendation technologies are able to predict the interests of consumers based on earlier purchases and online behaviour. Based on these predictions, online retailers could steer consumers towards their preferred products better than traditional retailers can. Moreover these recommendation systems can find interesting products for consumers that they otherwise would not have found. Yet there is also evidence that recommendation systems mostly provide consumers with high-sales recommendations and thus actually increase the concentration of sales (Fleder & Hosanagar, 2009).

Even though evidence for a lower concentration of sales is provided for some markets, contrasting beliefs about whether the long tail theory actually holds specifically for the music and film industry can be found in the academic literature. Instead of an increased demand for niche movies and music, sales in these markets appear to concentrate even more on the top selling hits in the presence of ecommerce options (Ordanini & Nunes, 2016; Tan, et al., 2017). Furthermore Elberse and Oberholzer-Gee (2007) are critical of the idea of the long tail theory in the film industry. Even though they show that DVD and VHS sales have shifted towards the long tail from 2000 to 2005 as a result of online retailing, they also find that that the number of titles that went without sales during this period increased significantly. Thus the tail of the sales distribution is full of film titles that generate practically no sales. Consequently they argue that the tail of the sales distribution has not started bulking, on the contrary the tail has become longer and flatter than before. Moreover they find that the importance of individual best-sellers has grown from 2000 to 2005. Success has become even more concentrated on the hit titles, as they find that the number of titles in the top 10% of weekly sales has dropped by more than 50% percent. While the demand of niche products seems to have increased in absolute terms, hit products have become relatively more popular. According to Elberse and Oberholzer-Gee this trend is typical for winner-take-all markets. Elberse (2008) finds similar trends for the music industry. Even though a shift towards the end of the tail is present in this industry as well, the concentration of sales becomes larger too. Thus the tail of the distribution becomes longer, but at the same time it also becomes flatter. Elberse suggests that the best-selling hits reach lower sales volumes, yet the smaller group of top hits account for a larger percentage of the overall demand. Elberse explains this trend by stating that the demand for niche products is mainly driven by heavy consumers. People that consume more titles tend to occasionally pick an obscure title, yet their most frequent choices are still the most popular titles. Thus: "The implication is that there is no segment with a particular taste for the obscure; rather, customers with a large capacity for content venture into the tail" (Elberse, 2008).

The main underlying reasons provided by the literature on the lack of a long tail in the film industry are relatively similar to the ideas of the superstar effect. Music and films both have a social aspect, which makes consuming the same as others more valuable and heterogeneity in quality leads to good titles outselling bad titles (Elberse & Oberholzer-Gee, 2007; Elberse, 2008). A different explanation for the lack of a long tail in the film industry is provided by Kumar, et al. (2014). The authors suggest that incomplete information regarding movie quality plays a role in the skewness of sales that can be found in the film industry. Their research results show that the broadcast of a movie on a pay-cable channel leads to higher DVD sales for movies that are relatively less popular. They argue that this result is caused by poorly informed consumers. As movies are a classic experience good, the quality and pleasure one derives from a film can only be evaluated after watching it. Thus watching a film on a pay-channel where there are no per-item costs, consumers become better informed. During the theatrical window only a small amount of films are available and as a result producers focus their promotion this small group.

The impact of streaming

The concept of streaming already appeared in the 1990’s, yet over time what is considered streaming has slightly differed. When it was first introduced the term streaming was merely used to describe the technology that allowed people to play media from home without needing to completely download the requested file beforehand. Instead streaming allowed consumers to watch or listen to content at the same time as it arrives. Moreover the streamed content is discarded after playing, therefore no actual content remains on the device and the content cannot be duplicated without authorization. (Krikke, 2004). However currently streaming usually refers to a media service which allows consumers on-demand access to a large media catalogue (Herbert, et al. 2018). Sometimes streaming content in the film industry is therefore also referred to as Video-on-Demand (VOD). Streaming technology and the introduction of streaming services has had significant consequences in the entertainment industry (currently it’s mainly used in the music, film and television sectors), nevertheless the impact has not been the same for every segment (Herbert, et al. 2018).

In the music industry the introduction of streaming has been the most impactful. With Spotify as frontrunner, this disruptive technology has had significant effects on the business models and revenue streams in the music industry. Streaming is the dominant form of consumption in the current day and age, yet only in recent years has it helped the music industry turn from a long period of declining revenue numbers into a period of growing revenue. 2016 was the first year since 1998 that the overall music industry in America recorded double digit growth and this trend continued in 2017 (Christman, 2017, 2018). The increasing revenue from paid subscription in the market seems to be mainly responsible for this growing trend. According to The RIAA paid subscriptions were responsible for 17.2% of the total revenue in 2015, the following year this was already 29.2% and in 2017 39.5% of all revenue in the music industry was accrued by paid subscriptions (RIAA, 2020). Earlier evidence for dropping revenue numbers in the music industry as a result of the characteristics of digital consumption were already presented by Elberse in 2010. More recently Naveed, et al. (2017) show that live music and live music revenues have also surged again. They suggest that the increasing trends in both digital consumption and live consumption could be related. Moreover they argue that this development could be the music industry’s answer to the decreasing revenue from physical sales.

The impact of streaming services has also grown in the film and television industries, yet streaming services are still not as dominant in these specific industries. As most of the major streaming services provide both film content and television content, there is a degree of overlap between the media forms in terms of streaming. Nonetheless the impact of streaming on these industries still differs somewhat. Herbert, et al. (2018) suggest that this difference is partially due to the influence of major institutions in the film industry:

> Nevertheless, movies and television remain distinct industrially, as the media conglomerates that commonly own them continue to differentiate these two media divisionally, and related institutions, such as the Motion Picture Association of America and the Federal Communications Commission, distinguish film from television in a variety of ways. Partly as a consequence of these industrial, institutional definitions, streaming has likewise affected each industry somewhat differently. (p.352).

Additionally the authors claim that the initial relationship between Netflix and the film studios was more hostile than the relationship between Netflix and the television studios. Film studios were afraid that the subscription model of Netflix, with low costs per additional film watched, would lower the price consumers would be willing to pay to rent or buy an individual movie. When Netflix started to offer their streaming service to customers, film studios were able to negotiate more lucrative deals with Netflix. Yet still they prioritized releasing their films themselves in more profitable release windows. On the other hand television studios did not consider Netflix a competitor. Instead Netflix added a new revenue stream, because no method to monetize easy immediate access to television shows existed yet.

In the television industry the introduction of video on demand (VOD) has had significant effects on the traditional players in the market. Cable companies have had to change their strategy to compete with new entrants such as Netflix (Burroughs, 2019). While the amount of literature on the effects of streaming in the film industry is still somewhat limited, there are some authors who suggest that online distribution and consumption has been a disruptive force in the film industry as well (Cunningham, et al., 2010). Their research tries to answer several key questions related to the rates of change in the film industry due to online distribution. The authors have continually examined the business practices and content of a substantial number of online film distributors in 2008 and 2009. While a number of these distributors were Hollywood based, the online nature of the distributors means that there was no main focus on one particular market for most of their research. However the authors have dedicated a part of their research particularly on the changes in the Australian market. The authors' case study provides several conclusions on the impact of online distribution on the film industry. First of all, the rate of change in online distribution has increased in the latter parts of the 2000's, mainly due to the spread of high-speed broadband. Secondly there are some new players in the market that have started competing with traditional distributors. Moreover these new players have introduced some new business models that the traditional distributors needed to adapt to. These business models include advertising-supported channels and subscription models. The main consequence of these new business models is that new players offered content at lower prices to increase their market share. Traditional Hollywood based distributors presented their content at premium value and thus had to adapt to the lower-price online models. A third conclusion of the paper is the fact that a co-evolution has taken place of the market and non-market sectors. Companies like Youtube have tried to exploit their online presence to compete with traditional distributors by creating and licensing their own high-value content. The general conclusion seems to be that new entrants in the film distribution market have brought along new business models that traditional distributors have had to adapt to. Yet these conclusions are based on data and observations from a time when most streaming service providers were relatively new.

More recent data showed that streaming services had surpassed physical video sales for the first time in 2016, moreover in this year streaming subscriptions amounted to more revenue than VOD sales (Wallenstein, 2017). Other recent insights into the disruptive nature of streaming as a form of online distribution are provided by Yu, et al. (2017). The authors make use of a natural experiment in the U.S market to determine whether streaming media cannibalizes traditional media sales. In 2015 the entertainment network Epix changed their licensing partner from Netflix to the far less popular streaming service Hulu. As Hulu had a far lower market share than Netflix at the time of the switch, the availability of the content of Epix became much lower. This allowed the authors to research the causal effect of the availability of content on streaming services on the physical sales of the content. The results of their research provide academic evidence that streaming services cannibalize physical DVD sales. In the three months after the switch the physical sales of the content of Epix increased by 24.7%. Moreover their results show that the cannibalization effect is especially likely for movies that have been released in stores only recently and for movies that have had strong theatrical performance.

Alternative determinants of box office revenue

The research into determinants of box office revenue consists of a substantial body of literature. As there are many factors that can potentially influence the financial success of a movie, this body of literature consists of research in several different areas.

The superstar effect developed by Rosen (1981), as discussed earlier, can be applied to the success of blockbuster films as products. However films are projects with many contributors, where the individual contributors can be considered superstars as well. Therefore one well-studied determinant of movie success is the influence of star actors. The body of literature on the influence of star actors does not provide conclusive results. Ravid (1999) shows that stars do not influence the box office success of a movie, however bigger budgets do. He argues that higher budgets will lead to higher revenues, regardless of whether this budget is spent on star actors or not. The inclusion of star actors is generally merely an indication that the budget of a film is sizeable. Additionally De Vany and Walls (1999) find that only a very small amount of stars have a real impact on the chance that a movie becomes a hit. Still they argue that the inclusion of a star in a movie is far from a guarantee for success. On top of that, causality might even be reversed, actors who are deemed superstars often receive more lucrative and better offers and therefore their chances of starring in a successful movie can be higher. On the other hand, there is also academic evidence that stars do have a strong impact on the financial success of a movie and might be worth millions of dollars in revenue (Elberse, 2007; Wallace, et al., 1993). Still the impact might not be the same for all stars, as some studies show that age and gender significantly impact the influence of a star on movie success (Treme & Craig, 2013).

One other determinant that has become increasingly more important in recent years is the impact of social media. Social media has allowed consumers to easily share their opinions and sentiments with everyone in the world. As a result information about products, such as quality, is being spread much faster than before the presence of social media. The concept of online word-of-mouth (WOM) and its impact on product sales has therefore been studied quite extensively in recent years. Research has shown that online WOM and box office revenue are quite closely related. Box office revenue and the valence (degree of positivity of reviews) of online WOM can positively influence the volume (number of user reviews) of feedback online. In turn the increased volume can lead to higher box office revenue again (Duan, et al., 2008). Moreover the average online rating of a movie can also directly influence the box office success of a movie, as WOM valence seems to have an effect on box office revenues as well (Chintagunta, et al., 2010). Even though research has shown that the online sentiment, in the form of the valence as well as the volume of user reviews, play an important role in the decision-making process of consumers, the opinions of experts cannot be completely disregarded. Kim, et al. (2013) show that both online WOM, in terms of volume and valence, and reviews by experts are significant factors of box office success in the U.S domestic market. On the other hand only the frequency of online WOM influences the success of a movie in the international market, while expert reviews have no discernable impact. This suggests that reliance on online WOM might be more important for the international market than the North American market. Contrasting these results are the findings by Basuroy, et al. (2020). They show that expert reviews have a larger influence on box office success than online WOM. Moreover they find that the influence of expert reviews depends on the degree of agreement between experts and the average consumer. The impact of the opinion of experts is stronger when experts and consumers share a similar sentiment about the film.

Social media has also allowed movie studios to use more direct and less costly marketing strategies. Furthermore social media can provide studios with simple metrics that quickly show the sentiment and anticipation of consumers. One example of such a metric is the number of likes on Facebook. Likes on Facebook prior to the release of a movie can show the preference of consumers and are able to tracked in real-time (Ding, et al., 2017). According to these authors the number of pre-release likes on Facebook has a significant positive relationship with the box office revenue of a film. More specifically each 1% increase in the number of likes before the release of a film is associated with an increase of 0.2% in box office revenue. Moreover the authors find that the impact of likes becomes stronger, the closer to the release of the film. Thus the latest likes are more impactful than early likes. Facebook likes are therefore able to easily tell film studios the level of anticipation for a movie. Additionally Oh, et al. (2017) show that online engagement of customers on social media platforms positively relates to the future box office success of a film. Therefore the use of social media platforms as marketing tools should help studios to achieve better economic performances.

Cultural diversity and welfare economics

The relatively substantial body of literature on the long tail phenomenon has so far been mostly focused on determining the effects of the phenomenon and whether it actually exists in all industries. Most of the studies present a descriptive study of these effects and either find evidence for or against a more evenly spread distribution of sales, which would resemble a longer tail. Even though Brynjolsson, et al. (2003) showed that consumer welfare has grown significantly through the increased market efficiency of online channels, few papers have addressed the welfare economic debates regarding changes in diversity due to online availability.

Product variety and diversity play a key role in the welfare economics as one of the basic issues regarding welfare economics is whether the product variety provided by the market is socially optimal (Dixit & Stiglitz, 1977). As the authors also state monopolistic markets would provide positive profits to suppliers, however this would reduce the product diversity and thus would lead to a welfare loss. A monopolistic market is often classified as a market failure, as prices become higher than the appropriate competitive price would be. Therefore an appropriate degree of competition and variety would result in the highest total welfare.

Measuring diversity in the cultural markets, or cultural diversity, can be difficult in and of itself as there are no real established metrics for diversity. Based on literature from the subjects of biodiversity and technological diversity three different dimensions of diversity can be distinguished: variety, balance and disparity (Moreau & Peltier, 2004; see also Benhamou & Peltier, 2007). For cultural products variety generally refers to the number of unique products or titles. Balance refers to the pattern of distribution of these unique titles, were balance is highest if the market share of each title is the same. Lastly disparity refers to the difference between the unique titles, which can be measured by different genres or languages for example. Moreau and Peltier (2004) used these three different dimensions to measure the cultural diversity in the film industry of six different regions in the period 1990-2000.[1] The authors find that the United States and France display the greatest number of variety i.e. number of unique films released. However the leaders in terms of balance, the distribution of market share for each film, are the European Union and France. The authors therefore highlight that measuring cultural diversity heavily relies on which dimensions are used.

The effect of online distribution channels on one of the dimensions of cultural diversity, the degree of balance, is already addressed in the literature regarding the long tail theory. Based on the contrasting theories it is expected that either the long tail theory holds and niche products become relatively more popular, leading to a higher balance of unique products or the converse happens and blockbuster titles become more popular, leading to a more skewed balance. The dimension of balance is also studied empirically by Peltier, et al. (2016). The authors find that the market concentration of books is lower online than offline. Moreover they find that for literature books dominant firms lose a part of their market share to smaller publishers. Therefore showing that in terms of balance, diversity increases by online sales in the book industry.

For the film industry Napoli (2019) suggests that technological innovations have led to a greater diversity of content distribution and content aggregators. Catering to only one particular niche has become less costly, which in turn has also increased the competition of content distributors. Due to low costs of entry there have appeared many content distributors that try to focus on a distinct niche that is not provided by the larger aggregators, such as Netflix. The author states that this development is still somewhat different than the original idea of the long tail by Anderson (2004). Currently no supplier is able to serve the complete long tail of less popular products, instead the long tail is divided among many different competitors that all provide slightly different content. This growing fragmentation can also increase the search costs for consumers, which would diminish the attention on the long tail. While online distribution channels should allow consumers to watch more diverse products, Huffer (2017) finds that film viewers are more likely to watch foreign films trough traditional offline channels, such as cinemas or other public screenings, than through online channels. While the supplied diversity might grow, this could mean that consumed diversity has not necessarily increased.

[1] The regions used in this research are the United States, France, the European Union, Hungary, Mexico and South-Korea.

III. Methodology

Data source

The main data source for this research is Boxofficemojo (IMDbPro, 2021). Boxofficemojo is a subsidiary of IMDd, which is one of the largest online databases regarding film and television information. Boxofficemojo is ran by IMDbPro, which is a subdivision of the IMDB company, that focuses more on providing information to industry professionals instead of the average consumer or film enthousiast. Boxofficemojo tracks the theatrical box office revenues of all films in dollars released in North America, as well as a large number of other countries they receive data from. The website publishes daily box office numbers, but also provides breakdowns per week, month, season or year. Boxofficemojo collects their data from several different sources, such as film studios and distributors. As the site is based in North America it can be expected that their domestic data sources might be more reliable than their international sources. The data collection of Boxofficemojo relies on a variety of sources, such as: film studios, distributors or sales agents and is not gathered first hand by the website. The revenue that is listed on Boxofficemojo only includes revenue that is earned from cinematic releases i.e. the sales of cinema tickets for a film. Thus this is the only form of revenue that is considered in this research. Boxofficemojo is considered to be one of the most reliable free sources of box office numbers and is widely used in academic research on the determinants of box office success (Ding, et al., 2017; Karniouchina, 2011; Treme & Craig, 2013).

The decision to use yearly data is motivated by the efficiency of data collection. The data for this research was taken from Boxofficemojo manually. Even though the usage of monthly data would allow for significantly more observations, it is questionable whether the final results would be different of more reliable. If automated data collection of a downloadable box office database would become available, future research could provide additional analyses using smaller time intervals and thus more data points. One important concern to consider when using quarterly, monthly or weekly data is the concept of seasonality. The concept of seasonality plays a significant role in the motion picture business. Einav (2007) finds that the seasonality effect in the movie business is largely caused by an amplification effect. Movie distributors often release their biggest movies during specific periods, namely during the beginning of summer and during the Christmas holiday season. This strategy increases the demand for movies during these periods where demand already is somewhat larger. Because the amount of movies as well as production budgets of these movies differs between the high-demand periods and low-demand periods, using such data without removing seasonality would lead to biased results.

The website allows for two different viewing options, charts can either be viewed by in-period releases and calendar grosses. Charts that present in-period releases show revenues only from movies released in that certain period and exclude revenues earned in that particular period by movies released outside the period. Charts by calendar grosses show all results within a given time period, regardless of whether these revenues are earned by movies released before this period. This research uses data reported by calendar grosses, as it is concerned with trends over time and not movie-specific trends. By using in-period release data, revenue that is accrued in a particular year could be attributed to the previous year, which would result in an unrealistic representation of a trend over time. An example of the differences between the two options is presented in table 1. As the example shows Rogue One: A Star Wars Story is a higher grossing movie than Finding Dory, yet Finding Dory grossed a higher revenue in 2016. Thus using in-period releases can paint a biased picture of the popularity of a film during a particular year.

This research focuses on the North American market instead of looking at the worldwide movie theatre market. Therefore only domestic data from Boxofficemojo is considered. As Canada is considered domestic as well, this means that the gathered dataset consists of box office numbers from the United States as well as Canada. The North American market is the largest market in the world by a substantial amount and North America is also leading in the amount of produced movies each year. Furthermore the decision to focus on this particular market is made, because of the method used to determine the distinction between the baseline period (pre-period) and the post-period in

Table 1

Comparison between in-period release option and calendar gross option

Movie	In-period release		Calendar gross	
	2016	2017	2016	2017
Finding Dory	486,295,561	0	486,295,561	0
Rogue One: A Star Wars Story	532,177,324	0	408,235,850	123,941,474

Source: Boxofficemojo

Notes: Revenue numbers are presented in dollars. Revenue numbers are not adjusted for inflation.

this research. As the identification for the distinction between these two periods the introduction of the streaming service by Netflix is used. Netflix can be seen as the first major streaming service provider and has been considered leading in the rise of the streaming technology (Burroughs, 2019). Thus the pre-period consists of the years before Netflix introduced their streaming service and the post-period includes all years after the launch of their streaming service. However Netflix did not introduce this service worldwide immediately, instead opting to spread their service gradually to other markets than the United States. For the purpose of this research there is no definite point in time that would signal a pre-streaming period and a post-streaming period worldwide, meaning it would be impossible to create a pre-period and a post-period for a study based on global data. This distinct point in time does exist for the United States and thus this point can be used to determine the cut-off point of the pre-period. Netflix first introduced their streaming service in the United States in 2007. Therefore this year is considered as the first year of the post-period. Netflix only introduced this service in Canada in 2010, however the Canadian market is considerably smaller than the United States' market. Thus the Canadian box office numbers will only be a marginal portion of the total domestic data in the sample. Therefore it can still be expected that the results will present a reliable picture of the North American market.

Noteworthy is the fact that the adoption of streaming services by consumers has evidently been a gradual process. Netflix reported 7,479,000 total subscriptions at the end of 2007, at this point focusing their services on the North American market (Netflix, 2008). At the end of the sample period the popularity of Netflix has grown significantly. Netflix reported a total of 67,662,000 paid memberships in the United States and Canada alone at the end of 2019 (Netflix, 2020). It is therefore reasonable to assume that the impact of streaming services has been a gradual process and the influence of streaming was perhaps not as strong in 2007 as it is now. To test whether the influence of streaming is similar over time, additional robustness checks are provided in appendix C.

Additionally data for the market share of film distribution companies is taken from The-Numbers (Nash Information Services, 2021). This is another website, owned and operated by Nash Information Services, LLC., that tracks box office data, both domestically and internationally. Similar to Boxofficemojo domestic data on this site includes Canada and calendar grosses are provided as well. However The-Numbers provides more free in-depth information about the film business, such as distributor related data. Boxofficemojo does provide more detailed industry data and distributor data to subscription members of their service IMDBPro. As the market share data gathered from The-Numbers is sufficient for the analysis regarding distribution studios, the decision is made to gather and use the free data instead of the data behind a paywall.

Both box office trackers gather data separately which results in some small differences between the reported revenues on the websites. However for most of the movies the revenues reported are similar and generally most of the differences are negligible in terms of size. The fact that both box office trackers present more or less similar results for each movie also points to the strength of the dataset, as it's supported by multiple different sources . A comparison between the reported revenue numbers by Boxofficemojo and The-Numbers is presented in table 2. The differences in reported revenue for the titles Ralph Breaks the Internet and Mary Poppins are likely to be caused by delayed or premature reporting of revenues during December and January. This seems to be the case as the differences in returns are reversed in the following year, when Boxofficemojo reports slightly higher revenues equal to the difference in reporting in 2018, for both movies than The-Numbers. As the total reported revenues for these movies are similar The-Numbers seems to be a reliable additional source of box office revenue. Moreover these difference in yearly number are small enough compared to the total market size that they are not likely to influence the final results.

Data sample for the concentration of individual movies

The data sampled from Boxofficemojo for this research consists of all movies that were released and generated box office revenue during the period of 1996 until 2019 in North America, a total of 24 years. The number of observations in the dataset is 16,076. As Boxofficemojo only lists movies that have generated box office revenue during the year, this dataset does not contain any observations with revenue of zero. Thus all observations can be used and there is no missing data. Avengers: Endgame accounts for the highest observation in the dataset, with a gross revenue of 858,373,000 dollars in the year 2019. The lowest observation in the dataset is the movie Zyzzyx Rd, which only grossed 30 dollars in 2006. Because the data is taken from the calendar gross chart instead of in-period release, this does not mean that 16,076 unique films were released in theatres during this period.

Table 2

Differences in reported revenue for Walt Disney films in 2018

Movie	Boxoffice revenue	
	Boxofficemojo	The-Numbers
Black Panther	700,059,566	700,059,566
Avengers: Infinity War	678,815,482	678,815,482
Incredibles 2	608,581,744	608,581,744
Ant-Man and the Wasp	216,649,740	216,648,740
Solo: A Star Wars Story	213,767,512	213,767,512
Ralph Breaks the Internet	177,616,854	187,152,693
Mary Poppins Returns	105,930,461	138,817,262
Star Wars Ep. VIII: The Last Jedi	102,963,014	102,963,014
A Wrinkle in Time	100,478,608	100,478,608
Christopher Robin	99,215,042	99,215,042
The Nutcracker and the Four Realms	54,581,769	54,785,758
Coco	29,891,816	29,891,816
Thor: Ragnarok	3,833,139	3,833,139
Total		

Sources: Boxofficemojo and The-Numbers

Notes: Revenue numbers are presented in dollars. Revenue numbers are not adjusted for inflation.

The sample period contains 13,380 unique movies. Some movies generated revenue in multiple years and thus occur more than once in the dataset. The most obvious reason that certain films are present more than once in the dataset is because they are released later in the year and therefore continue their theatrical run in the next year. One other reason why a particular film might appear more than once is the occasion of a re-release.

An initial overview of the yearly data is presented in table 3. The data in table 3 shows an impressive growth of the film industry in the last twenty years. During the years of the sample period the revenue earned per year has nearly doubled. The table shows that the number of movies screened in cinemas each year has increased with nearly 200% as well. This might be a reason for the increased revenues, however it can also be seen as a result of a possible growing demand for movies. The clear increase in the number of movies that are being shown each year shows that the absolute long tail has grown. Currently more movies reach an audience and are able to accrue revenue than in the past. However the data in the table does not provide any useful information on the relative long tail.

The absolute data for individual movies is not as relevant for this research, as the long tail theory primarily is concerned with the distribution of products instead of absolute product sales. As concentration of revenue is more important than absolute revenue for this

Table 3

Summary of the total box office revenue per year and highest grossing film of the year

Year	Number of films screened in cinema	Total box office revenue	Highest grossing movie of the year	Revenue of highest grossing movie
1996	306	5,648	Independence Day	306
1997	310	6,156	Men in Black	251
1998	334	6,726	Titanic	488
1999	448	7,378	Star Wars: Episode I -The Phantom Menace	430
2000	439	7,512	How the Grinch Stole Christmas	252
2001	413	8,111	Harry Potter and the Sorcerer’s Stone	289
2002	570	9,166	Spider-Man	404
2003	667	9,211	Finding Nemo	340
2004	700	9,365	Shrek 2	441
2005	676	8,838	Star Wars: Episode III – Revenge of the Sith	380
2006	746	9,209	Pirates of the Caribbean: Dead Man’s Chest	423
2007	775	9,657	Spider-Man 3	337
2008	725	9,629	The Dark Knight	531
2009	646	10,590	Transformers: Revenge of the Fallen	402
2010	651	10,557	Avatar	466
2011	730	10,174	Harry Potter and the Deathly Hallows: Part 2	381
2012	807	10,823	The Avengers	623
2013	826	10,922	Iron Man 3	409
2014	849	10,360	Guardians of the Galaxy	333
2015	846	11,126	Jurassic World	652
2016	856	11,377	Finding Dory	486
2017	852	11,073	Star Wars: Episode VIII – The Last Jedi	517
2018	993	11,889	Black Panther	700
2019	911	11,321	Avengers: Endgame	858

Source: Boxofficemojo

Notes: Revenue numbers are presented in millions of dollars. Revenue numbers are not adjusted for inflation.

research, adjusting for inflation over the sample period is not necessary. Moreover this research looks at a possible trend over time and thus aggregate yearly observations are necessary. Therefore annual measures of concentration for the most viewed films are calculated manually from the dataset. First off, a list containing the top 10 and top 20 films with the most revenue in a year is created. This list is based on the dataset gathered from

Boxofficemojo. Focusing on only the best performing movie of the year would probably lead to less reliable results as the use of only one movie is likely to produce more outliers. Still an analysis of the number one movie of the year is provided in appendix D as an additional robustness check.

As mentioned earlier it is possible that certain films earn revenue in multiple years, most often if they are released in the last months of the year. One concern could be that movies that are very popular at the end of the year will miss out on being on the top 10 or top 20 list, because their revenue is split between two years. However according to de Vany and Walls (1999) most movies earn their money in the first weeks after the initial release and only a small number of movies have long 'legs', meaning they earn a reasonable amount of revenue after the first weeks of release. They also state that the maximum amount of revenue for 65% to 70% of all movies is earned in the first week and that on average 60% of revenue for a film is earned in the first three weeks. Therefore it is quite unlikely that a film will miss out on being a top earner in a year despite being one of the most popular films. On the other hand there is a small number of films with such extraordinary 'legs' that they are present in the top lists of multiple years. In this sample there are 13 films that appear on the top 20 list for two successive years and of these films only six films manage to make the top 10 list on two occasions. No additional adjustments are made for these films, as this research is concerned with the concentration of the top films over time. Therefore the demand and following revenue for every film should be measured in year it was accrued. This is consistent with other researches on the long tail that do not make adjustments for products that are popular in multiple years (Elberse & Oberholzer-Gee, 2007; Peltier & Moreau, 2012). Certainly the availability of films in cinema is a shorter period than the availability of the physical products used in the aforementioned researches, books and DVDs. Still the decision is made to remain consistent with existing literature in regards to this aspect of the methodology.

After creating these lists the percentage of total revenue earned by the top 10 and the top 20 most viewed films per year is calculated to create a measure of the concentration of the most popular films. This is simply done by dividing the earned revenue of the 10 and 20 most viewed films of the year by the aggregate box office revenue of all released films in North America in that same year. This step results in one observation per year for top 10 films as well as top 20 films, thus in total 24 observations for both variables. The annual percentage of revenue earned by these subgroups of top films as well as the mean and standard deviation for these groups are presented in table 4. A visual representation and the analysis of this data is provided in the results section.

Data sample for the concentration of film distribution companies

Additionally this research also uses a small sample of data gathered from The-Numbers to analyse possible trends in the concentration of sales among film distribution companies. This small sample consists of the market share data of the six biggest distribution companies over the sample period of 1996 to 2019. The selection of these six companies is based on their

Table 4

Annual percentage of revenue earned by the top films

Year	Top 10 films	Top 20 films
1996	27.69	41.78
1997	24.93	38.53
1998	28.12	44.07
1999	28.00	42.30
2000	23.04	37.69
2001	26.02	41.43
2002	25.69	40.38
2003	24.51	38.04
2004	26.79	40.01
2005	25.77	40.38
2006	23.32	35.86
2007	26.99	41.41
2008	25.96	39.29
2009	25.42	40.14
2010	29.42	42.87
2011	23.98	38.29
2012	29.75	44.25
2013	27.03	42.17
2014	24.14	40.88
2015	34.64	50.34
2016	32.34	47.86
2017	31.73	47.98
2018	32.60	47.98
2019	38.65	52.71
Mean	27.77	42.38
SD	3.78	4.16

consistently strong market share during this period. Based on their considerable market share and their key role in the industry, these six companies are considered the major studios and are often referred to as the ‘majors’ (Eliashberg, et al, 2006). The six selected distribution companies all had an average market share of at least 10% over the total sample period from 1996 until 2019. Therefore the distributors considered in this research are: Walt Disney, Warner Bros., Sony Pictures, Universal, Paramount Pictures and 20th Century Fox.[2] In 2019 20th Century Fox was acquired by Walt Disney, reducing the number of leading distributors to five. To account for this acquisition the market share of 20th Century Fox is added towards the total market share of Walt Disney for this year. Consequently the yearly market share data of 20th Century Fox consists of one less observation than the other distributors in this sample. The combined market share as well as the individual market share of these six major distributors is presented in table 5.

This research focuses on the performance of film distributors instead of film producers because of two reasons. First of all according to Eliashberg, et al. (2006) one key metric of success for distributors is the box office results. The importance of box office revenue is part of the general emphasis of revenue over profits in the film industry. However box office revenue is a less reliable metric of success for production companies. While distribution is generally done by one particular company, production of new films can be a joint project between multiple different production companies. In some cases major studios are responsible for a larger share of the funding that the smaller firms that take a production credit. Moreover not all production companies distribute their own movies. The second reason is related to data availability. Most box office trackers provide the distribution company responsible for the distribution of a film, yet production credits are not always provided. In the case of joint production it can be even more difficult to find out which companies can take credit for the film.

[2] The-Numbers lists the market share of subsidiaries separately from their parent companies. For the sample of this research subsidiaries that are fully owned by their parent company i.e. are not owned by multiple different distributors, are included in the market share of their parent company. This means that Warner Independent is included in the market share of Warner Bros, Paramount Vantage is included in Paramount Pictures, Sony Pictures Classics is included in Sony Pictures and Fox Searchlight is included in 20th Century Fox.

Table 5

Annual market share for the six major distributors

Year	Combined	Walt Disney	Warner Bros.	Sony Pictures	Universal	20th Century Fox	Paramount Pictures
1996	82.06	20.76	15.77	11.33	8.36	13.00	12.84
1997	79.36	13.93	10.55	20.31	9.67	10.91	13.99
1998	70.85	16.38	11.20	10.78	5.92	11.12	15.45
1999	75.71	16.95	14.21	9.06	12.73	11.17	11.59
2000	71.07	14.75	11.80	9.73	14.22	9.93	10.64
2001	69.88	10.87	14.86	10.47	11.44	10.95	11.29
2002	70.31	12.84	11.60	17.21	9.67	11.57	7.42
2003	71.75	16.56	12.64	13.55	11.82	10.02	7.16
2004	68.37	12.46	13.42	14.50	9.76	11.73	6.80
2005	75.83	10.45	16.86	10.99	11.28	16.54	9.71
2006	83.91	16.20	11.87	19.22	8.79	16.99	10.84
2007	81.72	14.41	14.77	13.34	11.28	11.91	16.01
2008	83.35	10.26	18.34	13.41	11.02	12.98	17.34
2009	84.23	11.33	20.03	14.07	8.44	16.03	14.33
2010	84.77	13.98	17.97	12.90	8.45	14.80	16.67
2011	83.99	12.04	17.98	13.45	10.08	11.11	19.33
2012	77.67	14.25	15.28	16.86	12.45	10.43	8.40
2013	77.06	15.97	17.07	11.25	12.83	10.81	9.13
2014	82.69	15.59	15.21	12.69	10.74	18.65	9.81
2015	85.29	20.98	14.06	9.19	22.04	12.75	6.27
2016	85.01	26.07	16.95	8.38	12.61	13.51	7.49
2017	81.72	21.74	18.51	9.91	13.77	12.96	4.83
2018	85.50	26.24	16.18	11.49	14.80	10.35	6.44
2019	80.98	38.31	13.95	12.13	11.58		5.01

Notes: all numbers in the table are percentages.

Empirical approach

To assess whether the introduction of streaming services as a form of e-commerce has had an effect on the concentration of revenue earned by the most popular films a time-series analysis is conducted. Even though there are authors who suggest that visual judgements are not necessarily worse than statistical judgements (Baer, 1977), most authors suggest that it is in fact beneficial to make use of statistical time-series analysis to provide more accurate results (Jones, et al., 1978; McCain & McCleary, 1979). One of the more common methods for time-series analysis is the auto-regressive integrated moving average approach. However one of the limitations of this method is the need for a large amount of data points. A frequent recommendation is to use at least 50 to 100 observations per experimental phase

(Hartmann, et al., 1980). As this research uses a sample with significantly less data points the use of this method is not advisable. Therefore this research uses an intervention analysis developed by Tryon (1982). This method is specifically designed for short time series. As Tryon states the main answer this method provides is whether a time-series contains any statistically significant trends. The use of this method requires two basic conditions to be met (Handke, 2012). First of all the nature of the research subject should resemble a natural or quasi experiment where the expected relationship between the independent and dependent variable should be strong. Secondly the amount of observations per period should not exceed 49, otherwise more common and more complex statistical methods would provide more thorough results.

Streaming technology has had a significant impact in several segments of the entertainment industry (Elberse, 2010; Herbert, et al. 2018, Burroughs, 2019). In the film industry the impact of streaming services and digital consumption can be felt more and more as well (Yu, et al. 2017). The new form of consumption has been called a disruptive technology, which has changed both the demand and supply side of the market (Cunningham, et al. 2010). The introduction of streaming therefore can be seen as an exogenous shock to the film industry that created a baseline period and a post-streaming period. This shock occurred naturally throughout the whole United States at the same time. Therefore the nature of this research does resemble a quasi-experiment. Still as mentioned earlier it is likely that the influence of streaming is gradual, as the number of paying subscribers of Netflix has been growing over the years after the initial launch. Therefore it is expected that the treatment is gradually transpiring and does not resemble a perfect natural experiment. The nature of the research further deviates from a true perfect experiment, as there is no real random assignment of participants. It is likely that consumers who would subscribe to Netflix generally are more interested in films and the film industry. Therefore these consumers might be more interested in cinema visits or information regarding films as well. The results of this research are thus likely not based on a perfect random assignment of consumers. Furthermore as stated before this research uses a sample with only 24 observations. Thus the two basic conditions of the approach by Tryon (1982) are met.

The C-statistic

The intervention analysis developed by Tryon (1982) makes use of the C-statistic, which was originally established by Young (1941). The formula of this C-statistic is presented in equation 1. The C-statistic compares two different measures of variance to determine whether there are any systematic departures from random variation in a time-series. The first measure that is used includes the mean of the sample. This measure is used as the denominator in the formula. The second measure of variance is differs in the sense that it does not include the mean in the calculation. This measure is used as the numerator in the formula.

$$C = 1 - \frac{\sum_{i=1}^{N-1}(X_i - X_{i+1})^2}{2\sum_{i=1}^{N}(X_i - X)^2} \qquad (1)$$

Tryon (1982) states that the standard error is completely reliant on the sample size that is used. This standard error Sc is presented in equation 2. As the standard error is completely dependent on the sample size, the possibility exists to reduce the error to nearly zero by simply increasing the sample size. Thus with too many data points, the results of the test will always be significant. Therefore more complex time series analyses will provide more reliable results if the sample consists of more than 50 observations.

$$Sc = \sqrt{\frac{N-2}{(N-1)(N+1)}} \qquad (2)$$

The Z-statistic can be calculated easily using the formula shown in equation 3. The null hypothesis for this method is that the observations in a time series vary randomly around their mean. A statistically significant result signals that the data points in the time series follow a systemic departure from random variation i.e. a trend or a slope. If a significant result is found for the baseline period of the sample, this means that a trend is already present in the baseline data. To properly analyse the two different periods, it is required that the pre-period does not contain a trend. Thus if a trend is present in the pre-period, detrending the dataset is a necessary step before being able to properly compare the two periods with the intervention analysis.

$$Z = \frac{C}{Sc} \qquad (3).$$

The Gini coefficient

As an additional tool to measure the inequality of sales and provide a clearer picture of the complete sales distribution, the Gini coefficient is used. The preceding analysis of the top performing movies focuses on the changes in revenue of a consistent amount of movies, the top 10 and top 20 of each year. Therefore this analysis provides results on the absolute best performing films of the year, the biggest winners of each year. Contrastingly the Gini coefficient is a relative measure and is based on revenue earned by different quantiles. This means that the Gini coefficient adapts to changes in size of the population. If the number of movies released per year increases, evidently a certain top quantile of best performing movies now includes more movies as well. Thus one of the advantages of using the Gini coefficient to analyse the long tail is the fact that it focuses on the relative share of sales and thus provides a more nuanced depiction of the concentration than an absolute metric (Brynjolfsson, et al., 2010).

The Gini coefficient is a measure often used by economists to describe the inequality of income and wealth distribution. In recent years the Gini coefficient has also been used in research on the long tail to determine the inequality of the sales distribution (Brynjolfsson, et al., 2011; Hinz, et al., 2011; Oestreicher-Singer & Sundararajan, 2012). The Gini coefficient is based on the Lorenz curve and can be seen as a summary statistic of this curve (Dorfman, 1979). The Lorenz curve is a graphical representation of the distribution of wealth or income and graphs percentiles of a population against their cumulative percentage of income of the total population. The line of perfect equality is often depicted graphically alongside the Lorenz curve, this diagonal linear line represents the distribution if the percentage of income is exactly equal to the percentage of the population receiving this income. The Gini coefficient can be calculated as by dividing the area between the line of perfect equality and the Lorenz curve by the total area to the right of the line of perfect equality. The Gini coefficient can range from 0 (perfect equality) to 1 (perfect inequality). Interpretation of the Gini coefficient can be difficult and requires caution. This is due to the fact that the Gini coefficient is merely a coefficient that describes the properties of a diagram. However the diagram it describes is not based on a specific model or distribution process (Fellman, 2018). Additionally there are many different shapes of the Lorenz curve that would result in the same coefficient. As a result of these constraints analysing the changes of the coefficient statistically is difficult. Higher coefficients do signal that inequality has increased, however they do not provide any insight into how much the inequality has increased specifically or how the distribution has changed. The use of a time series analysis on the Gini coefficient is therefore not a common practice. To overcome these difficulties previous researches on the long tail have used more complex regression analyses to test the changes of the Gini coefficient and equality over time (Hinz, et al., 2011; Oestreicher-Singer & Sundararajan, 2012). However due to the limitations of the dataset used in this research, this option is not very feasible. Among the most popular and most-cited papers in the body of literature on the long tail, other options to measure changes in the Gini coefficient over time could not be found for the purpose of this research. It is still possible to derive some additional insights into the concentration and distribution of sales based on visual and non-statistical analysis of the Gini coefficient and Lorenz curves (see also Brynjolfsson, et al., 2011). Therefore an analysis of the Gini coefficient is presented in section IV. The results provided by this analysis can be useful to understand the development of the concentration of sales more completely, however they will not provide any statistical confirmation of changes resulting from the introduction of streaming.

IV. Results

Intervention analysis of the top films

In figure 1 the percentage of the total revenue in a year earned by the top 20 most viewed films and the top 10 most viewed films is presented. The dotted vertical line represents the distinction between the pre-period and the post-period. The pre-period includes 11 observations from 1996 until 2006. The post-period consists of 13 observations from 2007 until 2019. From visual inspection of figure 1 it seems that there is not a very pronounced trend in the pre-period and if there were it seems that this trend would be downward. In the post-period there does seem to be an upward trend present, especially in the latter half of this period. This trend seems to be present for both the films in the top 10 as well as the films in the top 20. Moreover both lines follow a relatively similar pattern, suggesting both the top 10 and the top 20 follow the same trend. The difference between the last observation for the pre-period and the last observation of the post-period is 15.4 percentage points for the top 10 films. For the top 20 this difference is 16.9 percentage points. Still this means that the increase from the last pre-period observation to the last post-period observation for the top 10 is actually larger than for the top 20 with 65.7% compared to a 47% increase for the top 20.

There is no visual indication of an existing trend for the pre-period and according to the C-statistic in the analysis performed later on the pre-period does not show any

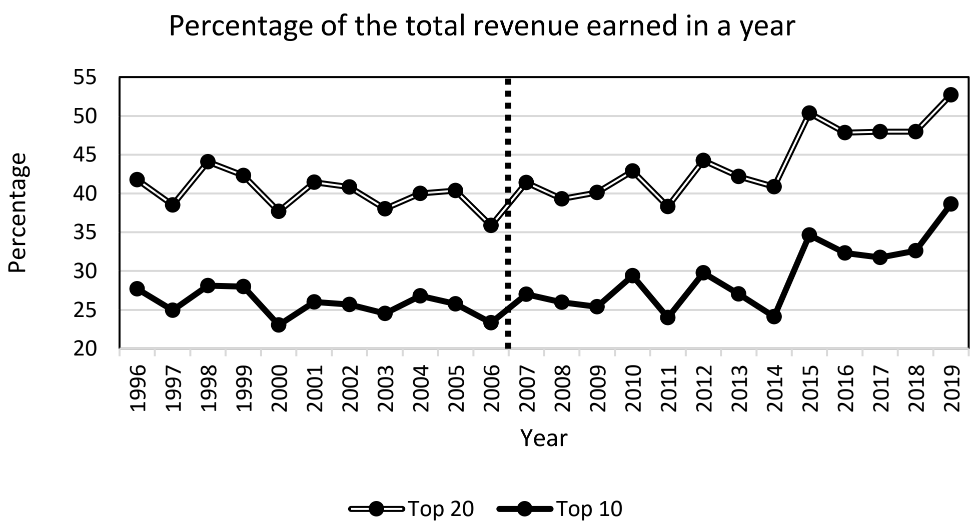


Figure 1.

significant deviations from the random variance, meaning the pre-period shows a stationary state. Therefore it is not necessary to de-trend the baseline data, before applying the C-statistic to the treatment series (Tryon, 1982).

In table 6 the results of the intervention analysis for the top 10 films are presented. The results in table 6 show that the Z-score for the post-period is significant at a 5% level (1.82).[3] Therefore the null hypothesis that there is no significant difference in the behaviour of the time series after the intervention can be rejected. The same tests are performed for the top 20 most viewed films in a year. The results of this test are presented in table 7. Similar to the top 10 there does not seem to be a trend in the pre-period based on visual inspection, on top of that the Z-score of (0.02) is not significant either. The analysis of the post-period actually provides even stronger evidence of an existing trend as the Z-score is significant at 1% level (2.41). Therefore in this case the null hypothesis can again be rejected. These results show that in contrast to the static pre-period, a significant upward trend is present in the post-period. This suggests that the treatment had a significant impact on the percentage of revenue that is earned by the top 10 films as well as the top 20 films.

Intervention analysis of major distribution companies

As an additional test for a possible trend in the concentration of box office revenue, the market share of the major distribution companies will be analysed. Instead of looking at individual movies as the product of the superstar theory, this analysis will look at the role of distribution studios as potential superstars. As stated before the sample of this analysis consists of the six major distribution studios in North America, being: Walt Disney, Warner Bros., Sony Pictures, Universal, 20th Century Fox and Paramount Pictures. The combined market share of these six distributors during the sample period is presented in figure 2. The dotted line represents the introduction of Netflix in the United States, which is taken as the split between the pre-period and the post-period. A first visual inspection shows that even though the market share has been relatively volatile over the years, the difference between

[3] The critical values for the Z-score are taken from Young (1941). A full table with all critical values for this test is provided in Appendix B.

Table 6

Intervention analysis for the top 10 films

	Year	Percentage of revenue earned by the top 10 films	D^2	
Pre-period	1996	27.69	7.60	Pre-period:
	1997	24.93	10.13	D^2 = 64.99
	1998	28.12	0.01	2SS(X) = 62.02
	1999	28.00	24.66	(The mean for the pre-period is 25.81)
	2000	23.04	8.88	
	2001	26.02	0.11	
	2002	25.69	1.37	C = -0.05
	2003	24.51	5.18	Sc = 0.27
	2004	26.79	1.03	Z = -0.17
	2005	25.77	6.02	not significant
	2006	23.32	13.51	
Post-period	2007	26.99	1.07	
	2008	25.96	0.29	Post-period:
	2009	25.42	16.00	D^2 = 249.23
	2010	29.42	29.58	2SS(X) = 467.11
	2011	23.98	33.32	(The mean for the post-period is 29.43)
	2012	29.75	7.40	
	2013	27.03	8.37	
	2014	24.14	110.24	C = 0.47
	2015	34.64	5.27	Sc = 0.26
	2016	32.34	0.37	Z = 1.82
	2017	31.73	0.76	significant at 5% level
	2018	32.60	36.56	
	2019	38.65		

Notes: all the numbers in the table are rounded to two decimals. For 11 observations the critical value for the Z-statistic is 2.21 at a 1% significance level and 1.64 at a 5% significance level. For 13 observations the critical value for the Z-statistic is 2.22 at a 1% significance level and 1.64 at a 5% significance level (Young, 1941).

the shares at the first and the final year of the sample is not that sizeable. Perhaps surprisingly the market share of the majors is smaller in 2019 than in 1996, with 80.98% and 82.06% respectively. Visual inspection of the data also reveals that there might be a trend present in the pre-period. This trend seems to follow a u-shape, an initial drop in the combined market share is followed by two years of growth. As a result the combined market share is slightly higher in the last year of the pre-period than the first year of this period, 83.91% against 82.06% respectively. Moreover the C-statistic for the pre-period data also suggests that the observations do not randomly vary around the mean of the pre-period. Thus to successfully apply the intervention analysis is it required to first de-trend the data.

Table 7

Intervention analysis for the top 10 films

	Year	Percentage of revenue earned by the top 20 films	D^2	
Pre-period	1996	41.78	10.53	Pre-period:
	1997	38.53	30.65	D^2= 112.17
	1998	44.07	3.14	2SS(X) = 112.64
	1999	42.30	21.28	(The mean for the pre-period is 40.08)
	2000	37.69	14.03	
	2001	41.43	0.38	
	2002	40.81	7.69	C = 0.004
	2003	38.04	3.87	Sc = 0.27
	2004	40.01	0.14	Z = 0.02
	2005	40.38	20.46	not significant
	2006	35.86	30.88	
Post-period	2007	41.41	4.51	
	2008	39.29	0.72	Post-period:
	2009	40.14	7.46	D^2= 193.31
	2010	42.87	20.97	2SS(X) = 504.12
	2011	38.29	35.45	(The mean for the post-period is 44.32)
	2012	44.25	4.30	
	2013	42.17	1.68	
	2014	40.88	89.57	C = 0.62
	2015	50.34	6.19	Sc = 0.26
	2016	47.86	0.01	Z = 2.41
	2017	47.98	0.00	significant at 1% level
	2018	47.98	22.45	
	2019	52.71		

Notes: all the numbers in the table are rounded to two decimals. For 11 observations the critical value for the Z-statistic is 2.21 at a 1% significance level and 1.64 at a 5% significance level. For 13 observations the critical value for the Z-statistic is 2.22 at a 1% significance level and 1.64 at a 5% significance level (Young, 1941).

Tryon (1982) suggests two possible methods for de-trending the dataset in order to apply the intervention analysis. The first method is to find an appropriate regression technique to secure the best line of fit and from here study the residuals of this regression. The second method is to apply differencing to the dataset. For this research the second option will be used, thus the data is transformed by calculating the first difference for every observation point. The main advantage of this method is that there is no need for additional judgement of regression models and finding the best line of fit (Handke, 2012). Especially for the particular trend that seems to be present in the pre-period of the combined market share, finding an appropriate regression model with a good fit could turn out to be quite complex. As Tryon (1981) states just two atypical data points could severely affect the slope or intercept of the regression and can make the selection process far more complex. One main disadvantage of the differencing method is that there is a loss of observations (Handke, 2012). In this case one observation is lost by the differencing method.

In table 8 the first difference scores and the results of the subsequent intervention analysis are presented. The Z-score of the pre-period is not significant (0.44), which means there is no trend present in the first difference scores and de-trending of the dataset was successful. The Z-score for the post-period is not significant either (-0.61). This means that there is no trend in the first difference scores of the post-period either. As a trend was present in the pre-period, these results are interpreted a little different than in the previous tests. In this case a negative Z-score for the post-period suggests that there is no evidence that the observations in the post-period follow a trend that significantly deviates from the trend in the pre-period. Thus this means that the time series of the concentration of the combined market share in the post-period did not behave differently from the pre-period. This suggests that the treatment had no discernible effect on the combined market share of the six major distribution companies. The inconclusiveness of these results is largely due to the existing pattern in the pre-period.

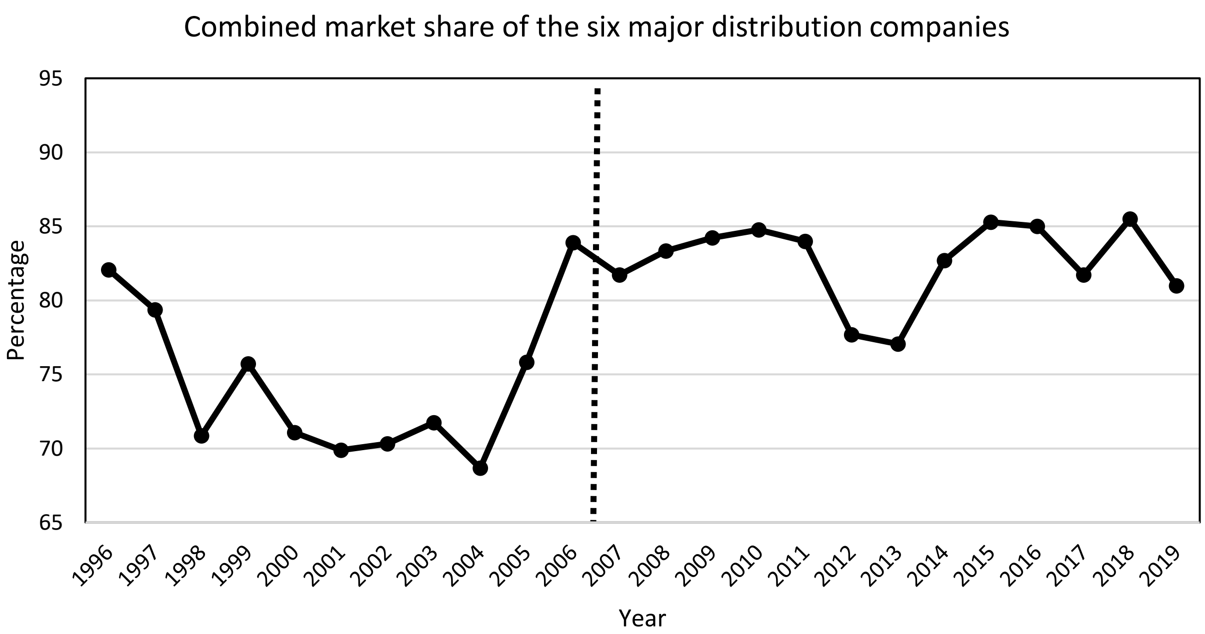


Figure 2.

Table 8

Intervention analysis for the combined market share of the six major distribution companies

	Year	Combined market share	1st difference	D^2	
Pre-period	1996	82.06	-	-	Pre-period:
	1997	79.36	-2.7	33.76	D^2 = 444.44
	1998	70.85	-8.51	178.76	2SS(X) = 508.46
	1999	75.71	4.86	90.25	(The mean for the pre-
	2000	71.07	-4.64	11.90	period is 0.18)
	2001	69.88	-1.19	2.62	
	2002	70.31	0.43	1.02	C = 0.13
	2003	71.75	1.44	20.43	Sc = 0.28
	2004	68.67	-3.08	104.86	Z = 0.44
	2005	75.83	7.16	0.85	not significant
	2006	83.91	8.08	105.47	
Post-period	2007	81.72	-2.19	14.59	
	2008	83.35	1.63	0.56	Post-period:
	2009	84.23	0.88	0.12	D^2= 264.66
	2010	84.77	0.54	1.74	2SS(X) = 228.83
	2011	83.99	-0.78	30.69	(The mean for the post-
	2012	77.67	-6.32	32.60	period is -0.23)
	2013	77.06	-0.61	38.94	
	2014	82.69	5.63	9.18	C = -0.16
	2015	85.29	2.60	8.29	Sc = 0.26
	2016	85.01	-0.28	9.06	Z = -0.61
	2017	81.72	-3.29	49.98	not significant
	2018	85.5	3.78	68.89	
	2019	80.98	-4.52		

Notes: all the numbers in the table are rounded to two decimals. For 10 observations the critical value for the Z-statistic is 2.20 at a 1% significance level and 1.64 at a 5% significance level. For 13 observations the critical value for the Z-statistic is 2.22 at a 1% significance level and 1.64 at a 5% significance level (Young, 1941).

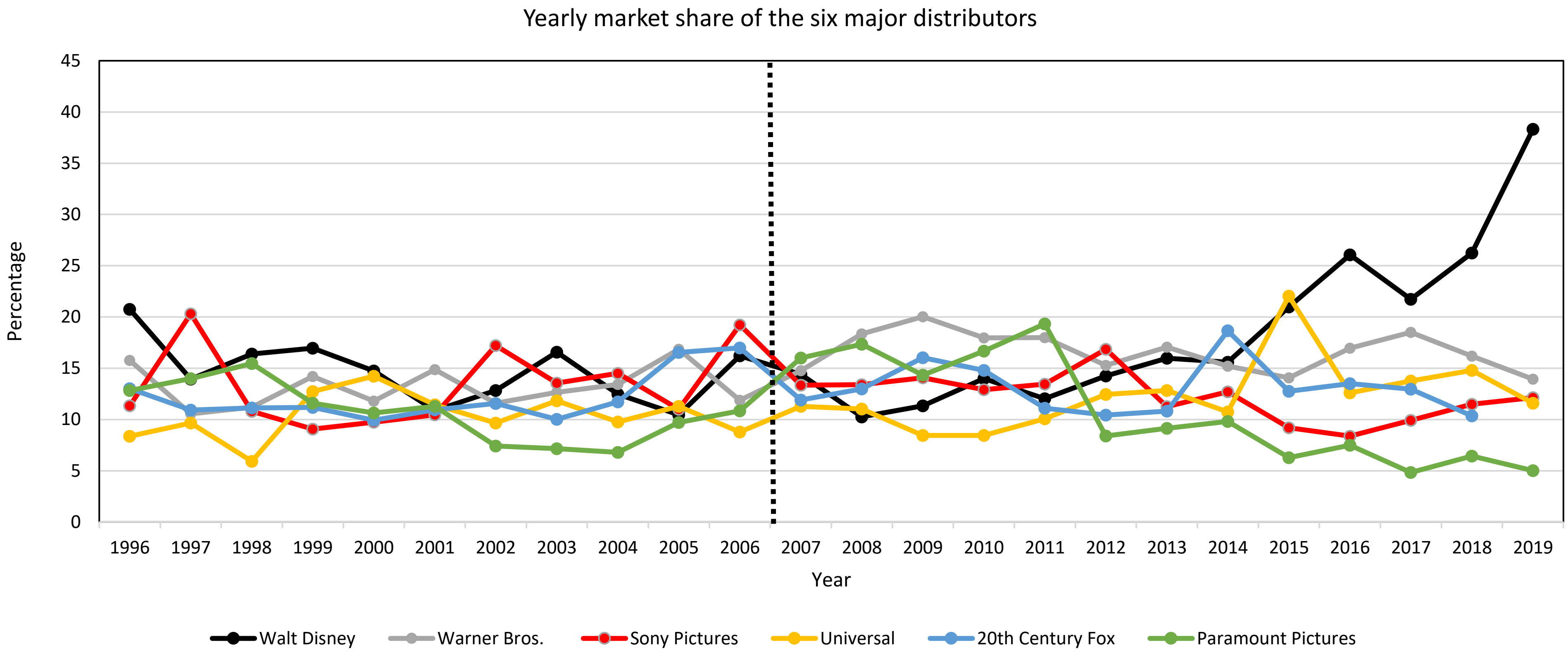
As a follow-up on the analysis of the combined market share, an analysis on the individual market shares of the six major distributors is performed. The market shares of the six majors over the sample period are presented in figure 3.[4] For clarity purposes individual graphs for each major distributor are also provided in figure 4. Based on an initial visual inspection is seems that the post-period has become more volatile than the pre-period. Moreover following the trajectory of Walt Disney in the post-period it seems that a the company has been able to create a significant gap between itself and the rest of the majors.

[4] To be noted that due to the acquisition of 20th Century Fox by Walt Disney, there is no data point of 20th Century Fox in the year 2019. Instead the market share of 20th Century Fox has been added to the market share of Walt Disney for 2019.

It is the only major distributor that has been able to increase their market share by a substantial amount in the post-period, from 14.41% to 38.31%. One likely reason for this spectacular growth of Walt Disney's market share from 2018 to 2019 is the acquisition of 20th Century Fox in 2019. Whereas the observations of four of the major distributors seem to be randomly distributed around the mean, visual inspection also shows that the data of two distributors exhibits a possible trend in the pre-period, namely: 20th Century Fox and Paramount Pictures. The market share of 20th Century fox seems to follow a small positive trend at the end of the pre-period, while the market share of Paramount Pictures seems to be dropping in the pre-period. The C-statistic for the pre-period of both companies provides additional proof that a trend is present. It is therefore necessary to de-trend the data for both these companies before the intervention analysis can be performed. De-trending is done by calculating the first difference scores of the data, similar to de-trending of the combined market share. As the data of the other distributors does not show an existing trend in the pre-period, de-trending is not required for this data.

The results of the intervention analysis for the market shares of all individual distributors are presented in table 9.[5] The results show that a significant difference in behaviour between the pre-period and the post-period is present for only two distributors. The Z-score is significant only for Walt Disney (3.19) at a 1% confidence level and for Sony Pictures (1.8) at a 5% confidence level. Therefore the null hypothesis that there is no difference in the behaviour of the time series after the intervention can be rejected. These results suggest that for these two distributors the treatment had a significant effect, while the treatment had no discernible effect for the other distributors. Based on figure 4, it seems that the market share of Walt Disney follows a positive trend, however the trend that is present for Sony Pictures actually seems to be a negative trend. A pre-period trend was present in the data of 20th Century Fox and Paramount Pictures, however the Z-scores of the post-period provide no proof that the observations in the post-period deviate from these trends. Therefore there is no evidence that the treatment at the point between the pre-period and the post-period had any effect on the development on the market share for both these distributors.

[5] To prevent an excessive amount of tables and results in this section, only the final result i.e. the Z-score of the C-statistic is provided in the table.



Yearly market share of the six major distributors
Percentage
45
40
35
30
25
20
15
10
5
0
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
Year
Walt Disney
Warner Bros.
Sony Pictures
Universal
20th Century Fox
Paramount Pictures

Figure 3.

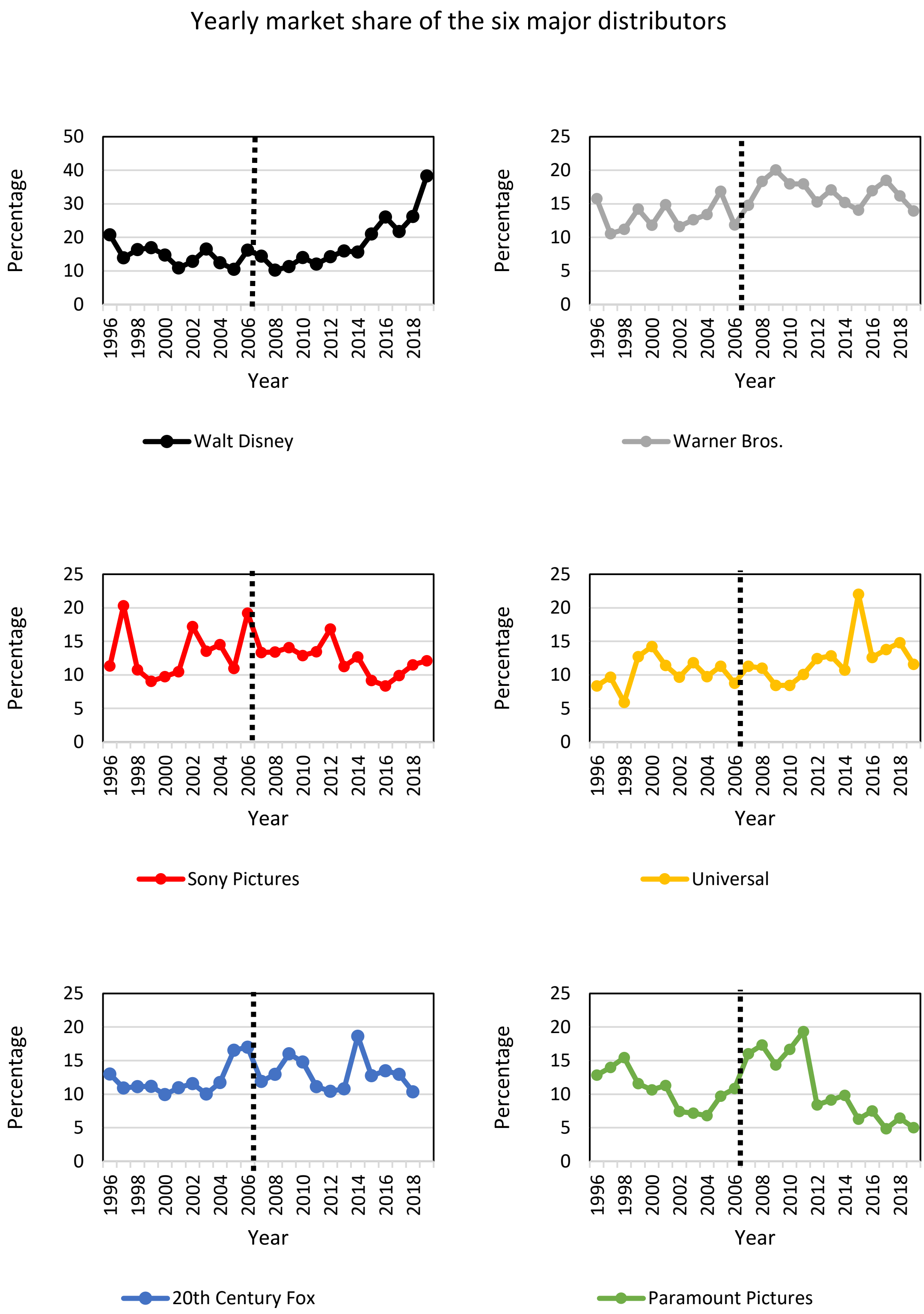

Yearly market share of the six major distributors
Percentage
Year
0
10
20
30
40
50
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
Walt Disney
0
5
10
15
20
25
Warner Bros.
Sony Pictures
Universal
20th Century Fox
Paramount Pictures

Figure 4.

Table 9

Intervention analysis for the major distribution companies

Distribution company	Pre-period Z-score	Post-period Z-score
Walt Disney	0.79	3.19**
Warner Bros.	-0.8	1.36
Sony Pictures	-0.06	1.8*
Universal	0.41	0.53
20th Century Fox	0.65	-0.88
Paramount Pictures	0.07	-0.82

Notes: All the numbers in the table are rounded to two decimals. * denotes significance at a 0.05 confidence level. ** denotes significance at a 0.01 confidence level. As a trend was found in the pre-period for both 20th Century Fox and Paramount Pictures, the intervention analysis for both distributors is performed on the first difference scores instead of the actual market share scores.

Analysis of the Gini coefficient

In table 10 the Gini coefficient of the distribution of revenue for every year in the sample period is presented. Furthermore the cumulative percentages of earned revenue for several quantiles are presented. As the Lorenz curve presents cumulative income earned (or in this case revenue) per quantile, it is common to present information regarding the bottom quantiles of the distribution e.g. the percentage of income earned by the bottom 10 percent. This research is more concerned with the top percentages, therefore numbers regarding the top quantiles are provided for clarity as well. Over the sample period the Gini coefficient has increased quite considerably, from 0.70 in 1996 to 0.92 in 2019. Thus according to this measure of inequality the inequality of the distribution of revenue for films has increased and has become more concentrated over the last 24 years. Furthermore the percentage of earned revenue by the bottom 90% follows a relatively steady decline from 1996 to 2019. Evidently this means that the top 10% of movies has received more and more of the total revenue. This also shows a more concentrated distribution of revenue. Interestingly the cumulative revenue earned by the bottom 10% and the bottom 50% show relatively identical patterns. For both groups the percentage of the total revenue dropped from 1996 to 2006. However in the period 2007 until 2019 an initial decline is followed by an increase in the latest years of said period. As a result the percentage of total revenue in 2019 is similar to that of 2006.

To visually analyse the development of the distribution of box office revenue graphs of the Lorenz curves are presented in figure 5. The yearly changes of the Gini coefficient are

Table 10

Gini coefficients and percentage of total revenue per quantile

Year	Gini coefficient	Bottom 10% quantile	Bottom 50% quantile	Bottom 90% quantile	Top 50% quantile	Top 10% quantile
1996	0.70 (306)	0.11‰ (31)	3.46% (153)	48.10% (275)	96.54% (153)	51.90% (31)
1997	0.71 (310)	0.12‰ (31)	2.97% (155)	50.31% (279)	97.03% (155)	49.69% (31)
1998	0.75 (334)	0.08‰ (33)	1.94% (167)	42.50% (301)	98.06% (167)	57.50% (33)
1999	0.80 (448)	0.04‰ (45)	0.79% (224)	35.31% (403)	99.21% (224)	64.69% (45)
2000	0.78 (439)	0.12‰ (44)	1.02% (220)	40.92% (395)	98.98% (219)	59.08% (44)
2001	0.79 (413)	0.07‰ (41)	0.98% (207)	38.59% (372)	99.02% (206)	61.41% (41)
2002	0.83 (570)	0.03‰ (57)	0.38% (285)	30.78% (513)	99.62% (285)	69.22% (57)
2003	0.85 (667)	0.03‰ (67)	0.26% (334)	26.81% (600)	99.74% (333)	73.19% (67)
2004	0.86 (700)	0.02‰ (70)	0.21% (350)	24.22% (630)	99.79% (350)	75.78% (70)
2005	0.86 (676)	0.02‰ (68)	0.21% (338)	26.95% (608)	99.79% (338)	73.05% (68)
2006	0.85 (746)	0.02‰ (75)	0.19% (373)	27.97% (671)	99.81% (373)	72.03% (75)
2007	0.87 (775)	0.02‰ (78)	0.16% (388)	22.62% (697)	99.84% (337)	77.38% (78)
2008	0.87 (725)	0.02‰ (73)	0.16% (363)	24.82% (652)	99.84% (362)	75.18% (73)
2009	0.86 (646)	0.02‰ (65)	0.16% (323)	24.98% (581)	99.84% (323)	75.02% (65)
2010	0.87 (651)	0.01‰ (65)	0.14% (326)	23.23% (586)	99.86% (325)	76.77% (65)
2011	0.87 (730)	0.02‰ (73)	0.13% (365)	23.05% (657)	99.87% (365)	76.95% (73)
2012	0.89 (807)	0.02‰ (81)	0.10% (404)	18.78% (726)	99.90% (403)	81.22% (81)
2013	0.90 (826)	0.01‰ (83)	0.08% (413)	15.82% (743)	99.92% (413)	84.18% (83)
2014	0.90 (849)	0.02‰ (85)	0.09% (425)	15.97% (764)	99.91% (424)	84.03% (85)
2015	0.91 (846)	0.01‰ (85)	0.10% (423)	13.15% (761)	99.90% (423)	86.85% (85)
2016	0.90 (856)	0.02‰ (86)	0.12% (428)	14.40% (770)	99.88% (428)	85.60% (86)
2017	0.91 (852)	0.02‰ (85)	0.11% (426)	13.12% (767)	99.89% (426)	86.88% (85)
2018	0.92 (993)	0.02‰ (99)	0.16% (497)	11.03% (894)	99.84% (496)	88.97% (99)
2019	0.92 (911)	0.03‰ (91)	0.20% (456)	10.98% (820)	99.80% (455)	89.02% (91)

Notes: all numbers are rounded to two decimals. Earned revenue of the 10% quantile is presented in permillages, while all other quantiles are presented in percentages. The figures in brackets present the number of films included in the quantiles (in case of the Gini coefficient this is the total number of films screened during the year). The figures in brackets are rounded up to the closest full number (in case of an odd number of total films the bottom 50% is rounded up, while the top 50% is rounded down).

relatively small and therefore it is likely that the Lorenz curves of most years will be relatively similar. To limit the amount of unnecessary figures and information and provide clearer results only the Lorenz curves of the first and last year of both the pre-period and post-period are presented. From visual inspection the increase in inequality is fairly clear. Based on visual inspection the difference between the beginning and the end of the pre-period seems to be more substantial than the difference in the post-period. Moreover visual inspection might suggest that the pre-period can be mainly characterized by a decline in the revenue of less popular products, as the tail has become significantly flatter. For the post-period it seems like the tail has remained relatively stable, while most of the change has occurred at the head of the distribution. Yet it is not possible to state whether the treatment had an effect on the development of the concentration of revenue. Additionally the visual analysis is not enough to determine whether any trends are present or statistically significant.

Due to the fact that a time series analysis of the Gini coefficient would likely not provide results that could be reliably interpreted, this research does not produce any statistically significant results based on the Gini coefficient. While the Lorenz curves and the Gini coefficient might seem to suggest that the decline of revenue earned by the absolute tail of the distribution has slowed down in the post-period, this cannot be confirmed statistically. Furthermore they provide no additional insights into the possibility of increased concentration of revenue as a result of the introduction of streaming. The data does show

Lorenz curves of revenue distribution

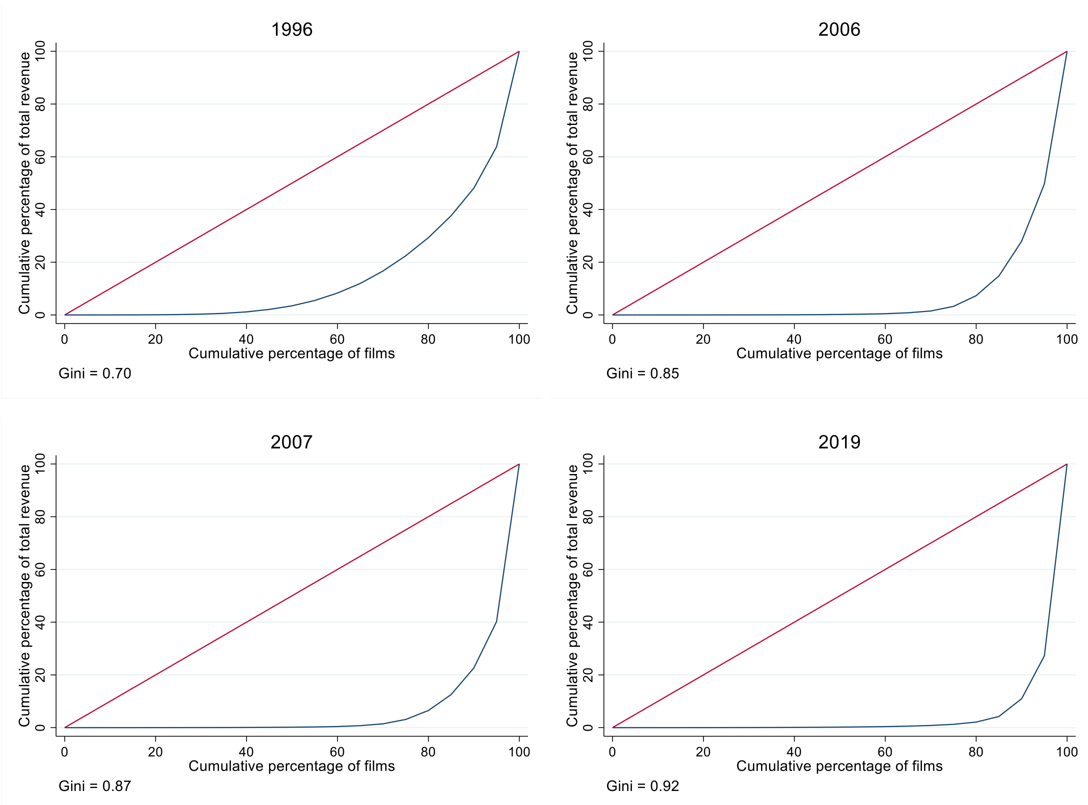


Figure 5.

that distribution of revenue has become more concentrated in the post-period, however it is not possible to state whether this has happened at an increased rate compared to the pre-period.

V. Conclusion

The main aim of this paper is to research how the introduction of streaming has affected the concentration of box office revenue. Streaming is a relatively new form of online consumption, but has become a lot more popular in recent years. With its growing influence streaming and streaming services could have an effect on the way movies are consumed and what type of movies are consumed. The main results of this paper show that the concentration of box office revenue has increased since 2007. The percentage of yearly total revenue earned by the top 10 and the top 20 films of each year has followed an increasing trend from 2007 until 2019 compared to the period before 2007. Therefore the trend of increasing concentration of box office revenue is consistent with the introduction of streaming in the film industry.

Moreover the results of this research show no evidence for a trend in the combined market share of the six major distributors after the introduction of streaming. Additional results on the individual market share of the major distributors show that only the market share of Walt Disney follows a positive trend since 2007. On the other hand the market share of Sony Pictures follows a negative trend since the introduction, while the market share of the other four distributors has not behaved statistically different since the introduction of streaming. Based on these inconclusive results it is not possible to state that the box office revenue has become more concentrated on the six majors. At most the conclusion can be drawn that Walt Disney has become the big winner and the biggest player on the market in the period from 2007 to 2019. Still the results of this research do not show whether this is due to the introduction of streaming or just better business strategies and investments. It is also possible that Walt Disney is more capable in responding to changes in the industry than other distributors. Also to be noted that these results are only based on the market share of box office revenue. The revenues generated by physical sales and VOD are not included. It could still be the case that the concentration of revenue earned by the major distributors from these products has increased after streaming was introduced.

This research also provides some exploratory results based on the analysis of the Gini coefficient and the Lorenz curves. While these results show that the overall inequality of the revenue distribution has steadily increased from 1996 until 2019, the percentage of revenue earned by the bottom groups seems to have stabilised since 2007. An initial decrease in this period has been followed by a slight increase back to the level it was at in 2006. Yet these suspicions are not confirmed statistically.

The main results of this research are in line with the theory of the superstar phenomenon. The biggest movies have become bigger since 2007, they earn a larger part of the total revenue and thus the distribution of revenue has become more concentrated. This rise in concentration is consistent with the development of new technological innovations in the film industry in the form of streaming. The implications of this result are that film producers should perhaps set their focus to a small amount of blockbuster instead of trying to spread their risk by making more smaller movies. One big hit will likely generate far higher revenues than a number of smaller mediocre performing movies. Moreover there are still opportunities for distributors outside of the six majors. While the concentration of individual movies has increased, the market share of the majors has not. Therefore it does not seem to be the case that only the major distributors are profiting from the introduction of streaming. Current businesses practices of the smaller distributors still seem to pay off and there is still room to in the market for these distributors.

There are some limitations of this research. First of all, the time-series analysis used in this research only shows evidence of potential trends over the years. As time is the only independent variable it is not possible to state what the actual causes of the present trends are. Even though there is academic literature suggests that the introduction of streaming has been a disruption in the entertainment industry, it is not possible to state whether the results are actually caused by the introduction of streaming or a different trend that is happening since 2007. It is possible that the growing concentration of revenue is caused by a different underlying reason. Moreover the mechanisms by which streaming potentially affects the concentration of box office revenue is not clear either. One possibility could be that the decrease of search costs for consumers has allowed them to easily watch smaller, less popular movies at home. On the other the larger marketing budget of bigger movies might still attract consumer to visit the cinema. A further possibility could be that cinema visits are increasingly seen as a special event. The cost per additional watched film is much lower on streaming services than it is in movie theatres. Therefore consumers might be willing to only visit movies of a certain quality and they might look at characteristics such as budgets to signal this quality. Another limitation is the fact that this research focuses on the revenue that is earned at the box office. Box office revenue is revenue earned by service which is time and place specific. Consumers can buy tickets online, however consumption of the film takes place at the predetermined location of the cinema. In general the long tail theory focuses mainly on the increase of niche products as a result of easier and cheaper distribution. While some aspects of the long tail theory are still applicable to the screening of films in cinemas, the long tail theory does not fully apply to physical sales or offline business. Therefore this research lacks a comparison between online sales through for example streaming or online purchases and physical consumption through cinema visits or physical DVD sales. Still this research at least shows the indirect results of the long tail theory. While trends in online consumption cannot be provided, a clear trend is apparent in the offline world of film consumption.

Future research could look to investigate whether the results provided by this research also apply to the online consumption of films. Moreover a comparison between box office revenue and online sales revenue could show potentially different determinants. It is possible that the motivations behind a cinema visit or streaming a movie at home are significantly different. In this area of research there are possibilities for both quantitative as well as qualitative research. Additional options for future research would be to investigate whether the growing concentration of revenue is merely on blockbusters or whether other identifying factors play a role as well. Perhaps the changes in concentration are different for different genres or perhaps all major producing companies have benefitted significantly from the growing concentration as opposed to the major distributors. Furthermore future research could address the dimension of welfare economics and provide normative assessments on the effects of streaming. The increased concentration of revenue could have significant effects on the welfare of consumers and suppliers. It is possible that this increased concentration will lead to less competition or more power to the biggest production companies. These developments could in turn result in market failures. Research on the dimension of welfare economics could provide useful insights for policymakers if the introduction of streaming does indeed lead to substantial market failures.

References

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

Anderson, C. (2004). The Long Tail. Wired Magazine. https://www.wired.com/2004/10/tail/

Baer, D.M. (1977). Perhaps it Would be Better Not to Know Everything. Journal of Applied Behavior Analysis, 10, 167-172.

Bakos, J.Y. (1997). Reducing Buyer Search Costs: Implications for Electronic Marketplaces. Management Science, 42(12), 1672-1692.

Basuroy, S., Ravid, S.A., Gretz, R.T., & Allen, B.J. (2020). Is Everybody an Expert? An Investigation into the Impact of Professional versus User Reviews on Movie Revenues. Journal of Cultural Economics, 44(3), 57-96.

Benhamou, F., & Peltier, S. (2007). How Should Cultural Diversity Be Measured? An Application Using the French Publishing Industry. *Journal of Cultural Economics, 32(2),* 85-107.

Brynjolfsson, E., Hu, Y.J., & Simester, D. (2011). Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. Management Science, 57(8), 1373-1386.

Brynjolfsson, E., Hu, Y.J., & Smith, M.D. (2003). Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. Management Science, 49(11), 1580-1596.

Brynjolfsson, E., Hu, Y.J., & Smith, M.D. (2006). From Niches to Riches: The Anatomy of the Long Tail. Sloan Management Review, 47(4), 67-71.

Burroughs, B. (2019). House of Netflix: Streaming Media and Digital Lore. Popular Communication, 17(1), 1-17.

Chintagunta, P.K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. Marketing Science, 29(5), 944-957.

Christman, E. (2017). Spotify’s Losses More than Double to $581M, Revenues Rise to $3B. Billboard. https://www.billboard.com/articles/business/7833686/spotify-2016-losses- financial-results-revenue

Christman, E. (2018). U.S. Music Industry Hits Highest Revenue Mark in a Decade, Fuelled by Paid Subscriptions. Billboard. https://www.billboard.com/articles/business/8257558/us-music-industry-2017-highest-revenue-in-decade-fueled-paid-subscriptions.

Clendenin, J.L. (2021). *The Cinerama Dome at Arclight Cinemas on Sunset Boulevard* [Photo]. Los Angeles Times. Retrieved from: https://www.latimes.com/entertainment-arts/business/story/2021-04-16/arclight-cinemas-whats-next

Cunningham, S., Silver, J., & McDonnell, J. (2010). Rates of Change: Online Distribution as Disruptive Technology in the Film Industry. Media International Australia, Incorporating Culture and Policy, 136, 119-132.

De Vany, A., & Walls, W.D. (1999). Uncertainty in the Movie Industry: Does Star Power Reduce the Terror of the Box Office. Journal of Cultural Economics, 23, 285-318.

Ding, C., Cheng, H.K., Duan, Y., & Jin, Y. (2017). The Power of the “Like” Button: The Impact of Social Media on Box Office. Decision Support Systems, 94, 77-84.

Dixit, A.K., & Stiglitz, J.E. (1977). Monopolistic Competition and Optimum Product Diversity. *The American Economic Review, 67(3),* 297-308.

Dorfman, R. (1979). A Formula for the Gini Coefficient. The Review of Economics and Statistics, 61(1), 146-149.

Duan, W., Gu, B., & Whinston, A.B. (2008). The Dynamics of Online Word-of-Mouth and Product Sales – An Empirical Investigation of the Movie Industry. Journal of Retailing, 84(2), 233-242.

Einav, L. (2007). Seasonality in the U.S. Motion Picture Industry. RAND Journal of Economics, 38(1), 127-145.

Elberse, A. (2007). The Power of Stars: Do Star Actors Drive the Success of Movies. Journal of Marketing, 71, 102-120.

Elberse, A. (2008). Should You Invest in the Long Tail? Harvard Business Review, 86, 88-97.

Elberse, A. (2010). Bye-Bye Bundles: The Unbundling of Music in Digital Channels. Journal of Marketing, 74, 107-123.

Elberse, A., & Oberholzer-Gee, F. (2007). Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales. Working Paper 07-015, Harvard Business School

Eliashberg, J., Elberse, A., & Leenders, M. (2006). The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions. Marketing Science, 25(6), 638-661.

Fellman, J. (2018). Income Inequality Measures. *Theoretical Economic Letters, 8,* 557-574.

Fleder, D.M., & Hosanagar, K. (2009). Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. Management Science, 55(5), 697-712.

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

Frey, B.S. (1998). Superstar Museums: An Economic Analysis. Journal of Economic Review, 22, 113-125

Handke, C. (2012). Digital Copying and the Supply of Sound Recordings. Information Economics and Policy, 24(1), 15-29.

Hartmann, D. P., Gottman, J.M., Jones, R.R., Gardner, W., Kazdin, A.E., & Vaught, R. (1980). Interrupted Time-series Analysis and its Application to Behavioral Data. Journal of Applied Behavior Analysis, 13, 543-559.

Herbert, D., Lotz, A.D., & Marshall, L. (2018). Approaching Media Industries Comparatively: A Case Study of Streaming. International Journal of Cultural Studies, 22(3), 349-366.

Hinz, O., Eckert, J., & Skiera, B. (2011). Drivers of the Long Tail Phenomenon: An Empirical Analysis. Journal of Management Information Systems, 27(4), 43-70.

Huffer, I. (2017). Social Inclusivity, Cultural Diversity and Online Film Consumption. *Cultural Trends, 26(2),* 138-154.

IMDbPro. (2021). *BoxOfficeMojo.* Retrieved from: https://www.boxofficemojo.com/?ref_=bo_nb_md_mojologo

Jones, R.R., Weinrott, M.R., & Vaught, R.S. (1978). Effects of Serial Dependency on the Agreement Between Visual and Statistical Inference. Journal of Applied Behavior Analysis, 11, 277-283.

Karniouchina, E.V. (2011). Impact of Star and Movie Buzz on Motion Picture Distribution and Box Office Revenue. International Journal of Research in Marketing, 28, 62-74.

Kim, S.H., Park, N., & Park, S.H. (2013). Exploring the Effects of Online Word of Mouth and Expert Reviews on Theatrical Movies' Box Office Success. Journal of Media Economics, 26(2), 98-114.

Krikke, J. (2004). Streaming Video Transforms the Media Industry. *IEEE Computer Graphics and Applications, 24(4),* 6-12.

Kumar, A., Smith, M.D., & Telang, R. (2014). Information Discovery and the Long Tail of Motion Picture Content. MIS Quarterly, 38(4), 1057-1078.

Lynch, J.G., & Ariely, D. (2000). Wine Online: Search Costs Affect Competition on Price, Quality, and Distribution. Marketing Science, 19(1), 83-103.

McCain, L.J., & McCleary, R. (1979). The Statistical Analysis of the Simple Interrupted Time-series Quasi-experiment. In T.D. Cook & D.T. Campbell (Eds.), Quasi-experimentation: Design and Analysis Issues for Field Settings. Chicago: Rand-McNally.

Moreau, F., & Peltier, S. (2004). Cultural Diversity in the Movie Industry: A Cross-National Study. *Journal of Media Economics, 17(2),* 123-143.

Napoli, P.M. (2019). Re-evaluating the Long Tail: Implications for Audiovisual Diversity on the Internet. In Albornoz, L.A., & Trinidad Garcia Leiva, M (Ed.), *Audiovisual Industries and Diversity: Economics and Policies in the Digital Era* (pp. 85-99). New York: Routledge

Nash Information Services. (2021). The Numbers. Retrieved from: https://www.the-numbers.com/

Naveed, K., Watanabe, C., & Neittaanmäki, P. (2017). Co-evolution Between Streaming and Live Music leads a Way to the Sustainable Growth of Music Industry – Lessons from the US Experiences. Technology in Society, 50, 1-19.

Oestreicher-Singer, G., & Sundararajan, A. (2006). Network Structure and the Long Tail of Ecommerce Demand. ICIS 2006 Proceedings.26.

Oestreicher-Singer, G., & Sundararajan, A. (2012). Recommendation Networks and the Long Tail of Electronic Commerce. MIS Quarterly, 36(1), 65-83.

Oh, C., Roumani, Y., Nwankpa, J.K., & Hu, H.F. (2017). Beyond Likes and Tweets: Consumer Engagement Behavior and Movie Box Office in Social Media. Information & Management, 54, 25-37.

Ordanini, A., & Nunes, J.C. (2016). From Fewer Blockbusters by More Superstars to More Blockbusters by Fewer Superstars: How Technological Innovation has Impacted Convergence on the Music Chart. International Journal of Research in Marketing, 33, 297-313.

Peltier, S., & Moreau, F. (2012). Internet and the ‘Long Tail versus Superstar Effect’ Debate: Evidence from the French Book Market. Applied Economics Letters, 19(8), 711-715.

Ravid, S.A. (1999). Information, Blockbusters, and Stars: A study of the Film Industry. The Journal of Business, 72(4), 463-492.

RIAA. (2020). *U.S. Sales Database.* Retrieved from: https://www.riaa.com/u-s-sales-database/

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

Stigler, G., & Becker, G. (1977). De Gustibus non est Disputandum. American Economic Review, 67, 76-90.

Tan, T., Netessine, S., & Hitt, L. (2017). Is Tom Cruise Threatened? An Empirical Study of the Impact of Product Variety on Demand Concentration. Information Systems Research, 28(3), 643-660.

Treme, J., & Craig, L.A (2013). Celebrity Star Power: Do Age and Gender Effects Influence Box Office Performance? Applied Economics Letters, 20(5), 440-445.

Tryon, W.W. (1982). A Simplified Time-series Analysis for Evaluating Treatment Interventions. Journal of Applied Behavior Analysis, 15(3), 423-429.

Wallace, W.T., Seigerman, A., & Holbrook, M.B. (1993). The Role of Actors and Actresses in the Success of Films: How Much is a Movie Star Worth? Journal of Cultural Economics, 17, 1-27.

Wallenstein, A. (2017). Home Entertainment 2016 Figures: Streaming Eclipses Disc Sales for the First Time. Variety. https://variety.com/2017/digital/news/home-entertainment-2016-figures-streaming-eclipses-disc-sales-for-the-first-time-1201954154/

Young, L.C. (1941). On Randomness in Ordered Sequences. Annals of Mathematical Statistics, 12, 153-162.

Yu, Y., Chen, H., Peng, C.H., & Chau, P. (2017). The Causal Effect of Subscription Video Streaming on DVD Sales: Evidence from a Natural Experiment. Information Systems of Economic eJournal.

Appendices

Appendix A – Films that are most popular in multiple years

Table A.1

Releases belonging to the most popular films in multiple years

Movies present twice on the top 20 list	Years
Aquaman	2018 & 2019 (13 & 20)
Avatar	2009 & 2010 (5 &1)
Cast Away	2000 & 2001 (19 & 16)
Frozen	2013 & 2014 (6 & 20)
Jerry Maguire	1996 &1997 (18 & 14)
Jumanji: Welcome to the Jungle	2017 & 2018 (18 & 7)
Meet the Fockers	2004 & 2005 (13 & 14)
Night at the Museum	2006 & 2007 (17 & 15)
Star Wars: Episode VII – The Force Awakens	2015 & 2016 (2 & 10)
The Lord of the Rings: The Fellowship of the Ring	2001 & 2002 (10 & 12)
The Lord of the Rings: The Return of the King	2003 & 2004 (4 & 14)
The Lord of the Rings: The Two Towers	2002 & 2003 (6 & 16)
Titanic	1997 & 1998 (7 & 1)

Notes: Numbers in brackets represent the ranking of the movie in the respective year

Appendix B – Complete overview of the critical values of the C-statistic for selected sample sizes at 1 percent level of significance

Table B.1

Critical values of the C-statistic for selected sample sizes at 1 percent level of significance

Sample size	Critical value	Sample size	Critical value
8	2.17	18	2.25
9	2.18	19	2.26
10	2.20	20	2.26
11	2.21	21	2.26
12	2.22	22	2.26
13	2.22	23	2.27
14	2.23	24	2.27
15	2.24	25	2.27
16	2.24	∞	2.33
17	2.25		

Notes: Taken from Young (1941). The critical value for the 5 percent level of significance is 1.64 for all sample sizes.

Appendix C – Additional robustness checks for the analysis of the top films using different post-periods

This appendix presents results of the intervention analysis for the top 10 and top 20 films using different post-periods. As explained in section III it is likely that the effect of streaming services on the concentration of box office revenue is gradual. The introduction of streaming has been a shock to the industry, however Netflix has gradually gained a larger amount of subscribers. As a result it could be possible that the complete effects of streaming services cannot be captured fully by one time series with unvarying periods. It might be the case that the influence of streaming has become more prominent over time.

In table C.1 the results of the tests for different post-period are presented. For each test the pre-period is consistent with the tests performed in section IV, thus ranging from 1996 until 2006. The post-period differs for each additional test. Each additional test uses a different starting year for the post-period. Therefore the first test is based on a post-period of 2008 until 2019. Observations that lie between the last year of the baseline period and the first year of the post-period are omitted from the analysis. As at least 8 observation points are needed to perform the intervention analysis, therefore no more than 5 observations can be omitted from the post-period. This means that the last test is based on a post-period from 2012 until 2019.

The results in table C.1 show that the value of Z is not consistent for different post-periods. While the results for the period of 2008 until 2019 are consistent with the results in section IV, for later periods the null hypothesis can only be rejected for the top 20 films. For the last additional period the null hypothesis cannot be rejected for the top 20 films either, meaning that the time series of the post-period does not behave different from the pre-period for the top 10 or top 20 films. The results do suggest that there is a trend present in the post-period for the top 20 films, however they provide little evidence for a trend in the post-period for the top 10 films. These results could also show that the introduction of streaming caused a shock with immediate effects in the year 2007 and the influence of streaming became weaker over time. However it could also be the case that the omission of observations led to less reliable results. For instance the latest period is based on a post-period consisting of the minimum amount of observations needed.

Table C.1

Intervention analysis for the top films for different post-periods

Post-period	Z value for top 10 films	Z value for top 20 films
2008-2019	1.72*	2.31**
2009-2019	1.52	2.04*
2010-2019	1.36	1.81*
2011-2019	1.57	1.87*
2012-2019	1.28	1.48

Notes: All the numbers in the table are rounded to two decimals. * denotes significance at a 0.05 confidence level. ** denotes significance at a 0.01 confidence level

Appendix D – Additional robustness checks for the analysis of the top film of the year

This additional robustness check focuses on the number one movie of the year. To test whether the trend of an increased proportion of the total revenue for the most popular movies is actually a result of one film becoming a larger winner, analysis of the film gaining the most revenue is performed as well. It could be that the number one movie of the year takes away attention and revenue from all the other movies and that the trend of the top movies taking a larger percentage of the cut is caused solely by a large increase in the cut of the most viewed movie. However the results of the intervention analysis of the top 10 and top 20 films already show that this is unlikely, as the evidence for a trend is stronger in the results of the top 20 than in the top 10. Still both analyses indicate that there is an existing trend in the post-period. In figure D.1 the percentage of total revenue earned by the best performing movie of the year is presented. The dotted line represents the break-point between the pre-period and post-period. Visual inspection of the graph does suggests that the variance might have increased in the post-period, yet an existing trend is more difficult to recognize. Only in the latest years a possible trend seems to present itself. As there does not seem to be a trend in the pre-period and no evidence for such trend is provided by the C-statistic for the pre-period, de-trending is again not required.

The results of the intervention analysis for the number one movie of the year are presented in table D.1. As stated this analysis does not provide evidence for a trend in the pre-period, because the Z-score is not significant (0.49). The result for the post-period is not significant either (0.85), therefore the null hypothesis that there is no trend cannot be rejected. Thus this analysis does not provide evidence for a trend in the revenue gained by the number one movie of the year. It is to be noted that the results of this robustness check are likely to be less reliable than the results of the tests performed for the top 10 and top 20 films, due to the increased risk of outliers in this particular analysis.

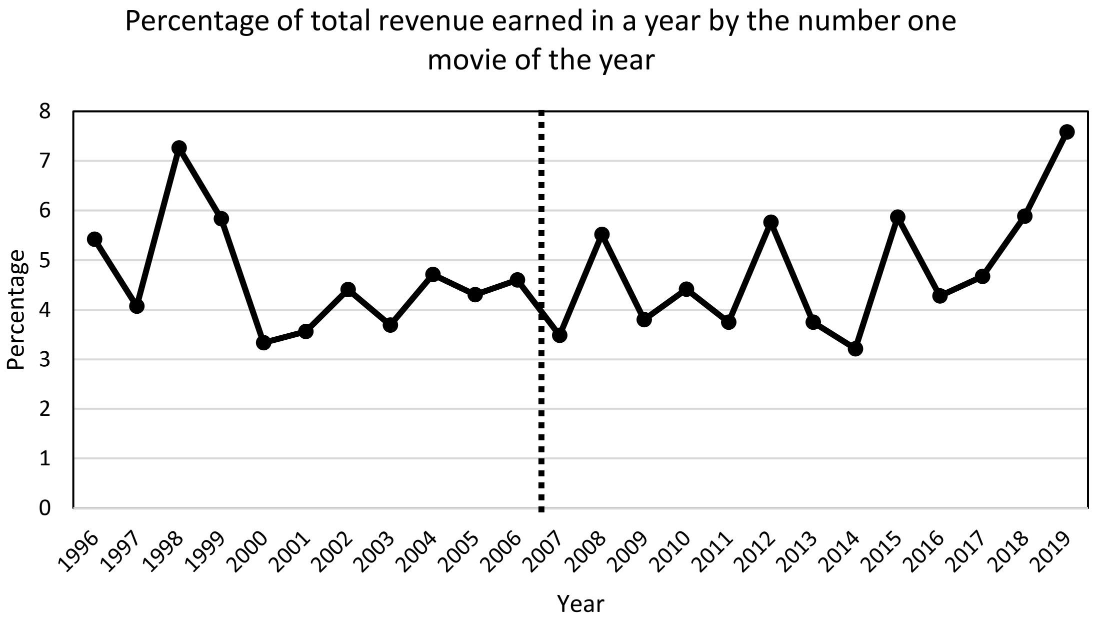


Figure D.1

Table D.1

Intervention analysis for the top film of the year

	Year	Percentage of revenue earned by the number one film	D^2	
	1996	5.42	1.82	Pre-period:
	1997	4.07	10.16	D^2= 22.83
	1998	7.26	2.04	2SS(X) = 26.33
	1999	5.83	6.23	(The mean for the pre-
	2000	3.33	0.05	period is 4.65)
Pre-period	2001	3.56	0.72	
	2002	4.40	0.51	C = 0.13
	2003	3.69	1.05	Sc = 0.27
	2004	4.71	0.17	Z = 0.49
	2005	4.30	0.09	not significant
	2006	4.60	1.24	
	2007	3.48	4.12	Post-period:
	2008	5.51	2.95	D^2= 30.33
	2009	3.80	0.38	2SS(X) = 38.78
	2010	4.41	0.44	(The mean for the post-
	2011	3.75	4.06	period is 4.77)
	2012	5.76	4.06	
	2013	3.74	0.28	C = 0.22
Post-period	2014	3.21	7.02	Sc = 0.26
	2015	5.86	2.52	Z = 0.85
	2016	4.27	0.16	not significant
	2017	4.67	1.48	
	2018	5.89	2.87	
	2019	7.58		

Notes: all the numbers in the table are rounded to two decimals. For 11 observations the critical value for the Z-statistic is 2.21 at a 1% significance level and 1.64 at a 5% significance level. For 13 observations the critical value for the Z-statistic is 2.22 at a 1% significance level and 1.64 at a 5% significance level (Young, 1941).