CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Bachelor thesis

2020 **Martin Derco**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Martin Derco

Determinants of overall performance of a motion picture at the domestic film market

Bachelor thesis

Prague 2020

Author: Martin Derco

Supervisor: RNDr. Michal Červinka, Ph.D.

Academic Year: 2019/2020

Bibliographic note

DERCO, Martin. Determinants of overall domestic performance of a motion picture. 42p. Bachelor thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies. Supervisor RNDr. Michal Červinka, Ph.D.

Abstract

A multi-billion film and television industry is a non-negligible component of both national and global economies, employing hundreds of thousands workers in the domestic market alone. An average major motion picture of recent years amounts to a hundred million U.S. dollars investment. The study explores the determinants of box office revenues. Firstly, the domestic film market is described, and related literature is reviewed. Secondly, using a unique cross-sectional dataset obtained from major publicly available sources we construct several models which should provide us with explanatory information on what factors relate to a theatrical success in terms of revenues. The log-log OLS regression is employed to estimate the impact of key determinants of film's profitability. In conclusion the evaluation of hypotheses is provided alongside with several suggestions on future research and film production.

Keywords

film industry, film, OLS regression, box office revenues

Abstrakt

Multimiliardový filmový a televizní průmysl je v dnešní době nedílnou složkou jak národních ekonomik tak i té globální. Jen na území Severní Ameriky vytváří každoročně stovky tisíc pracovních pozic. Průměrné náklady snímků v produkci velkých studií se pohybují v řádu sta milionu dolarů. Cílem práce je identifikace faktorů filmových zisků. Nejprve jsme popsali filmový trh na území Severní Ameriky a zmínili předchozí práce týkající se tohoto téma. Dále jsme za použití dostupných dat sestavili několik ekonometrických modelů zkoumajících faktory ovlivňující filmové tržby. Zlogaritmovanou transformací metody nejmenších čtverců jsme vypočítali regresi a získali charakteristiky jednotlivých nezávislých proměnných. V závěru práce jsou vyhodnoceny jednotlivé hypotézy, navrhnut možný postup pro další výzkum a formulována doporučení pro budoucí filmovou tvorbu.

Klíčová slova

filmový průmysl, film, metoda nejmenších čtverců, tržby

Declaration of Authorship

1. The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

2. The author hereby declares that all the sources and literature used have been properly cited.

3. The author hereby declares that the thesis has not been used to obtain a different or the same degree.

Prague 30.7.2020 **Derco, Martin**

Acknowledgments

I would like to express my gratitude to my supervisor RNDr. Michal Červinka, Ph.D., for providing me with valuable insights and assistance and his willingness to share his time. Additionally, special acknowledgments belong to my family for supporting me during the whole studies.

Bachelor’s Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author's name and surname:	Martin Derco
Supervisor:	RNDr. Michal Červinka, Ph.D.
Language:	English

Proposed topic:

Determinants of overall domestic performance of a motion picture

Preliminary scope of work:

Motion picture and television industries have fundamental impact on both local and national economies. For example, the filming of “Captain America: Winter’s Solider” boosted Ohio’s economy by $31 million (Fried, 2015). In addition, it employs local workers, brings film-induced tourism and foreign capital into shooting locations. Recently, the industry experience rapid changes, e.g. breaking sales records only days after being set and thus I find it meaningful to examine which factors affects theatrical revenues and also review past studies concerning this topic.

First, I intend to review past studies and literature, add recent data and subsequently use regression analysis. Preliminary analysis shows that the opening weekend box office performance of the US films at the domestic (North American) market becomes more and more reliable predictor for the overall revenues. Our primary objective is to examine the relationship between opening weekend box office performance and consequent success (in terms of revenue) of a motion picture using publicly available data and its development throughout past two decades. Additionally, I would like to review Mrs. Dvořáková´s thesis studying pre-release factors concerning European motion pictures as high-risk business projects and apply key elements of her study to the US market data including also period 2017-2019. Moreover, I would like to attempt at identifying other factors affecting opening weekend US box office itself such.

In my thesis I plan to use regression analysis on data from International Movie Database (IMDb.com) and Box Office Mojo (boxofficemojo.com) and other sources.

Outline

1. Introduction
2. Literature review
3. Data and variables
4. Methodology
5. Results & Interpretation
6. Summary and Conclusion
7. References

List of academic literature:

Woolridge, J. M. (2012) Introductory econometrics: a modern approach, 5th Edition, South-Western Cengage learning, Michigan State University

Dvorakova, A. (2017) Faktory finančního úspěchu filmu na evropském filmovém trhu Master thesis. Masarykova univerzita, Ekonomicko-správní fakulta. Supervisor Daniel Němec.

Barrie Gunter (2018), Predicting Movie Success at the Box Office, Palgrave Macmillan, Cham

Wenjing Duan,Bin Gu,Andrew B. Whinston (2008), “Do online reviews matter? — An empirical investigation of panel data”, Decision Support Systems, Elsevier, 45(4), 1007-1016

Lubbers, C. A., Adams W.J (2004) Merchandising in the Major Motion Picture Industry: Creating Brand Synergy and Revenue Streams. Journal of Promotion Management, 10(1/2), 55-63

Fried, N. (2015) “U.S. film and TV production drives economic growth in every corner of America” Retrieved from https://www.mpaa.org/u-s-film-and-tv-production-drives-economicgrowth-in-every-corner-of-america/

Contents

List of tables

List of figures

1. Introduction

The motion picture industry and specifically the North American market, that is also being referred to as the 'domestic' market, is a rapidly growing environment, generating over $11 billion in revenues, characterized with heavy competition and high level of risk element. Globally, the film industry amounted to $42.2 billion[1] in 2019, an increase of roughly $0.4 billion compared to the previous year. This is a trend we could have observed over the past many years starting (for the purpose of our thesis) with 2005's figure of $23.1 billion.

Despite the fact the industry itself grows in numbers it is also going through major changes in the past years. Secondary distribution channels such as online televisions and streaming channels are gaining on popularity, while film theatres face a downturn in physical attendances. Therefore, any possible risk mitigation strategies and success-breeding patterns in terms of theatrical releases are expected to gain on significance as theatre screens may slowly become a niche market regarding motion picture's quality.

In our thesis we seek to analyze a data sample of motion pictures with release dates between years 2005 and 2019 using the information available at the *International Movie Database* and identify which factors determine their success in terms of theatrical revenues. In our thesis we will primarily focus on the factors that are being decided in the process of film production and may be affected by the producers at least to some certain extent. These factors are also often referred to as the 'pre-release factors'. Furthermore, our objective is to evaluate the significance of opening weekend and its impact on further performance of motion pictures.

The general idea behind the analysis is that producers are able, to a certain extent, affect the opening box office success and thus, potentially, the overall economic performance of a motion picture by choosing felicitous attributes to attract desired attention of broad audience.

In the first part of this work we will describe domestic film market, its features and key players. Afterwards, we will go through several notable pieces of literature that paid attention to this topic and performed similar analyses on past data samples. Our set of hypotheses should be based on assumptions arising from past research given deeper

[1] Source: statista.com, https://www.statista.com/statistics/271856/global-box-office-revenue/

knowledge about the market itself, its trends and methodology previously used as well. Testing these hypotheses is expected to shed more light on the drivers of economic success of major motion pictures.

After we evaluate the most appropriate approach towards analyzing our unique data sample, we describe our data set and provide an empirical analysis. Afterwards, we will form a regression equation that will help us estimate the individual effects of various determinants. In the last part we will present our own conclusions and possible suggestions.

In our thesis we will refer to film theatres as ‘theatres’ only. By the term ‘domestic market’ we understand the geographical region of the North America including Puerto Rico and Guam. Information acquired from web articles will be sourced in the bottom notes, academical articles quoted will be in brackets. Terms ‘film’, ‘movie’ or ‘motion picture’ are considered synonyms and refer to the same object defined as ‘a recording of moving images that tells a story and that people watch on a screen or television’[2].

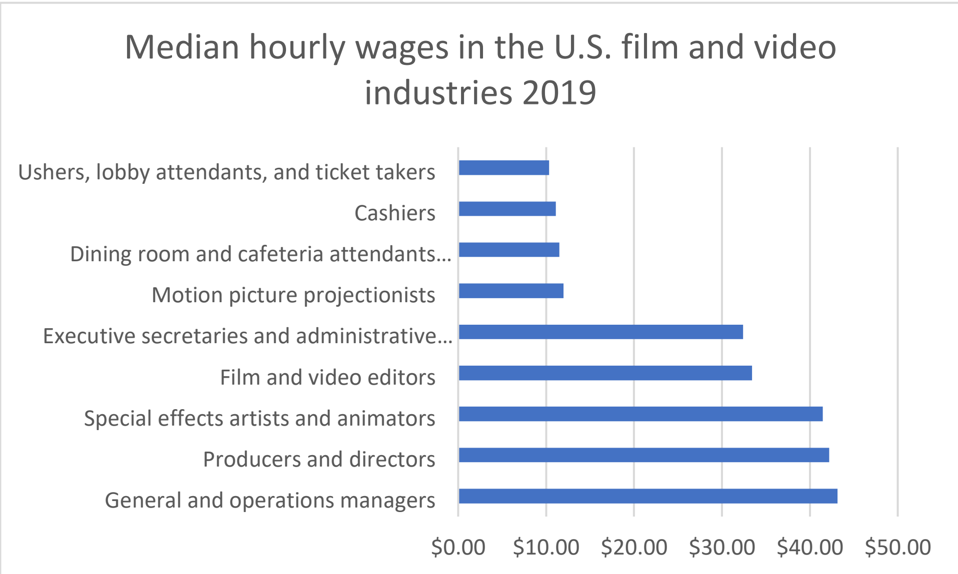
---

[2] Source: Merriam-Webster Colleagiate Dictionary, https://www.merriam-webster.com/dictionary/movie

2. Domestic market description

Film industry has a non-negligible impact on both culture and economy. According to statista.com TV and film production generated 85.1 thousand direct and 95.8 thousand spin-off workplaces over the 2018/2019 season in Canada and there were 456 thousand[3] people employed in the U.S. motion picture and sound recording industry in January 2020. This includes production jobs in theatrical feature film, television, broadcaster in-house, and freelance sales segments. Another interesting statistic is that average hourly earnings of all employees in the motion picture and sound recording industries was $37.78 in 2019 which is 60.8% higher than an average wage of U.S. worker that was roughly $23.5 in the same year[4]. Additionally, as reported by the *Economist*[5], the average cost of film production today is about $60 million. The cost of marketing averages approximately $40 million, making the total average cost of producing and marketing a major motion picture roughly $100 million. However, the impact films have on the economy goes beyond the money spent on the production and its theatrical release. For example, Warner Bros. made $500 million from selling Batman merchandise and $251.1 million from the theatrical release (Lubbers and Adams, 2004). In 2016, the industry made $262.9 billion from licensed merchandise and $54.6 billion from corporate trademarks (Szalai, 2017).

Figure 1: US median hourly wages, source: Statista.com



[3] Source: statista.com, https://www.statista.com/statistics/184412/employment-in-us-motion-picture-and-recording-industries-since-2001/
[4] Source: U.S. Bureau of labor statistics, https://tradingeconomics.com/united-states/wages
[5] Source: The Economist, 2016, https://www.economist.com/graphic-detail/2016/02/26/how-to-make-a-hit-hollywood-film

The Motion Picture Association (MPA) is an American trade union representing five major film studios in the United States – Walt Disney's Studios, Warner Bros., Universal Pictures, Sony Pictures, Paramount Pictures, originally known as the 'Big Six', until Disney's acquisition of the 20th Century Studios in 2019, as well as video streaming service Netflix – further on, we will exclude Netflix from our analysis and focus on film studios only. Releases of the MPA accounted for 83.7% of all earnings ($11.3 billion) at the box office in North America in 2019.

Walt Disney's studio was the first one to reach the $3 billion bar in 2016 and hit the record-breaking domestic box office gross of $3.8 billion in 2018. Disney's studios with domestic market share of 33.1% (or even close to 38% considering the market share of 20th Century Studios) strengthened its position of lead market player on both domestic and global markets in 2019. The studio is constantly solidifying its monopoly position by acquiring profitable companies and franchises like Fox, Marvel Studios or aforementioned 20st Century Studios.

Table 1: Domestic market share, source: Statista.com (2020)

Studio	Box Office Market Share through 2019
Disney *(Combined with 20th Century Studios)*	33.10% *(38.00%)*
Warner Bros.	13.80%
Universal	13.40%
Sony	11.70%
Lionsgate	6.80%
Paramount	5%
Other	11.20%

At $717 billion, the U.S. media and entertainment industry (M&E) represents a third of the global M&E industry including motion pictures, television programmes and commercials, streaming content, music and audio recordings, broadcast, radio, book publishing, video games and ancillary services and products, standing no.1 in the world.

The U.S. M&E is expected to reach more than $825 billion by 2023, according to the 2018-2023 Entertainment & Media Outlook[6] by PriceWaterhouseCoopers.

According to *The Numbers*[7] there were roughly 1.24 billion tickets sold in the domestic market in 2019, compared to 1.58 billion tickets in 2002 which was historically a peak, however the overall box office revenue has continuously increased mainly due to raising ticket price: inflation and the introduction of 3D technology brought average ticket prices up from $4.23 in 1990 to $9.11 in 2018. In fact, 2018 was the highest-grossing year of all time with ticket sales revenue over $12 billion which was likely caused by the highly anticipated releases that attracted enough consumers despite the admission prices being at their highest.

---

[6] Source: Media & Entertainment spotlight, https://www.selectusa.gov/media-entertainment-industry-united-states

[7] Source: The Numbers, https://www.the-numbers.com/market/

3. Literature review

In our thesis we would like to inspect key factors affecting monetary revenues of motion pictures originated from the domestic market. In this section we will present several essential studies focusing on the same topic.

3.1. Risk evaluation

Because films are so expensive to produce, the revenue they make is important in deciding if more films of the same type are made in the future (Berger and Raddick, 2017). In other words, studios and producers are likely to be interested in what factors affect the revenue of their future film. Film studios face multiple uncertainties connected with producing a motion picture. First obstacle on the way towards a new picture is raising sufficient funds. Developing a film may be described as sum of work surrounding the initial concept of the story or idea. Producers need to decide whether seek an original idea or rely on secondary source (literature, plays, previous films etc.). The whole creative and executive process encompasses significant costs, which will take considerably long time to return if so. There is roughly one in twenty film projects that makes it through development to production in Hollywood (Finney, 2015).

Another threat of present-day film industry relates to distribution. Modern digital technologies make it increasingly difficult for filmmakers to profit on their piece of work through the traditional distribution channels - cinemas and theatres. The audience shifts towards online renting and streaming of films rather than buying films post theatrical release. One of film industry-specific situation occurs in the marketing field. The life span of a film in the primary market is usually quite short – around 10 weeks – and thus the motion picture has relatively short time to earn a meaningful profit to its producers, which may result into a failure without a suitable marketing strategy (and sufficient advertisement budget). Finally, there is the risk connected with the audience perception of the final product. Consumer’s reaction is highly unpredictable and fan reviews spread quickly through the social networks, therefore, even if a film makes its way through all the aforementioned obstacles, the eventual success depends on the highly uncertain public acceptance (Finney, 2015).

There were several studies focusing on risk mitigation strategies in the process of filmmaking. An interesting finding comes from Emanuele Teti, suggesting concurrent production a reasonable strategy for Hollywood film companies to substantially mitigate risk of losing money in absolute terms. Parallel production of motion pictures with different costs appeared to be constantly successful as the financial loss coming from *n*-1 (*n*-2) films is, in many cases, offset by the revenue generated by one (or two) successful project included in the same sample (*n*) (Teti, 2013). However, this whole scheme is generally backed by high production budgets of American films generated by investors who, trusting the qualitative and business validity of the projects, invest sizeable amounts of money (Finney, 2015).

Teti used a dataset from AC Nielsen's database, that included 4,178 observations coming from the US film market between years 1988 to 1999. The analysis focused on theatrical revenues specifically and did not take into account other distribution channels as the objective was not to predict overall expected profitability, but rather assess indicators that can depict the scale or degree of dispersion of these expected values – that is, the risk and return trade-off the film companies must face. Films that generated extraordinary revenues (in some cases over 1,000%) were removed from the sample as these were mainly low budget projects that, for some reason, achieved the level of audience approval necessary to generate such returns. These observations collectively contribute very little in terms of money generated but distort the analysis of frequency distributions because of their extreme rates of return compared to their costs. The final dataset was based on a sample population of 1,636 observations.

A frequency distribution method was used, controlled for several factors describing the context of situation – year of release and revenues of the whole U.S. population in the specific years.

Figure 2: Frequency distribution of revenues, source: Teti, 2013

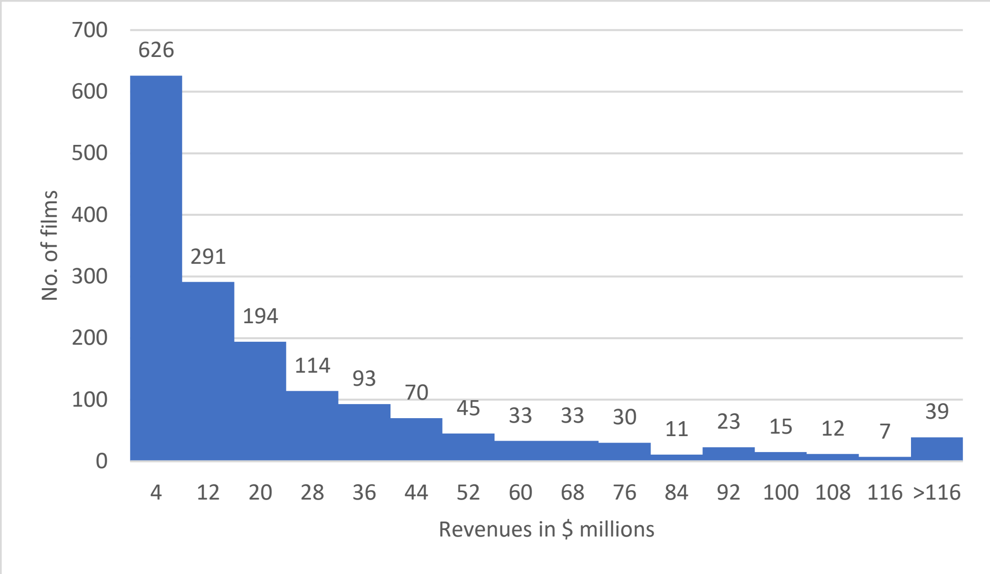


Figure 2 shows us that the distribution is highly right-skewed, indicating that a small number of films generate extremely high revenues, while most films produce revenues around or below the mean annual value for the given period. Furthermore Teti's results showed an extremely high kurtosis, which indicates that most of the variance is caused by infrequent 'extreme' events rather than frequent 'ordinary' ones - for instance, the highest kurtosis – found in 1997's data – can, thus, be mainly attributed to the presence of Titanic which, alone, generated revenues of $413 million, equal to the average revenue of 13.5 films in 1997.

The mean revenue of the entire population of 1,636 films from Figure 2 is equal to roughly $25 million, while the interquartile mean (IQM) produced a significantly lower outcome - $14.3 million, which corresponds better with median ($12.7 million) as well. These observations leads us to a thought that the films to the right of the median value 'weigh' much more heavily than the films to the left of it; in another words, the potential high revenues of the hits should more than balance the potential losses of the worse-performing films of the distribution. In each year, a small number of films achieved extraordinary returns while the majority delivered rates of return below mean value. Most of inspected films fell below the breakeven point and had negative rates of return. The IQM value and median value being quite different from the mean value, offers a clear evidence that the distribution is a highly irregular one. We will come across similar feature in our analysis as well.

In conclusion Teti demonstrated that considerable investments in production are a good way to increase the likelihood of larger box office takings but are not a guarantee of a positive return from the project. Applying this kind of logic allowed Hollywood film companies to make the best of a sector that, while extremely risky, is a great generator of money in absolute terms.

3.2.Success determinants

The economic reasoning of identifying the success determinants is quite straight forward. Usually, researchers seek to specify the key factors that affects the financial performance of a movie, to provide producers with guidance on how to increase the profitability of their film project.

To the best of our knowledge, the first one to come up with a multiple regression analysis in order to specify determinants of motion picture theatrical success was Litman in 1983. In his article, Litman examined a sample counting 125 films released in the USA between years 1972 and 1978. Domestic theatrical rentals were set as the dependent variable. There were five binary independent variables for film story types (genres) - Science Fiction, Drama, Action-Adventure, Comedy, and Musical. The original hypothesis tested in Litman's work was that science-fiction and action-adventure genres should generate higher profits than the other categories. Another set of independent binary variables stood for three[8] individual MPAA ratings (R, PG, G), expecting PG gross the highest rentals as it is not as restricting as R nor being stuck with Walt Disney's label signalling children-specific movie – we will go with a similar thinking in our work. Remaining explanatory variables were controlling for seasonality, presence of popular actors, involvement of major distribution studio(s) and budget estimation.

In conclusion, Litman's analysis highlighted several determinants of movie success for Hollywood film industry in 1980's and set a base for future studies in this field. Litman himself continued to study the movie industry and brought some new insights in 1989. In general, there were two fundamental conclusions coming from this study; as the number of motion pictures releases raised rapidly, advertisement costs increased, the role of marketing became principal, and the production budget variable

[8] There were no movies rated NC-17 (originally X) in the dataset

gained on significance. Secondly, pre-release factors had attained greater attention as almost every variable that represents some known factor or conveys information that reduces uncertainty seems to be correlated with financial success. (Litman 1989)

Another approach was chosen by De Vany and Walls (1999) who found out that box office revenues might be asymptotically Pareto-distributed and have infinite variance. Hence, expected revenue's forecast is imprecise and lacking in foundation. However, better choices might be achieved when the impact of certain variables on the probabilities can be predicted. De Vany and Walls treated movie project's box office revenues as probability distributions, and they strived to analyse the probabilities of extreme outcomes. They decided normal distribution unsuitable for this kind of analysis because when outcomes are normally distributed, the probability of extreme outcomes is vanishingly small.

Their principal objective was to describe the characteristics of a motion picture's revenue over different stages after its release (1 week after release, 5 weeks etc.), find its cause and describe its distribution.

Their results showed that weekly revenues were autocorrelated. A film that experienced recent increase in revenues is more likely to generate additional growth rather than a film which experienced growth in distant past. As to the distribution characteristic they found that total revenue distribution is a mixture of opening distribution – where majority of films gross humble revenues and several hits achieve massive success – and a log normal distribution, suggesting that if the film runs long enough it may acquire an independent life. Despite converging to log normality the revenue distribution never quite reaches it, because of fatter tails and mass points at the far right, where blockbusters are located – our data on revenues provide a very similar distribution.

De Vany and Walls extended this work in 1999 when they tried to answer our common question: How do stars, genre, release patterns, et cetera, alter the quantiles, extremals, probability mass, and survival functions of motion pictures, focusing specially on the star factor. They decided to calculate the probabilities a movie will earn revenues equal or greater than $50 million and then calculate how stars or other factors alter those probabilities.

For their work they used data including 2015 observations of movies released between years 1984 and 1996. Each actor, producer or director listed on Premier's list of

100 most powerful people in Hollywood or in James Ulmer's list of A and A+ people was considered a 'star'. Around 20% of their observations featured a 'star'.

First, the effect of a star presence on revenue distribution was tested. Revenue distribution was plotted as a log function of Rank (1=highest grossing movie in a specific week), presence of a 'star' (0 or 1) and their interception term over 6 two-year intervals. In their work they refer to this estimation as to 'Pareto rank law'.

Their results suggested Pareto rank law quite stable over the followed decade despite escalating production and advertising budgets. De Vany and Walls considered Pareto rank distribution an exceptionally good fit for all movies with or without star presence and thus concluded that the distinguishing factor causing movies to be strongly ranked in terms of revenues cannot be traced to stars (De Vany and Walls, 1999).

Finally, they decided to evaluate the effect of a star on a motion picture's profit. Profit was estimated a function of star, genre, sequel, rating and year, where profit was set as 0.5*revenue-budget[9] and measured in millions of dollars. The equation was estimated in levels rather than logs, because of negative profits for the majority of sample. Very poor fit with a low value of R squared was delivered for were profits expectable, everyone would make them (De Vany and Walls, 1999).

In conclusion De Vany summarizes movie industry as profoundly uncertain business. The probability distributions of box office revenues are characterized by heavy tails and infinite variance. In his opinion past success does not imply one in future and forecasts on expected are meaningless as the possibilities do not converge to the mean. Producers may take precautions and position a movie to improve its chances of success, but after release, the audience decides its fate.

Some of other, more advanced methods used to analyse film industry, is path analysis. First to use the path analysis method to analyse the determinants of theatrical success were (Thurau et al., 2006).

Path analysis method is an extension of multiple regression. Its aim is to provide estimates of the magnitude and significance of possible connections between sets of variables (Webley and Lea 1997).

[9] Profit did not include foreign and other non-theatrical revenues. The 0.5 figure is a rough estimate of the average rental rate. This equation was a crude approximation to profits in the domestic film market (De Vany and Walls 1999)

They defined a set of hypotheses concerning interrelations between dependent variables included in two models with a differing dependent variable – long term domestic box office and profitability. They tested their hypotheses against the sample of 331 films released in the USA in the years 1999-2001. Their model evaluated mutual relationships among *Profitability, Short Term Box Office (STBO), Long Term Box Office (LTBO), star power, director power, cultural familiarity* (more likely a sequel factor), *production costs, advertising expenses, summer release* binary variable*, reviews, awards* and *consumer perceived quality.*

Thurau claims, their model explains around 30% of profitability and 80% for LTBO which would be quite remarkable, while on the other hand there are several notable limitations connected with this method. First, path analysis is most likely to be useful when researcher already has (one or) a small set of hypotheses to test, all of which can be represented in a single path diagram. It has little use at the explanatory stage of research. Second, the analysis is not able to establish the direction of causality. (Webley and Lea, 1997)

Another modern approach were the neural networks (Delen and Sharda, 2006). Neural networks (NN) are known to be biologically inspired analytical techniques, capable of modelling extremely complex non-linear functions. In their study they used a data sample of 834 movies released between 1998 and 2002. The data was discretized in nine categories according their box office revenues ranging from a 'flop' (grossing below $1 million) to a 'blockbuster' (grossing over $200 million). Seven different independent variables were chosen to describe box office revenues.

Each independent variable (category) had a set of possible binary pseudo-variables i.e. an R rated movie had a pseudo-variable R value set to 1, and G, PG, PG-13 and NR set to 0. Competition variable was set "High" for release months June and November, "Medium" for May, July and December and "Low" for the rest of months. The only exception was variable Genre, where individual movies may be assigned as many categories as the movie can be classified in – as these do not exclude one another. In total 26 pseudo (representing seven categories) and nine output (representing revenue categories) variables were used in this study.

The NN model's aim was to categorize each film into one of the nine box office categories. There were two metrics used to assess model's accuracy. A percentage of a correct classification, 'bingo' rate, and 1-away correct classification rate. Results were plotted in Table 2.

Table 2:NN confusion matrix, source: Sharda and Delen, 2006

		actual categories									avg.
		1	2	3	4	5	6	7	8	9	
predicted categories	1	37	35	5	4	0	0	0	1	2	
	2	33	37	13	14	0	1	0	1	1	
	3	5	13	28	21	1	4	8	7	4	
	4	15	3	16	38	0	2	3	4	9	
	5	0	0	6	13	55	30	7	3	2	
	6	0	1	2	3	31	26	19	13	4	
	7	0	0	8	5	5	12	24	21	10	
	8	0	0	5	2	3	7	24	20	16	
	9	0	0	9	1	2	7	8	22	43	
BINGO		0.411	0.416	0.304	0.376	0.567	0.292	0.258	0.217	0.473	**0.369**
1-away		0.774	0.955	0.620	0.713	0.887	0.764	0.720	0.685	0.648	**0.752**

Blue-highlighted cells showed the correctly classified 'bingo' observations, while others show misclassifications. The directly neighbouring cells shows misclassifications of one class, these were used to calculate one-away correct classification rate. Table 2 used summarized data for all the five years assuming the predictors for financial success had not changed significantly over the followed period – this assumption was later validated.

The results of their model show, that neural networks are able to predict the 'success' category of a motion picture before its theatrical release with 'bingo'-exact accuracy of 36.9% and with a possibility of one-away correct classification the model achieves even 75.2% accuracy. In their work Sharda and Delen even provided accuracy comparison with alternative techniques – Logistic regression, Discriminant analysis and Classification and regression tree – achieving roughly 6% improvement compared to logistic regression. However, in their own work they claimed that even though providing better predictive ability, neural network applications in marketing provide lesser explanatory value in general.

3.3. Recent studies

One of more up-to-date studies by Lubbers and Mika, 2018 focused on genre and MPAA rating's effect on box office revenues. Set of 457 observations was analysed through simple correlation using the Pearson's correlation coefficient. A crosstabulation table with chi-square calculation was used to evaluate the relationship between movie genre and box office. To deal with the continuous character of box office data in comparison with categorical character of genre and MPAA rating data, they divided the box office into 5 categories from low to high. The chi-square test showed that there is a statistically significant difference between the movie genre and its box office success.

Their results support previous studies stating that there is a significant relationship between genres and box office revenues (Redfern, 2012). However, the effect may differ based on how movies are divided into genre classes. Often movies are assigned to more than one genre or a genre-hybrid (two genres combined into one, such as RomCom). Depending on which categorization is used it may affect the results.

Furthermore the MPAA analysis suggested rating to be a statistically significant determinant of box office revenue as well. These results support findings that the PG-13 rating reaches the highest average box office (Austin et al., 1981). A possible explanation for this is that the MPAA rating limit the audience that goes to the movies, and the PG-13 rating is a 'happy medium' where both children and adults can enjoy the entertainment. The PG-13 category is the one that the majority of audiences feel comfortable attending, thus including the widest consumers base.

Another recent study, though still only a preprint version, is (Hyatt and Johnson, 2020). While focusing on success in terms of 'positive reviews' rather than from financial point of view, the method they used might be of potential interest for our paper as well. Being only a preprint version, we choose not to comment on the results but rather focus solely on methodology used. In their work they decided to model 2SLS to estimate the impact of variables on average customer reviews and the next marginal customer review simultaneously. Their model used residuals from the first stage as an additional independent variable in the second stage, with an aim to bridge the two stages and incorporate average rating into the future rating estimation. In general, their conclusions were in accordance with previous studies, while the second model's stage provided remarkably poor explanatory value.

Last but not least we paid attention to (Adela Dvorakova, 2017), a master thesis which presented a motion picture as a high-risk investment and focused on possible risk-mitigation strategies. Primary objective of her work was to find out what pre-release factors affect the future profit of a motion picture and to what extent. While Dvorakova focused on the European market we will examine the domestic market. Typically, production budget, genre MPAA rating, length of a movie, major studios/distributors, star or director power, and release date are listed as key pre-release factors of success, while opening weekend box office and domestic box office reckon among post-release factors.

Dvorakova has recourse, again, to the multiple linear regression analysis which is, to our best concern, the most frequent method of predicting the success of a motion picture in current literature. According to the set of hypotheses provided, her objective was, among others, to test whether films of certain genres perform better than others.

Dvorakova used a data sample of 1457 films released between years 2000 and 2010 in the territory of the European Union. An interesting feature of European market is a low number of releases in the summer and pre-Christmas period, which is probably caused by increased and undesired competition from the North American distributors. There are also major differences between projects produced on a purely national level and movies that were co-produced with major distribution studios and Hollywood studios, especially in terms of budgets.

In terms of results Dvorakova divided her work into two parts; she ran the same regression on a full sample of 1457 films including observations co-produced with major non-EU producers, and then again on a sample of purely EU motion pictures. In her results she compared the outcomes.

4. Data review

In the following section we will go through the sources of our data and briefly describe them, and we will provide list of eligibility conditions applied on our dataset. Furthermore, we will divide our data into two categories according to their character and describe individual data characteristics.

4.1. Data source

The data set contains both domestic and worldwide budget and revenue estimates as well as various other available theatrical characteristics such are genre, MPAA rating, or running time. The original data source is International Movie Database (IMDb.com).

According to the server's information, IMDb is the world's most popular and authoritative source for movie, TV and celebrity content. As of December 2019, IMDb has approximately 6.5 million titles (including episodes) and 10.4 million personalities in its database, as well as 83 million registered users. IMDb was launched on October 17, 1990 and is currently owned by Amazon.

IMDb's subsidiary, *Box Office Mojo*[10], is a website that tracks box office revenues as well as many other statistics on 3-day basis. It follows performance of recent releases, overall gross profits, top-10 gross profits and percentual change. The completeness of data is not flawless, there are some missing data in the dataset, especially for low budget, independent films, however, the data contained in IMDb is currently the most reliable source of information on film financial data. Furthermore, it provides general movie data such as release date, year of production, running time, MPAA rating, or user reviews. On top of that it also contains information related to specific distributing companies or individual personalities related to movie industry such are distributors, producers, composers etc. As an additional data source, we will use server *The Numbers*[11]. This website provides movie industry data and research services for clients from the whole film industry.

[10] Source: www.boxofficemojo.com
[11] Source: Nash Information Services, LLC, www.the-numbers.com

4.1.1 Eligibility and dataset adjustments

For a work to be eligible for inclusion in the database it must be of ‘general public interest’ and should be available to the public or have been available in the past. General public interest is assumed if a work has been:

- released in cinemas,
- shown on TV,
- released on video or the web or prints have been made available to the public,
- listed in the catalogue of an established video retailer; (i.e. Amazon.com),
- accepted and shown on film festivals,
- made by a (now) famous artist or person of public interest,
- made famous for some reason and is widely talked about/referenced in media or the 'film community' or is now of general historic interest for some reason. (IMDb, 2020)

Our original dataset contained 11,607 films produced and released in the domestic market between years 2005 and 2019.

The criteria for inclusion in the data set suitable for future analysis were the following:

- the entry is not a documentary, TV show, broadcasting, video game,
- the data on theatrical revenues (box office) of the film is available,
- the data on production budget of the film is available.

Observations not satisfying aforementioned criteria were removed from the original dataset and were not used in our analysis.

Dataset was intentionally deprived of 2020’s data due to two main reasons – Movies released in the beginning of year 2020 might not had ended their screening period in the time of analysis and thus their box office data would have been skewed, secondarily, movie industry suffered tremendous losses caused by COVID-19 pandemic and so 2020’s data would have been presumably an unwanted outlier.

Further on, we needed to clear the dataset from all the observations with incomplete data. We had to check on the inclusion criteria and afterwards null the observations with zero running time, no MPAA ratings or other faulty values that would bias our results. After disposal of unsuitable observations, we were left with a dataset of 2353 films.

4.2 Financial data

Domestic box office - The term box office is used to refer to the commercial success of a movie in terms of revenue made from movie theatre ticket sales at domestic market. Specifically, a first run release domestic box office revenue is a gross profit for the first 133 screening days of an original release (some movies had multiple re-relases) of a motion picture. The box office, and more specifically the prediction of a movie's box office, have been a popular topic for researchers ever since motion pictures became an industry.

Domestic opening weekend box office is commonly defined as a gross profit between the first Friday and first Sunday of release (and including Thursday previews) (The Numbers, 2020).

Opening weekend box office estimate has allegedly great impact on overall profit of a film. A controversy from 2014 around Transformers 4: Age of Exctinction supports this hypothesis. Paramount, the studio backing the film, was accused of overestimating box office totals during the film's opening weekend. Paramount said the movie ranked in \$100.38 million, in line with expectations that it would exceed \$100 million, but box office watchers came up with a different tally — closer to \$97.5 million (Busch, 2014). Being first to hit the \$100 million opening weekend bar in 2014 – a powerful marketing point - was supposed to affect the film 's future international box office success.

Even though later studies revealed that real opening weekend was closer to \$98 million, this marketing move was undoubtedly a success ranking Age of Extinction the highest grossing motion picture of summer 2014 and the second highest of the year globally, despite its poor reviews.

Budget - Motion Picture Association of America divides budget into two separate figures – Production budget and Prints and Advertising (P&A) budget. Most commonly production budget is being published, while P&A costs are only estimated. Production budget refers to amount of money it cost to make the movie including pre-production, film and post-production, but excluding distribution costs. The average cost of a major studio movie was about $65 million when the MPAA stopped tracking the number in 2006 and has risen since then. The most expensive films each year commonly cost more than $200 million and there are a few films costing over $300 million.

P&A budget consists of prints cost – physical copy of a motion picture, and distribution costs. Most of the money is spent on TV, but radio, newspapers and magazines, the Internet and in-theatre advertising are also very important. The average P&A budget for a major studio release was nearly $36 million the last time the MPAA reported the figure in 2007, therefore it is often being estimated as 50% of the production budget. They have since stopped tracking advertising numbers. However, films with $100 million advertising budgets are now common, especially for tentpole releases. Even midlevel releases will spend $40 million to $50 million on advertising. (MPAA, 2019)

4.3 Qualitative data

Genre – one of key variables that affects box office is naturally movie genre. fantasy/science fiction movies used to generate the highest box office (Redfern, 2012). In fact, of the top 20 grossing movies between 1991 and 2010, eight were fantasy/science fiction. On the other hand, as Redfern notes, the highest grossing genre at that time was family with a total gross of $23.6 billion. To present some recent data, there are eight sci-fi movies and 15 adventure-ranked movies among the top 20 Most Profitable Movies (vast majority of films has more than one genre characteristic), based on absolute profit on worldwide gross (The Numbers, 2019).

To stress the potential effect genre variable may carry, we can provide an example of the period of 1972-1978, when Science Fiction-Horror genre was clearly a very popular genre and adopting this story type led to an increase in rentals of nearly $6 million (Litman 1983). Even though this effect was connected with more than 40 years old trend which has little meaning in recent film market, we might find a similar trend connected with different genre in our analysis.

MPAA ratings - The Motion Picture Association of America's (MPAA) film-rating system categorizes movie content as suitable for certain audiences.

- G – general audience, all ages admitted
- PG – parental guidance suggested, some material may not be suitable for children
- PG-13 - parents strongly cautioned, some material may be inappropriate for children under 13
- R – restricted, under 17 requires accompanying parent or adult guardian
- NC-17 - adults only, no one 17 and under admitted

MPAA rating affects various other variables of a motion picture. Directors' contracts almost always contain a requirement that the film qualify for a MPAA rating of no more restrictive than R, PG-13, etc., unless otherwise stipulated by the studio (Ellis and Conaway, 2015). Movies rated PG-13 had higher average gross domestic revenues per movie than movies in any other ratings category. There are fourteen PG-13, five PG and one G rated movies among the top 20 Most Profitable Movies, based on absolute profit on worldwide gross (The Numbers, 2019).

PG-13 rating in particular is expected to be an important factor for box office success. According to some past studies, PG-13 rating can result in $15 million - $34 million more at the box office compared to the more restrictive R rating (Ellis and Conaway 2015).

To our best concern there have been only few studies evaluating **running time**'s effect on profitability. However, British consumers consider optimal movie length to be below 2 hours (91-120 minutes) (YouGov, 2015) nevertheless, an average of top 20 grossing film's running time was 128.55 minutes in the same year (127 mins in 2019)[12].

Release date is the last of our observed independent variables. There are commonly two seasons being referred to in relation with movie releases. The 'tentpole' peak season and the 'dump' season.

Pre-Christmas period (November through December) and pre-summer season (May and June) are typically referred to as tentpole season. These months are also connected with major U.S. holiday like Christmas or Memorial Day. Motion pictures of high expected revenues are being released during these months, often being accompanied by large budgets and heavy promotion and supported by distribution of related

[12] Based on our dataset

merchandise such as toys or videogames. These projects are expected to carry their studio's profitability

Summer months (July and August) and the beginning of the year are called the dump season. January is usually considered a dump season in most of the service and entertainment industries, while August and September are usually characterized as movie-specific dump season. Films released during these seasons are often of lower commercial and critical expectations, films with less prominent stars or of worse marketable genres like horror movies.

5. Hypotheses and variables

Dataset available for our thesis allows for evaluation of several hypotheses related to potential economic performance of a motion picture on a sample containing domestic films. In the following section we will go through the set of hypotheses we want to evaluate in our study and relevant variables that will help us obtain the answers.

5.1.Hypotheses

First, the relation between domestic opening weekend box office and overall domestic gross profit will be tested. Strong correlation is expected to arise here, as opening weekend is one of the basic metrics determining film's success.

Based on previous research, qualitative and financial data will be tested in relationship with revenue. It is suggested that at least some of key attributes, such are budget, genre, MPAA rating or running time are supposed to influence film's performance. As the highest-grossing list[13] is dominated by recent films, it is logical to test whether this is caused by steadily increasing production and marketing budgets, just as it is expected, that certain genres are more popular among consumers than others.

The set of hypotheses may be summarized as follows:

H1: Opening weekend box office positively affects gross film's revenue

Variable *domestic_opening* captures the financial performance of a motion picture over the first weekend when it is being screened. It is presumable that a movie that performs well during it's opening weekend will attract more moviegoers in the following period of time including those who were undecided during the advertising phase or were not interested into this particular picture at all. As we have already mentioned before an opening weekend box office may be taken as a motion picture's selling point as well.

[13] List of top grossing movies by boxofficemojo.com

H2: Higher production budget leads to better performance of a motion picture

Variable *budget* stands for a sum of total production costs. Greater amount of money spent for advertising should lead to greater awareness of the certain motion picture among consumers and thus is expected to attract a wider base of moviegoers to the cinemas. Similarly, a greater production budget gives producers' possibility to hire renowned actors, enhance cinematic effects, shoot in desired locations around the world and boost the quality of a motion picture in general which should consequently attract additional moviegoers. Another reasoning may be that a higher production budget may be used to produce the film with advanced technology such as for example IMAX 3D technology and thus charge higher ticket prices.

H3: Films of certain genres perform better than others

Genre binary variables were included since we expect some genres to attract more consumers than others. According to statista.com the most numerous age categories of moviegoers (in terms of domestic market) are 25-39 and 60+ respectively. Even though these categories are not subject of any MPAA restrictions, some genres (horror, crime, thriller...) may restrain other age groups and different genres like drama or romantic may not attract another numerous group of consumers – families with young children. Therefore, we expect certain genres to be more popular among wider audience and accordingly more profitable.

H4: MPAA rating is key variable affecting film's performance

Mpaa dummy variables were added because MPAA restrictions have impact on the population of moviegoers allowed to see a particular motion picture. As long as certain MPAA ratings restrict specific groups of consumers from purchasing a ticket the *mpaa* dummy variables is expected to be statistically significant in our model.

H5: Films with running time below 2 hours perform better

In recent years, popular movie releases often tend to be over 2 hours and sometimes even over 3 hours long. Based on previous research excess movie length may limit consumer's experience. To prove YouGov's 2015 study claiming optimal movie length to be below 2 hours right, variables *running_time* and *running_time_sq* were applied. *Running_time* shall show us whether the length of a motion picture affects its future financial performance and its square product shall help us find the optimum that we expect to be below 120 minutes.

H6: Films released before Christmas and summer holidays perform better

Following previous studies, we assume similar trend will appear in our analysis. We expect movies in our sample to perform better during the peak seasons and on the other side provide poorer results in terms of revenue during the 'dump' months.

5.2.Independent variables

The collected data on variables may be separated into two aforementioned groups: numerical, such as running time and financial data on budget and gross revenues, and qualitative data on MPAA rating and genres.

Other variables added into our model were dummy variables controlling for seasons. We added season variable for pre-Christmas releases as this factor was found significant in past researches and for summer and pre-summer seasons as well.

Opening weekend box office variable – *domestic_opening* – shall show us to what extent gross domestic revenues of a motion picture are affected by its performance during the first weekend since release.

Movie budget was found to have non-negligible impact on production costs that are positively related to marketing expenses (Prag and Casavant, 1994) which leads to a conclusion that movies with greater budget are advertised more which should presumably affect future revenues (H2).

Another independent variable, the MPAA rating, should reveal us whether future revenues were affected by viewer restrictions. This variable was transformed into 5 separate dummy variables – *pg13*, *pg, r, g,* and *nc17* – which should serve better to our purpose. In the model itself only 4 MPAA dummy variables were used in order to avoid the dummy variable trap. Statistically PG-13 is the most frequent value among observations, thus it will be used as the base group.

Running_time variable stands for the length of a movie expressed in minutes and its square product helps us to visualize its effect more accurate.

Last group of variables are binary variables for individual movie genres. There are in total 18 genre variables in our model that control for their specific effects on movie profitability. As these are binary variables that do not exclude each other and rather provide a combined effect all of them were included into the model.

5.3. Dependent variable

For our purpose, the domestic box office first run release revenue will serve as a dependent variable. The individual movie revenues will be collected from *Box Office Mojo*[14], a website that tracks box office revenue in a systematic, algorithmic way. The analysis will cover the movies from dataset that closed their first run theatrical release in the years 2005 through 2019.

[14] Source: www.boxofficemojo.com

Table 3: Independent variables

Type	Variable	Description
numerical	*budget*	Production and Advertisement budget
	domestic_opening	Domestic openenig weekend box office
	running_time	Running time expressed in minutes
	running_time_sq	Square product of *running_time*
categorical	*pg13*	Dummy variable for PG-13 rating
	pg	d.v. for PG rating
	r	d.v. for R rating
	g	d.v. for G rating
	nc17	d.v. for NC-17 rating
	summ	d.v. for summer releases
	spring	d.v. for pre-summer releases
	christm	d.v. for pre-Christmas releases
	action	control variable for genre action
	adventure	c.v. for genre adventure
	animation	c.v. for genre animation
	biography	c.v. for genre biography
	comedy	c.v. for genre comedy
	crime	c.v. for genre crime
	drama	c.v. for genre drama
	family	c.v. for genre family
	fantasy	c.v. for genre fantasy
	history	c.v. for genre history
	horror	c.v for genre horror
	musical	c.v. for genre musical
	mystery	c.v. for genre mystery
	romance	c.v. for genre romance
	scifi	c.v for genre sci-fi
	short	c.v. for genre short
	thriller	c.v. for genre thriller
	war	c.v. for genre war
	western	c.v. for genre western

6. Descriptive statistics

In this section we will present summary statistics of our data, beginning with financial data through all the remaining categories. We will evaluate distribution of our data as well as representation of individual characteristics in our sample.

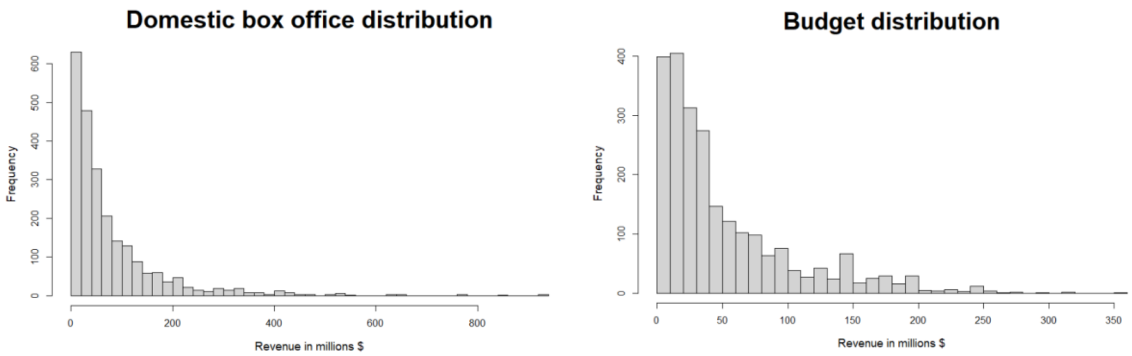
Table 4 shows us summary statistics on our financial data – domestic box office revenues expressed in millions of dollars, budget estimates expressed in millions of dollars and summary statistics for a new variable created as a product of revenue-budget ratio.

Table 4: Revenues and budgets summary statistics in millions of U.S. dollars

Revenues and budgets						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
Domestic box office revenues	0.0022	18.32	43.34	75.75	94.13	936.66
Budget	0.0011	16.00	35.00	53.18	73.00	356.00
Revenue/budget	0.000	0.620	1.199	8.532	2.252	7194.587
n=2353						

At our first sight it is obvious that both revenues and budgets are extremely skewed right judging by the great difference between the mean and the median value, which confirms previous studies (Teti 2013, De Vany 1999). The fact is even more obvious from the graphical representation of their distribution in Figure 3.

Figure 3: Distribution histograms in millions of $



Extremely high values of some observations draw the mean value far from the more accurate median which is typical for movie data samples.

Last row from Table 4 shows us summary statistics for the revenue-budget ratio, which was used as an approximation for profitability. The median value being relatively close to 1 indicates, that nearly half of observations in our sample are close to or below the breakeven point. Minimum value of 0 means, that there was at least 1 film in our list that lost (almost) all its production budget. Table 5 divides our dataset into profitability categories. Roughly 42% of all the movies in our sample could not recoup their production costs. According to MPA, the average advertising costs may be estimated as 50% of production costs. Applying such estimation and a certain amount of simplification we may consider film ‘a success’ if it generates revenue twice the amount of its production costs. Therefore, there are around 30% of ‘successful’ motion pictures among our observations.

Table 5 : Profitability quantiles

Profitability

	>1	>1.5	>2	>10
count	1364	941	681	59
percentage	57.97%	39.99%	28.94%	2.51%

n=2353

Table 6 gives us summary on genre characteristic distribution in our data sample. Majority of movies are being characterized as genre-hybrid and has more than one genre characteristic. It is of no difference in our data set.

Slightly over half of all observations were classified as ‘drama’ which is a common feature with other studies and might be probably explained, with a dash of exaggeration, that there is a bit of drama in every movie and that producers tend to use as much relevant genre characteristics as possible – such as keywords – for easier traceability. Apart from that, great diversity is being presented by our dataset in terms of genres with a sufficient representation of each one.

Table 6: Genres

Genres		
genre	count in n	percentage
action	591	25.12%
adventure	701	29.79%
animation	194	8.24%
comedy	889	37.78%
biography	198	8.41%
crime	370	15.72%
drama	1188	50.49%
family	362	15.38%
history	124	5.27%
fantasy	409	17.38%
horror	258	10.96%
musical	75	3.19%
mystery	282	11.98%
romance	479	20.36%
sci-fi	355	15.09%
thriller	712	30.26%
war	103	4.38%
western	41	1.74%

n=2353

Next characteristic examined in our study is seasonality. According Table 7, nearly one quarter of all our films were released in the pre-Christmas period, specifically in November and December, around one fifth during the summer dump season, and around 13% of our observations has a release date in May or June. With a simple logic that around 8.3% of movies would be released each month when distributed evenly, these numbers suggest increased competition before Christmas and during summer months but slightly decreased competition in the pre-summer season.

Table 7: Seasonality

Seasonality		
	count in n	percentage
pre-summer	307	13.05%
summer	427	18.15%
pre-Christmas	586	24.90%
n=2353		

Last summary statistic depicting MPAA ratings distribution in our data sample is presented in the following Table 8.

Table 8: MPAA rating

MPAA ratings		
rating	count in n	percentage
G	25	1.06%
PG	414	17.59%
PG-13	988	41.99%
R	922	39.18%
NC-17	4	0.17%
n=2352		

Roughly 80% of our data sample is represented by films with restrictions PG-13 or R which is very common for recent film market as majority of tentpole releases are of no lower than PG-13 restriction. These restrictions allow everyone to attend the theatre, while giving producers nearly a 'free hand' in terms of movie content such as adult activity, harsh language, intense graphic violence, drug abuse and nudity (MPA 2020). Movies rated G or NC-17 have rather poor representation in our sample which was expected as G rating stands for children-specific movies which are not common among recent theatrical releases and has moved to different platforms such are web-TVs. On the other side, NC-17 rating stands for features like 'extreme horror violence, sadistic graphic violence, bizarre sexuality/nudity or R rated material involving teen characters' which has been proved negatively affecting film revenues and is rather a feature of 'artistic' pictures in cinematography.

7. Methodology

In the following section we will describe the methodology we follow in our analysis and its theoretical background necessary for correct application and execution.

7.1. The OLS Model

First, we will recall the Ordinary Least Square (OLS) regression, formulate the model used in the numerical simulations and discuss the assumptions required for analysis. Further on we will discuss several limitations that may arise while using OLS and treatment methods applied to deal with them. We operate on a cross-sectional data set. In order to analyse the effect on predefined explanatory variables on the financial performance of a motion picture several models will be introduced:

$$y_i = \beta_0 + \boldsymbol{X}\beta + u_i, i = 1, \dots n \qquad \text{(a)}$$

where y_i stands for the domestic revenue for the motion picture i, **X** is the vector of independent variables including financial and qualitative factors and u_i represents the error term or disturbance, which is assumed to be normally and identically independently distributed and n is the number of observations in our data sample.
Regarding validity of the cross-sectional model, the Multiple Linear Regression (MLR) assumptions must be satisfied. According to Gauss-Markov theorem, if these assumptions are met, OLS estimators of parameter β in (a) are best linear unbiased estimators (BLUE):

- MLR.1 – parameters must be linear
- MLR.2 – random sampling
- MLR.3 – no perfect collinearity
- MLR.4 - E(u_i) = 0 zero mean error
- MLR.5 - var(u_i) = E($u_i{}^2$) = σ^2 homoskedasticity

Even though we used quadratic forms of some variables, the linearity assumption is satisfied as the model is linear in parameters. Random sampling assumption was satisfied by reducing the total amount of observation by the eligibility conditions. Inclusion of maximum possible explanatory variables into the model allows to account for zero correlation with the error term. Additionally, no mutual collinearity was found among our independent variables through the VIF (variance inflation factor) method. The VIF for slope coefficient j is simply $VIF_j = 1/(1 - R_j^2)$, precisely the term in $Var(\hat{\beta}_j)$ that is determined by correlation between x_j and the other explanatory variables and value of VIF > 10 is usually signalling multicollinearity is present (Woolridge, 2012).

In the next step we test for normality of our residual. The Shapiro-Wilk normality test tests the null hypothesis that the data tested are normally distributed. If the p-value is less than the chosen alpha level, the null hypothesis has to be rejected.

Furthermore, we need to test whether heteroskedasticity presence do not affect our model. This is important for OLS estimators to be BLUE. To evaluate this, we will use the standard Breusch-Pagan test. Prior to any testing the consistency of estimators needs to be checked – for that we will inspect our summary statistics such are range, variance or standard deviation over the years observed.

Breusch-Pagan test

The purpose of Breusch-Pagan test is to evaluate whether the variance of error term from our regression is dependent on the values of our independent variables. In such case, heteroskedasticity would be present.

First, the OLS regression is run to obtain the $\hat{u}$ residuals. Then, $\hat{u}^2$ is regressed on all our independent variables. With an application of the F-test, we assign the null hypothesis to homoskedasticity and its alternative to heteroskedasticity. In case we reject the null hypothesis at 1% significance level, heteroskedasticity is present in our model violating the MLR.5 assumption. In such case we need to treat for its presence using robust standard errors.

8. Results

In this section we present the estimated results of our regression analysis and try to interpret them in terms of our data. The R software was used to proceed the data and estimate the coefficients. We used the OLS regression to estimate the effect of several determinants on box office revenues and try to discover recent pattern of motion picture's profitability. Similarly to other studies following this topic, we log-transform the variables on financial data in order to normalize for the outlying observations with extreme values, which are typical for movie data sets.

8.1. OLS regression results

Here, we present the results of the OLS regression model. We will try to analyse the impact on several factors possibly affecting the overall revenues motion pictures generate in the domestic market. Through our analysis we will follow the theoretical background provided in the previous chapter.

First of all, we apply the variance inflation factor method to test for multicollinearity in our regression. VIF test provides us with indexes measuring to what extent the variance in the estimated regression coefficients increased due to collinearity. Usually a VIF value greater than 10 signals problem. In our case, none of the coefficients in any of presented models have this issue, except for square product of *running_time,* which is logical, thus we can conclude that multicollinearity is not present. Results of our VIF tests are available in Appendix B.

Second, we check the normality of our residual running a Shapiro-Wilk test. Even though the test provided a low p-value and therefore the null hypothesis cannot be accepted, according to Gauss-Markov theorem, the OLS estimate is BLUE as the errors

- Have zero mean
- Are uncorrelated
- Have constant variance

No condition of normality is stated (or even any condition that the errors are IID) and these three conditions have been met – we tested for residuals correlation through autocorrelation function (ACF) and partial ACF. In order to reduce the deviation from normality we log-transformed our dependent variable. The t-test is described as a robust test with respect to the assumption of normality therefore, some deviation shall not cause any significant problems. However, we are aware that we must interpret the results cautiously.

Since assumptions MLR.1 to MLR.4 are satisfied, we can claim our estimates are consistent and further application of tests is possible. All of our parameters are linear, there is no perfect collinearity among our variables as provided with the VIF method and we try to control for as many as possible independent variables in our model in order to satisfy the zero conditional mean assumption.

In the next step we run the Breusch-Pagan test on all our models to find out, whether our error term suffers from heteroskedasticity. Regrettably, all Breusch-Pagan tests provided low p-value therefore we must reject the null hypothesis as our models exhibit heteroskedasticity. In order to control for that, we must apply robust standard errors for valid inference.

The following Table 9 presents an overview of all our log-log regressions. Independent variables that appeared insignificant in all our models were removed from the overview to make it clearer. Full results are available in Appendix A. First five columns show results for the whole sample as we expect the results to be relatively consistent over the time and the last three columns give us the comparison among three 5-year periods from our sample to give us the opportunity to discuss the possible trends over the last 15 years. First column represents results of regression without the control variables for seasonality, MPAA rating and running time that were added in columns (2), (3) and (4) respectively. Full regression ran on full sample is represented by column (5) that was highlighted.

Last two rows in our table give us results for R squared and adjusted R squared, which are measures for goodness-of-fit of our model, providing us information on how much variance in our regression is explained by our independent variables. As we can see our values of these statistics are quite high (even compared to previous studies), which is mainly caused by the first two variables – *log(budget)* and *log(domestic_opening)* as they have a great deal on explaining our dependent variable alone as expected. The adjusted R2, which is adjusted for the number of covariates, shows the best fit for the model with all the control variables.

Using the F-test helps us determine the joint significance of our variables used in the model. The test provided us with p-value equal to 0.000002, therefore we cannot reject the null hypothesis, 'all coefficients are equal to 0 at 1% significance level'. Therefore, even though some of the variables are independently insignificant, we may keep them in our model.

The main object of our attention will be the highlighted full model, estimated by the following equation:

$$
\begin{aligned}
log(domestic) = {} & \beta_0 + \beta_1 log(budget) + \beta_2\, log(domestic\ opening) + \beta_3 action \\
& + \beta_4 adventure + \beta_4 animation + \beta_5 biography + \beta_5 comedy + \beta_6 crime \\
& + \beta_7 drama + \beta_8 family + \beta_9 fantasy + \beta_{10} history + \beta_{11} horror \\
& + \beta_{12} musical + \beta_{13} mystery + \beta_{14} romance + \beta_{14} scifi + \beta_{15} thriller \\
& + \beta_{16} war + \beta_{17} western + \delta_1 spring + \delta_2 summer + \delta_3 christmas + \alpha_1 PG \\
& + \alpha_2 R + \alpha_3 NC + \alpha_4 G + \gamma_1 runningtime + \gamma_2 runningtime^2 + \mu
\end{aligned}
$$

Table 9: Regression results

log-log OLS regression result

independent v.	dependent variable							
	log(domestic)							
			full sample			2005-2009	2010-2014	2015-2019
	(1)	(2)	(3)	(4)	(5)			
Constant	3.43700***	3.94210***	3.88970***	3.83590***	**3.02070*****	1.2808	3.76290***	5.60420***
	(0.552320)	(0.545640)	(0.572130)	(0.558110)	**(0.739860)**	(1.363700)	(1.317200)	(1.118200)
log(budget)	0.27389***	0.22248***	0.22105***	0.14762***	**0.14379*****	0.18458**	0.0975	0.15980***
	(0.038410)	(0.038140)	(0.039060)	(0.041070)	**(0.041120)**	(0.082590)	(0.074010)	(0.050880)
log(domestic_opening)	0.59205***	0.60535***	0.60876***	0.60593***	**0.60642*****	0.61403***	0.64548***	0.47293***
	(0.020200)	(0.019640)	(0.019870)	(0.019540)	**(0.019490)**	(0.035200)	(0.028700)	(0.047550)
action	-0.27044***	-0.22056***	-0.21660***	-0.19140***	**-0.19331*****	-0.16211*	-0.19325**	-0.26827***
	(0.053630)	(0.053980)	(0.053940)	(0.053720)	**(0.053650)**	(0.087870)	(0.097600)	(0.087030)
animation	0.27411***	0.27416***	0.26850***	0.39617***	**0.41629*****	0.21518	0.63320***	0.29850***
	(0.064260)	(0.066430)	(0.066100)	(0.069800)	**(0.072410)**	(0.149240)	(0.134390)	(0.100590)
biography	0.36488***	0.33690***	0.34677***	0.31230***	**0.30206*****	-0.18182	0.80345***	0.16845***
	(0.111610)	(0.105220)	(0.106310)	(0.104820)	**(0.104940)**	(0.169000)	(0.213100)	(0.147050)
crime	-0.20737***	-0.18797***	-0.20439***	-0.19477***	**-0.19572*****	-0.24218**	-0.17783*	-0.10329
	(0.059470)	(0.059030)	(0.059970)	(0.058800)	**(0.058730)**	(0.108670)	(0.102370)	(0.091160)
drama	0.11316**	0.08995**	0.09077**	0.01177	**0.00786**	0.04467	-0.05222	-0.05794
	(0.044950)	(0.044810)	(0.044980)	(0.046170)	**(0.046270)**	(0.086360)	(0.085990)	(0.072000)
family	-0.25486***	-0.22886***	-0.15445	-0.06098	**-0.05179**	-0.14424	0.07869	-0.06562
	(0.056070)	(0.059290)	(0.102490)	(0.101000)	**(0.101160)**	(0.204920)	(0.165700)	(0.168500)
fantasy	-0.06918	-0.07058	-0.06247	-0.0596	**-0.06111**	-0.05377	0.04757	-0.21636***
	(0.047630)	(0.048420)	(0.048100)	(0.046740)	**(0.046710)**	(0.086610)	(0.091390)	(0.074800)
history	0.1953	0.22386*	0.21921*	0.16357	**0.16806**	0.32221	0.09215	0.06211
	(0.134110)	(0.126540)	(0.126930)	(0.124820)	**(0.124780)**	(0.198270)	(0.244180)	(0.212130)
horror	-0.26861***	-0.23892***	-0.25354***	-0.15558**	**-0.14195****	-0.29161**	-0.12433	-0.03055
	(0.062240)	(0.061590)	(0.061920)	(0.062570)	**(0.062420)**	(0.126420)	(0.100920)	(0.113780)
musical	0.35595***	0.24522*	0.25040*	0.18931	**0.17683**	-0.06673	0.16449	0.37088**
	(0.130740)	(0.133140)	(0.133200)	(0.130730)	**(0.130700)**	(0.181010)	(0.306630)	(0.162310)
romance	-0.14334**	-0.13652**	-0.12791**	-0.13994**	**-0.14464****	-0.16671	-0.04161	-0.24025**
	(0.056810)	(0.055860)	(0.057500)	(0.056830)	**(0.056840)**	(0.113300)	(0.089690)	(0.095390)
sci.fi	-0.09425**	-0.07684*	-0.07186*	-0.07932**	**-0.08041****	-0.31226***	0.04648	0.02018
	(0.040790)	(0.039460)	(0.039620)	(0.039140)	**(0.039150)**	(0.074550)	(0.066360)	(0.074400)
spring1		0.17719***	0.17627***	0.16450***	**0.16408*****	0.12912	0.20179**	0.15560**
		(0.054140)	(0.054270)	(0.052880)	**(0.052880)**	(0.111830)	(0.093050)	(0.069920)
christm1		0.50318***	0.50493***	0.46992***	**0.47478*****	0.45714***	0.60533***	0.25122***
		(0.054190)	(0.054160)	(0.052390)	**(0.052650)**	(0.103800)	(0.095840)	(0.080380)
PG			-0.07176	-0.03773	**-0.04193**	0.04908	-0.29733*	0.09434
			(0.097960)	(0.094970)	**(0.094970)**	(0.188380)	(0.151950)	(0.164960)
R			0.06171	0.028	**0.02853**	0.10396	0.02102	-0.12282*
			(0.048580)	(0.048490)	**(0.048480)**	(0.093410)	(0.084310)	(0.072000)
running_time				0.01213***	**0.02692*****	0.04594***	0.01602	0.01536
				(0.001360)	**(0.008630)**	(0.015450)	(0.013810)	(0.014290)
running_time_sq					**-0.00006***	-0.00015**	-0.00002	-0.000001
					(0.000030)	(0.000060)	(0.000060)	(0.000060)
Observations			2353			782	877	694
R^2	0.725	0.738	0.739	0.748	0.748	0.744	0.785	0.686
Adj. R^2	0.722	0.735	0.736	0.745	0.745	0.734	0.778	0.669

Note: *p<0.1; **p<0.05; ***p<0.01

8.1.1. Interpretation of results

To summarize our variables and their effect on domestic box office revenue we will go through the table from the top to the bottom. As expected, box office positively influences box office revenue, specifically, increasing budget by 1% results in 0.14% increase in box office revenue. As suggested by previous studies, sufficient production budget gives motion picture a better chance to 'make it through' (Teti, 2013), increase its quality by investing into popular cast, hight-tech effects, audio and video quality and allows for wider opening on a higher number of screens around the country (De Vany and Walls, 1996). Even though our *budget* variable does not account for advertisement budget, we can logically assume that with increased production budget the P&A budget increases as well as the P&A budget is being casually estimated as roughly 50% of production budget. Based on this assumption, greater budget gives distributors option to invest heavily into promotion and extend the media hype. Greater investments into advertisement was found to influence remarkably consumer's quality assessment (Thurau et al., 2006).

Domestic opening is another variable that positively affects box office revenues. As 'success breeds success' it was just logical that this variable would have non-negligible impact on total revenues, but the extent might be surprising as the variable itself explains roughly 40% of variance in the dependent variable. Great opening weekend is considered a big selling point in the movie industry and has the ability to give a kick to the specific movie in terms of attracting undecided audience into the theatres, attracting second-channels (TV, online streaming channels, etc.) distributors to purchase screening rights or raise awareness in other markets (the rest of the world).

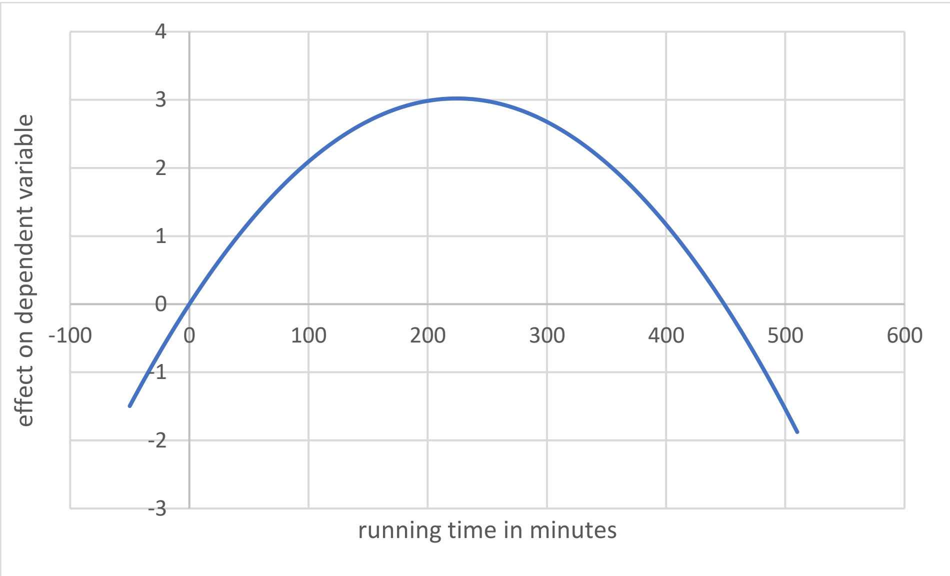
Interpretation of effects of individual genres is a bit more complex. In our main model, seven genres (out of 18 in total) provide significant effects on our dependent variable. However, regarding that majority of observed movies are characterized with more than one single genre (often three or even four), the total effect is rather a combination of these individual genre effects. Nevertheless, in order to comment on genre-specific effects, *action* variable provided a surprising result having a negative coefficient and thus causing a 19% decrease in total revenue. This effect may be caused by excessive financial burden action movies carry in terms of expensive special effects etc. Same story goes with *sci.fi* variable even though to much less extent (around 8% decrease). However, this might be also connected with decreasing popularity of this specific genre as major motion pictures such are Marvel's stories are dropping this tag and rather use a 'fantasy' characteristic as 'science' is not an object of these stories, really. *Animation* and *Biography* provided significant positive effect on revenues. Animation in general is a very popular genre as many animated movies are released every year and scoring great positions in top-charts. Another reasoning behind the positive effect might be the attempt of producers to attract older audience to animated movies (85% of animated movies in our sample are rated PG). Finally, animated movies have without doubt less production costs in terms of special effects, cast etc. (even though main characters are dubbed by famous actors, the human factor is still significantly reduced). *Biography*'s strong positive effect granting roughly 30% increase in revenues might be a result of recent year's trend. There were many tentpole releases recently following the lives of famous personalities, singers or bands etc. which may account for this effect in combination with *musical* which appeared significant in some of the models as well (Bohemian Rhapsody, Rocketman, ...). *Horror, romance* or *crime* are genres usually connected with negative effect on revenues, being frequently connected with the 'dump' release season as well.

Next, we will focus on the seasonality effect on box office revenues. Originally, we had 3 categories: *christm*, representing pre-Christmas releases, *spring*, representing spring tentpole season (May and June) and *summ*, representing the summer 'dump' season. As the *sum*m variable was found insignificant in all the models it was dropped out. Remaining control variables provided explicit results. Releasing a movie in the pre-Christmas period results in over 47% increase in term of revenues compared to other months even accounting for the increased competition and spring releases gross additional 16% in revenues holding other variables constant. Again, this effect was observed in previous studies and can be explained by studios placing their highly expected releases into these months.

Similar case arises with the MPAA rating dummy variables. Originally, we had 5 variables, (G, PG, PG-13, R, NC-17) intentionally excluding the PG-13 variable as a 'base'. As G and NC-17 were found insignificant in all our models, we decided to drop them from the summary. Even though PG and R control variables were not found to significantly affect theatrical revenues of motion pictures in our sample as well (in our full model) nevertheless, as their t-value was close to 1.5 in our regression we decided to remain them. *R* variable has a positive coefficient close to 0.03, indicating almost no change in comparison to PG-13 rating which satisfies our previous assumption of these two rating categories being the most profitable ones. *PG* has a negative coefficient of only 0.04 which results in better-than-expected performance of this category in comparison with the base.

Finally, running time element was examined in our regression. Both *running_time* and *running_time_sq* appeared to have a significant impact on our dependent variable (cubic product was not found to be statistically significant or affecting the result). To interpret the results better a graphical depiction of the quadratic function is provided:

Graph 1: running time coeffitient function plot



Combining the results from our regression and from the plot of the function, each additional minute of running time increases domestic box office revenue up to roughly 225 minutes when it would start to decrease however, longest film from our dataset has a running time of 201 minutes therefore, the conclusion is that the longer the running time, the higher the revenue. This effect may be easily assigned to the prolonging average running time of highly expected theatrical releases while on the other side, shorter average running time of less profitable pictures.

As to the inter-annual comparison we can conclude that the general trend remained relatively consistent with some differences mainly in significance levels of individual factors. The general trend that can be observed are constantly increasing values of the overall revenues as the industry is steadily growing, introducing new technologies and grossing record amounts year after year.

8.2.Hypotheses evaluation

H1: Opening weekend box office positively affects gross film's revenue

Opening weekend box office revenues were found positively affecting total revenue of a motion picture at 1% significance level. An increase of opening weekend box office by 1% results in 0.6% increase in total domestic revenue.

H2: Higher production budget leads to better performance of a motion picture

Hypothesis number 2 was confirmed as well as production budget positively correlates with our dependent variable at 1% significance level.

H3: Films of certain genres perform better than others

At least one of listed genre variables was consistently improving theatrical revenues and at least one of them was affecting them negatively in all our models provided. Therefore, we are able to confirm hypothesis number 3 as well however, as we have already mentioned the individual effects of specific genres are rather tricky to evaluate and one shall focus more on their combined product in order to assess their true effect on the dependent variable. Even though there can be no doubts that certain genres solely decrease the potential profitability of a motion picture and relate to rather humbly performing pictures of mixed qualities.

H4: MPAA rating is key variable affecting film's performance

Hypothesis number 4 cannot be supported even at 10% significance level as none of our independent dummy variables for MPAA rating were found statistically significant in our model. However, limited conclusions could be made about their effect.

H5: Films with running time below 2 hours perform better

Even though running time with its square product was found statistically significant in our regression, it provided quite the opposite result than we expected, therefore we cannot support this hypothesis either. Therefore, we may conclude the YouGov's 2015 study results were likely based on subjective feeling of respondents that might not felt comfortable attending over 120 minutes long movie however, in terms of financial performance this finding was irrelevant.

H6: Films released before Christmas and summer holidays perform better

Finally, both parts of hypothesis number 6 can be supported at 1% significance level as both peak seasons attribute significantly to box office revenues and thus we may conclude that this specific seasonality holds over decades through many studies performed on movie industry.

Overall, all the hypotheses were tested resulting in confirmation of 4 out of total 6 hypotheses. Hypotheses number 4 could not be supported due to statistical insignificance of its corresponding variables and hypothesis number 5 had to be rejected as the opposite was proven.

9. Conclusion

The objective of this thesis was to examine the factors affecting the box office revenues of motion pictures released at the domestic market using the updated data including recent years. In the beginning, the North American film industry market was described and characterized. Based on the literature review provided, some basic assumptions were set and several attitudes towards analysing our data were presented. These thoughts were used to create our set of hypotheses and to choose our methodological approach.

Four of our total six hypotheses were confirmed providing a comparison to previous research and possibly shedding light into the recent aspects of the industry. Results obtained in our analysis may serve for future evaluation and possible decision making of film producers.

The novelty of our analysis lies within incorporating running time into our model which has been, to our best concern, rarely observed in similar studies alongside with considering joint effects of genres on the revenue-generating process.

As the industry itself grows every year some features remain constant over time while the others follow trends that may change and therefore we find it meaningful to provide such analysis with as up to date data as possible in order to keep the results meaningful for present production.

In terms of recommendations our thesis provided evidence that special emphasis should be put on arranging for a successful opening weekend as this has showed to be an important indicator of future motion picture's performance. However, alternative or advanced methods like simultaneous model equation method could be used to explain the opening weekend's effect on total revenues in more detailed way. Moreover, the explanatory power of our model could be enhanced by incorporating additional data on P&A expenditures which might be a possible subject for future analysis.

List of references

Articles

Berg, J., Raddick, M.J. (2017) First You Get the Money, Then You Get the Reviews, Then You Get the Internet Comments: A Quantitative Examination of the Relationship Between Critics, Viewers, and Box Office Success. Quarterly Review of Film and Video, 34(2), 101 – 129

Dantes, A., Finney, A., (2015) The international film business: a market guide beyond Hollywood. CHOICE: Current Reviews for Academic Libraries, 52(10), 1663-1663

De Vany, W., Walls D.,(1996) Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry. The Economic Journal, 106(439), 1493-1514

De Vany, W., Walls D. (1999) Uncertainty in the Movie Industry: Does Star Power Reduce the Terror of the Box Office? Journal of Cultural Economics, 23(4), 285

Dvorakova, A. (2017) Faktory finančního úspěchu filmu na evropském filmovém trhu Availible at: https://is.muni.cz/th/ug3z4 Master thesis. Masarykova univerzita, Ekonomicko-správní fakulta. Supervisor Daniel Němec.

Ellis, D., Conaway B., (2015), Do MPAA Ratings Affect Box Office Revenues? Academy of Business Research Journal, 1, 64-88

Hennig-Thurau, T., Houston B.M, Walsch G., (2007) Determinants of motion picture box office and profitability: an interrelationship approach. Review of Managerial Science, 1(1), 65-92

Hyatt, A., Johnson, D. K. N. (2016) Why Do You Like What You Like? The Impact of Previous Netflix User Ratings on Subsequent User Experience. Colorado College Working Paper, availible at SSRN: https://ssrn.com/abstract=2807475 or http://dx.doi.org/10.2139/ssrn.2807475

Litman, B. R., (1983), Predicting Success of Theatrical Movies: An Empirical Study. The Journal of Popular Culture, 16: 159-175

Litman, B. R.,Kohl, L.S. (1989) Predicting Financial Success of Motion Pictures: The '80s Experience. Journal of Media Economics, 2(2), 35-50

Lubbers, C. A., Adams W.J (2004) Merchandising in the Major Motion Picture Industry: Creating Brand Synergy and Revenue Streams. Journal of Promotion Management, 10(1/2), 55-63

Lubbers, C., Mika, D. (2018) User-generated reviews, genre, mpaa rating and time in theater as correlates of movie box-office revenues. Quarterly Review of Business Disciplines. 5. 163-182.

Prag, J., Casavant J., (1994). An Empirical Study of the Determinants of Revenues and Marketing Expenditures in the Motion Picture Industry. Journal of Cultural Economics, 18(3), 217

Redfern, N. (2012) Genre trends at the US box office, 1991 to 2010. European Journal of American Culture, 31(2), 145-167

Sharda, R., Delen D., (2006) Predicting box-office success of motion pictures with neural networks. Expert Systems With Applications, 30(2), 243-254
Simonet, T. (1981) M. P. A. A. Ratings and the Box Office: Some Tantalizing Statistics. Film Quarterly, 35(2), 28-30

Teti, E. (2013) The dark side of the movie. The difficult balance between risk and return. Management Decision, 51(4), 730 – 741

Waterman, D, Jayakar, Kp., (2000) The competitive balance of the Italian and American film industries. EUROPEAN JOURNAL OF COMMUNICATION, 15(4), 501-528

Website articles

Busch, A. (2014). Box Office: 'Transformers' Controversy Erupts As Paramount Stakes Claim To $100M Three-Day Gross, No One's Buying It. Retrieved from: https://deadline.com/2014/06/transformers-late-night-estimate-8-75m-on-path-to-mid-90m-opening-797118/

Dahlgreen, W. (2015). Note to producers: the ideal movie length is under 2 hours. Retrieved from: https://yougov.co.uk/topics/lifestyle/articles-reports/2015/09/26/note-producers-perfect-movie-length-under-2-hours

Szalai, G. (2017) Licensed merchandise sales hit $262.9B in 2016, boosted by "Star Wars." The Hollywood Reporter. Retrieved from:
http://www.hollywoodreporter.com/news/licensedmerchandise-sales-hit-2629b-2016-boosted-by-star-wars-1005108

Webley, P., Lea, S., (1997) Topic 3: Path Analysis. In: University of Exeter: Department of Psychology, Available from:
https://people.exeter.ac.uk/SEGLea/multvar2/pathanal.html

Websites

Box Office Mojo (2020), About Box Office Mojo. Boxofficemojo.com. Available from: http://www.boxofficemojo.com/about/?ref=ft

Box Office Mojo (2020), Box Office Tracking by Time. Boxofficemojo.com. Available from: http://www.boxofficemojo.com/about/boxoffice.htm

IMDb.com (2020), IMDb Datasets. Available from: https://www.imdb.com/interfaces/

MPAA.org (2019), 2019 Theme report. Available from: https://www.motionpictures.org/wp-content/uploads/2020/03/MPA-THEME-2019.pdf

Numbers (2017), Domestic movie theatrical summary 1995 to 2017. Retrieved from: http://www.the-numbers.com/market/

Statista.com (2020), Watson, A., Market share of leading film studios in North America from in 2019. Retrieved from https://www.statista.com/statistics/187171/market-share-of-film-studios-in-north-america-2010/

Statista.com (2020), Watson, A., Median hourly wages of selected occupations in U.S. motion picture and video industries as of May 2019. Retrieved from https://www.statista.com/statistics/199976/hourly-wages-of-the-largest-occupations-in-motion-picture-and-video-industries-in-the-us/

The Economist (2016), How to make a hit Hollywood film. Retrieved from https://www.economist.com/graphic-detail/2016/02/26/how-to-make-a-hit-hollywood-film

US Bureau of Labor Statistics, (2017). Consumer Price Index Data from 1913 to 2017. US Inflation Calculator, US Bureau of Labor Statistics. Available from: http://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from1913-to-2008/

Bibliography

Woolridge, J. M. (2012) Introductory econometrics: a modern approach, 5th Edition, South-Western Cengage learning, Michigan State University

List of appendices

Appendix A: Full regression results

log-log OLS regression result

independent v.				dependent variable				
						log(domestic)		
		full sample		2005-2009		2010-2014		2015-2019
Constant	3.43700***	3.94210***	3.88970***	3.83590***	3.02070***	1.2808	3.76290***	5.60420***
	(0.552320)	(0.545640)	(0.572130)	(0.558110)	(0.739860)	(1.363700)	(1.317200)	(1.118200)
log(budget)	0.27389***	0.22248***	0.22105***	0.14762***	0.14379***	0.18458**	0.0975	0.15980***
	(0.038410)	(0.038140)	(0.039060)	(0.041070)	(0.041120)	(0.082590)	(0.074010)	(0.050880)
log(domestic_opening)	0.59205***	0.60535***	0.60876***	0.60593***	0.60642***	0.61403***	0.64548***	0.47293***
	(0.020200)	(0.019640)	(0.019870)	(0.019540)	(0.019490)	(0.035200)	(0.028700)	(0.047550)
action	-0.27044***	-0.22056***	-0.21660***	-0.19140***	-0.19331***	-0.16211*	-0.19325**	-0.26827***
	(0.053630)	(0.053980)	(0.053940)	(0.053720)	(0.053650)	(0.087870)	(0.097600)	(0.087030)
adventure	0.05269	0.04743	0.05554	0.00951	0.01144	-0.03405	0.04788	0.01319
	(0.050420)	(0.049420)	(0.049950)	(0.048980)	(0.048860)	(0.093810)	(0.083000)	(0.080060)
animation	0.27411***	0.27416***	0.26850***	0.39617***	0.41629***	0.21518	0.63320***	0.29850***
	(0.064260)	(0.066430)	(0.066100)	(0.069800)	(0.072410)	(0.149240)	(0.134390)	(0.100590)
biography	0.36488***	0.33690***	0.34677***	0.31230***	0.30206***	-0.18182	0.80345***	0.16845***
	(0.111610)	(0.105220)	(0.106310)	(0.104820)	(0.104940)	(0.169000)	(0.213100)	(0.147050)
comedy	-0.07863	-0.07217	-0.07686	0.04285	0.04505	-0.03591	0.0554	0.1204
	(0.056600)	(0.055590)	(0.055690)	(0.056140)	(0.055940)	(0.111010)	(0.096740)	(0.082020)
crime	-0.20737***	-0.18797***	-0.20439***	-0.19477***	-0.19572***	-0.24218**	-0.17783*	-0.10329
	(0.059470)	(0.059030)	(0.059970)	(0.058800)	(0.058730)	(0.108670)	(0.102370)	(0.091160)
drama	0.11316**	0.08995**	0.09077**	0.01177	0.00786	0.04467	-0.05222	-0.05794
	(0.044950)	(0.044810)	(0.044980)	(0.046170)	(0.046270)	(0.086360)	(0.085990)	(0.072000)
family	-0.25486***	-0.22886***	-0.15445	-0.06098	-0.05179	-0.14424	0.07869	-0.06562
	(0.056070)	(0.059290)	(0.102490)	(0.101000)	(0.101160)	(0.204920)	(0.165700)	(0.168500)
fantasy	-0.06918	-0.07058	-0.06247	-0.0596	-0.06111	-0.05377	0.04757	-0.21636***
	(0.047630)	(0.048420)	(0.048100)	(0.046740)	(0.046710)	(0.086610)	(0.091390)	(0.074800)
history	0.1953	0.22386*	0.21921*	0.16357	0.16806	0.32221	0.09215	0.06211
	(0.134110)	(0.126540)	(0.126930)	(0.124820)	(0.124780)	(0.198270)	(0.244180)	(0.212130)
horror	-0.26861***	-0.23892***	-0.25354***	-0.15558**	-0.14195**	-0.29161**	-0.12433	-0.03055
	(0.062240)	(0.061590)	(0.061920)	(0.062570)	(0.062420)	(0.126420)	(0.100920)	(0.113780)
musical	0.35595***	0.24522*	0.25040*	0.18931	0.17683	-0.06673	0.16449	0.37088**
	(0.130740)	(0.133140)	(0.133200)	(0.130730)	(0.130700)	(0.181010)	(0.306630)	(0.162310)
mystery	0.01289	0.04691	0.04829	0.03188	0.03383	0.0073	0.15796	-0.15475
	(0.060570)	(0.059730)	(0.060010)	(0.058410)	(0.058230)	(0.111650)	(0.103120)	(0.105270)
romance	-0.14334**	-0.13652**	-0.12791**	-0.13994**	-0.14464**	-0.16671	-0.04161	-0.24025**
	(0.056810)	(0.055860)	(0.057500)	(0.056830)	(0.056840)	(0.113300)	(0.089690)	(0.095390)
sci.fi	-0.09425**	-0.07684*	-0.07186*	-0.07932**	-0.08041**	-0.31226***	0.04648	0.02018
	(0.040790)	(0.039460)	(0.039620)	(0.039140)	(0.039150)	(0.074550)	(0.066360)	(0.074400)
thriller	-0.04305	-0.01625	-0.02442	0.02173	0.01858	-0.05629	0.02345	0.06948
	(0.053190)	(0.052070)	(0.052090)	(0.051550)	(0.051590)	(0.103910)	(0.091600)	(0.081540)
war	-0.11072	-0.15365	-0.1781	-0.21076	-0.21009	-0.25996	-0.19447	-0.08092
	(0.135490)	(0.133060)	(0.133680)	(0.132000)	(0.131920)	(0.174400)	(0.253300)	(0.265140)
western	-0.02647	-0.06656	-0.06131	-0.14581	-0.12854	-0.24441	-0.34856*	0.41578
	(0.176930)	(0.173700)	(0.173030)	(0.171440)	(0.170520)	(0.307280)	(0.209810)	(0.396750)
spring1		0.17719***	0.17627***	0.16450***	0.16408***	0.12912	0.20179**	0.15560**
		(0.054140)	(0.054270)	(0.052880)	(0.052880)	(0.111830)	(0.093050)	(0.069920)
summ1		-0.00152	-0.00088	-0.00084	-0.00048	0.06349	-0.04806	-0.0399
		(0.050530)	(0.050820)	(0.050320)	(0.050340)	(0.107010)	(0.080680)	(0.078200)
christm1		0.50318***	0.50493***	0.46992***	0.47478***	0.45714***	0.60533***	0.25122***
		(0.054190)	(0.054160)	(0.052390)	(0.052650)	(0.103800)	(0.095840)	(0.080380)
PG			-0.07176	-0.03773	-0.04193	0.04908	-0.29733*	0.09434
			(0.097960)	(0.094970)	(0.094970)	(0.188380)	(0.151950)	(0.164960)
R			0.06171	0.028	0.02853	0.10396	0.02102	-0.12282*
			(0.048580)	(0.048490)	(0.048480)	(0.093410)	(0.084310)	(0.072000)
G			0.51741	0.32181	0.35464	0.47452	0.14535	0.23412
			(0.664080)	(0.547390)	(0.542620)	(0.486700)	(0.558880)	(0.207610)

NC.17			0.0605	0.15195	**0.16191**	0.29044	NA	NA
			(0.197810)	(0.201870)	**(0.202220)**	(0.202190)	NA	NA
running_time				0.01213***	**0.02692*** **	0.04594***	0.01602	0.01536
				(0.001360)	**(0.008630)**	(0.015450)	(0.013810)	(0.014290)
running_time_sq					**-0.00006***	-0.00015**	-0.00002	-0.000001
					(0.000030)	(0.000060)	(0.000060)	(0.000060)
Observations		2353		782		877		694
R^2	0.725	0.738	0.739	0.748	0.748	0.744	0.785	0.686
Adj. R^2	0.722	0.735	0.736	0.745	0.745	0.734	0.778	0.669

Note: *p<0.1; **p<0.05; ***p<0.01

Appendix B: VIF method

Variables	(1)	(2)	(3)	(4)	(5)
log(budget)	2.253717	2.360224	2.448722	2.75401	2.780159
log(domestic_opening)	1.948451	1.977367	2.022788	2.024172	2.025501
action	1.825697	1.840968	1.847824	1.852547	1.853402
adventure	1.84748	1.848412	1.86305	1.877222	1.878008
animation	1.898217	1.900894	1.957984	2.001833	2.036172
biography	1.357599	1.360106	1.368015	1.371272	1.38032
comedy	1.727758	1.728537	1.740351	1.860135	1.86141
crime	1.303873	1.305979	1.342597	1.343034	1.343167
drama	1.763331	1.76831	1.774405	1.829886	1.83416
family	2.26436	2.269914	4.405328	4.445772	4.458099
fantasy	1.41295	1.415815	1.426794	1.426836	1.427204
history	1.304601	1.305733	1.323092	1.328589	1.329721
horror	1.526214	1.52857	1.557127	1.59044	1.610797
musical	1.081498	1.092033	1.100971	1.105066	1.110449
mystery	1.284269	1.287679	1.291718	1.292728	1.293176
romance	1.279345	1.280477	1.311121	1.311956	1.315963
scifi	1.359716	1.36251	1.367543	1.367797	1.367969
thriller	1.747802	1.759275	1.77162	1.787604	1.789951
war	1.160245	1.1638	1.180952	1.182539	1.18256
western	1.027426	1.029001	1.033446	1.037794	1.043519
summ		1.176661	1.179216	1.179216	1.179237
spring		1.159346	1.159709	1.160268	1.160289
christm		1.299355	1.300284	1.308437	1.313405
pg			3.600703	3.606672	3.60953
r			1.448957	1.458586	1.458659
g			1.307257	1.310383	1.311554
nc17			1.038796	1.041105	1.043156
running_time				1.80149	86.15165
running_time_sq					82.18662