STOCKHOLM SCHOOL OF ECONOMICS
Department of Economics
659 Degree project in economics
Spring 2022

Lost in Translation

A study exploring how the origin of the story affects a movie's financial success

Theo Renaudin (24589)

Abstract

The purpose of this study is to investigate how the Return-on-Investment of a movie is impacted by the origin of the movie's story, for instance an existing fictional book, a real-life event or an original script. The objective is to better understand the potential financial success of a movie at the very early pre-production stage, minimising investment risks and optimising portfolio strategy. The scope of the study is all French movies released between 2017 and 2019. Following the work of preceding research, a multivariate regression is conducted including the most prominent variables which have been studied during the past decades, with the addition of different dummy variables for "origin of the story". The first contribution of this study is that its findings are consistent with previous studies focusing on different regions. The second contribution is that the study shows that the origin of the story of a movie is correlated with different variables. The third contribution is the finding that the origin of the story is correlated with profitability.

Keywords: Movie production, origin of story, ROI, prediction, decision-making

JEL: Z110, M110

Supervisor: Sampreet Goraya
Date submitted: 10 May, 2022
Date examined: 25 May, 2022
Discussants: Ida Lennström, Ida Nordenadler
Examiner: Johanna Wallenius

Acknowledgments

Thank you to my closest friends for motivating me when my morale has been low. Thank you to my supervisor Sampreet for always telling me what I need to hear. Thank you to my mother for covering walls with calendars and post-its. Thank you to my father for his impressive knowledge of the field, words cannot describe how helpful he’s been.

Table of content

Definitions

ROI	Return on investment. Calculated with the following formula: $ROI = \frac{revenue - production\ costs}{production\ costs}$
BUDGET	A variable in the final regression, denoting the production cost of the movie
AWARDS	A variable in the final regression, denoting the number of César nominations and Oscar nominations the movie has received
CRITIC	A variable in the final regression, denoting the professional critic review the movie has gotten on the website Allociné
MAJOR	A variable in the final regression, denoting whether or not a movie was distributed by one of the major distributors in France
HOLIDAY	A variable in the final regression, denoting whether or not the movie was released in holiday season
DRAMA	A variable in the final regression, denoting whether or not the movie is of the drama genre
ACTION	A variable in the final regression, denoting whether or not the movie is of the action genre
SEQUEL	A variable in the final regression, denoting whether or not the movie is a sequel to a previously released movie
BOOK	A variable in the final regression, denoting whether or not the movie is based on a book
COMIC	A variable in the final regression, denoting whether or not the movie is based on a comic book
MOVIE	A variable in the final regression, denoting whether or not the movie is based on a movie, for example a remake
PLAY	A variable in the final regression, denoting whether or not the movie is based on a play
TRUE	A variable in the final regression, denoting whether or not the movie is based on a true story
ORIGINAL	A variable in the the preliminary regression, denoting whether or not the movie has an original story

I. Introduction

The movie industry is an essential provider of worldwide entertainment and has seen constant growth since its birth more than 100 years ago. In 2019, the theatrical box-office had reached $42,2 billion globally. The Covid-19 pandemic significantly impacted the industry and reduced global box-office to $11,8 billion in 2020 and $21,3 billion in 2021 (Motion Picture Association (MPA), 2019-2021). Despite the competition of online video subscriptions, theatrical box-office is expected to be back to pre-pandemic level and growth rates (Gower Street Analytics; 2022).

As the movie industry has matured, it has become increasingly important for investors and producers to understand as early as possible which movies will succeed and which movies will not. Driven by industry demand, the research field of predicting movie success has made great advancements in the past decades. However, with the existing knowledge, no producer has the ability to know that investing in a certain movie will generate a positive return on investment (ROI). This is even more true at the planning stages of the movie, where very little is known about the movie and the main investment decision has to be taken. More knowledge about the factors that affect the revenue of a movie is necessary in order to achieve the goal of confidently predicting if a movie will be successful early in the decision-making process.

In recent years, an increasing number of movies have been based on books, comics and true stories (The Numbers, 2021, https://www.the-numbers.com/market/sources). In other words, movies that are not based on original screenplays but on stories that have already been written or told in other contexts. There are reasons to believe that a movie will be more successful if its story is based on an already existing story. For instance, a movie based on a popular book is likely to become successful as its story has already been read and liked by many readers (Simonton, 2009). However, the origin of the story and how it affects movie success has been studied very little in the literature. This creates a gap that this study intends to address.

Therefore, this study is conducted to understand if and how much the origin of the story of a movie affects its success. The purpose is to gain further knowledge of what makes a movie financially successful based on data that is available early in the decision-making process. We thus ask the following question:

Does the origin of the story have a considerable impact on the success of a movie?

Here, the origin of the story refers to where the storyline of the movie comes from, whether it is based on a fictional book, a previous movie, a comic, a true story, a play or if it is an original screenplay. By ‘considerable impact’, it is meant that the effect of the storyline is large enough to be worth considering when making the decision of investing in a movie. Finally, the success of a movie can be

described in several ways (financial performance, awards, critical reception). In this study the success of a movie is seen from a financial perspective, thus analysing the movies' return on investment (ROI).

In order to answer the research question, the study is conducted as a stepwise regression analysis based on a previous study by Pangarker and Smit (2013). By using the same control variables, which are some of the most prominent variables in the literature for movie predictions, and by adding variables for the origin of the story, the author will be able to analyse the regression to understand if and how much the origin of the story affects the success of a movie.

The study contributes to the literature in 3 ways: First of all, its findings are consistent with previous studies focusing on different regions. Indeed, Pangarker and Smit show that Major distributor, Award nomination and Sequel all have a positive impact on revenue in the US. According to the results of this study, the same variables have a positive effect on ROI in France. The second contribution is that study shows that the origin of the story of a movie is correlated with different variables. For example, movies based on books are positively correlated with being nominated for awards, while movies based on comic books are positively correlated with higher budgets. The third contribution is the finding that the origin of the story is correlated with profitability, however the correlation changes depending on what origin it is. Movies based on books and comic books both have strong negative correlations with profitability for example. However, these results aren't significant above the 0,05 level, and thus aren't conclusive.

The data consists of information on all the movies released in France between 2017 and 2019. Information such as production costs, date of release, genre, critics review and more has been collected through different online databases taken from CBO box-office, JP's Box-Office as well as Allociné. The data on number of award nominations, on whether the movie is a sequel or not, and on the origin of story have all been collected manually by a systematic review of each movie's Wikipedia page.

The remainder of the paper is structured as follows: The background and literature review (section 2) provides an overview of the movie industry and presents the relevant research in the literature together with this paper's contribution. The theoretical framework (section 3) defines how the success of a movie is measured and why the origin of the story could matter for this success. The data part (section 4) provides a description of the scope of the data, how it was collected and how it was prepared for the regression. The method (section 5) provides information on how the multivariate regression was conducted. The section 6 presents the results of the analysis and the section 7 discusses what can be interpreted from those results.

II. Background and Literature Review

The movie industry

With regards to the number of movies produced and the box-office revenue, the movie industry is dominated by the United-States. In 2019, the last year before the Covid-19 pandemic, the box-office revenue from movies produced in the United-States represented 71,5% of the worldwide box-office revenue and 30,6% of the total number of movies produced (see table 1).

Table 1: Number of Movies, Total Worldwide box-office and Share of Worldwide box-office for 2019 - 10 largest countries in box-office

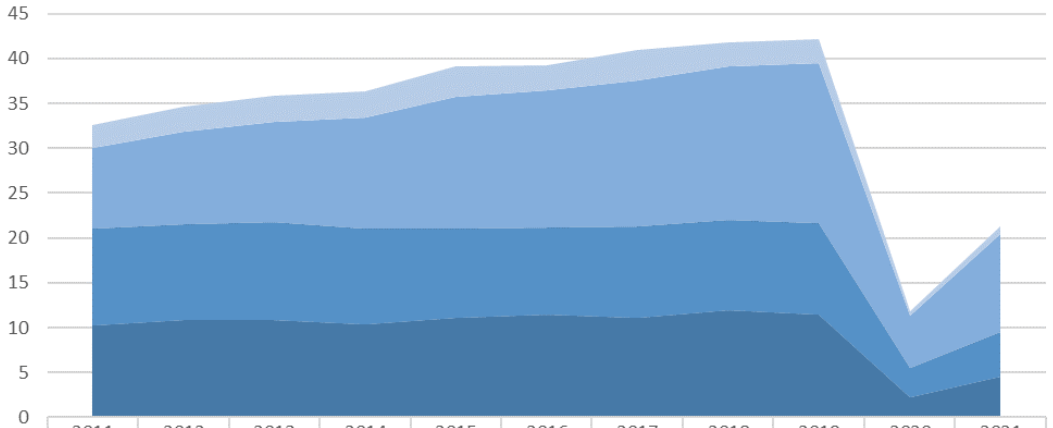
Production Country	Number of Movies 2019	Total Worldwide box-office 2019	Share 2019
United States	1,638	$27,747,447,471	71.460%
China	278	$6,106,681,794	15.727%
Japan	141	$2,258,425,823	5.816%
United Kingdom	342	$2,048,057,725	5.275%
India	270	$1,202,698,675	3.097%
Republic of Korea	153	$1,123,042,747	2.892%
France	271	$757,743,082	1.951%
Hong Kong	33	$748,082,214	1.927%
Canada	169	$601,264,666	1.548%
Australia	87	$373,555,625	0.962%

Note: Source: https://www.the-numbers.com/movies/country-breakdown/2019

The United States dominates the world film industry and most research studies have used US domestic data rather than data from other countries (Simonton, 2009). Only occasionally will a study examine box-office revenue or production cost for another specific country (e.g., Hand, 2001; Lee, Kyung Jae, Chan, 2009; Ruus, Sharma, 2019) or include world-wide performance statistics (e.g., Litman & Ahn, 1998).

Even if movies produced in the United-States represent more than 75% of the world box-office revenue, the worldwide box-office revenue is distributed more equally between EMEA, Asia-Pacific and North America. During the period 2013-2019, box-office revenue increased mostly in Asia Pacific (see graph 1).

Figure 1: Yearly box-office revenue in billions USD by main region - 2011/2021

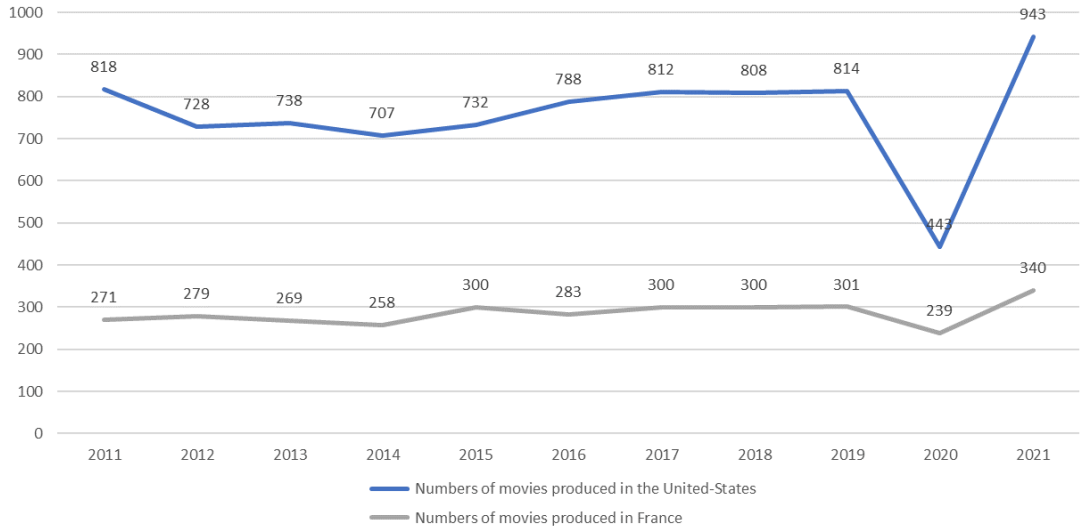


	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Latin America	2,6	2,8	3	3	3,4	2,8	3,4	2,7	2,7	0,5	0,9
Asia Pacific	9	10,4	11,1	12,4	14,6	15,3	16,2	17,1	17,8	5,8	10,9
EMEA	10,8	10,7	10,9	10,6	10	9,8	10,2	10,1	10,3	3,3	5
US/Canada	10,2	10,8	10,9	10,4	11,1	11,4	11,1	11,9	11,4	2,2	4,5

Note: Source: Motion Picture Association https://www.motionpictures.org/

Looking at North America and EMEA, the box-office revenue has been stable between 2011 and 2019. This is also the case in number of movies produced as shown in by graph 2 for France and the United-States.

Figure 2: Number of Movies Produced for Future Theatrical Release by Year - 2011/2021 - France and United-State



Note: Source: Motion Picture Association https://www.motionpictures.org/, CNC https://www.cnc.fr

Predicting a movie's financial success

In 1983, William Goldman, Oscar-winning writer of screenplays for "Butch Cassidy and the Sundance Kid" and "All the President's Men" wrote in his book *Adventure in the Screen Trades* that

“Nobody knows anything… not one person in the entire motion picture field knows for a certainty what’s going to work”. 1983 was also the year when Litman developed the first multiple regression model attempting to predict the financial success of a movie. In this ground-breaking study, he provided evidence that the independent variables for production costs, critics’ ratings, science fiction genre, major distributor, Christmas release, Academy Award nomination, and winning an Academy Award were all significant determinants of the revenue of a movie.

The importance and potential benefits of predicting the success of a movie has grown steadily during the past 50 years (McKenzie, 2009; Ruus & Sharma, 2020) in proportion with the increase of the film industry’s revenue and financial stakes (Vogel; 2014). The amount of research on the subject of predictability of a movie’s success has therefore also seen the same growing evolution. Several reviews of previous research have been published with the goal of gathering previous results and conclusions (Eliashberg & Leenders, 2006; Simonton, 2009: Eliashberg, Weinberg, & Hui, 2008; McKenzie, 2012).

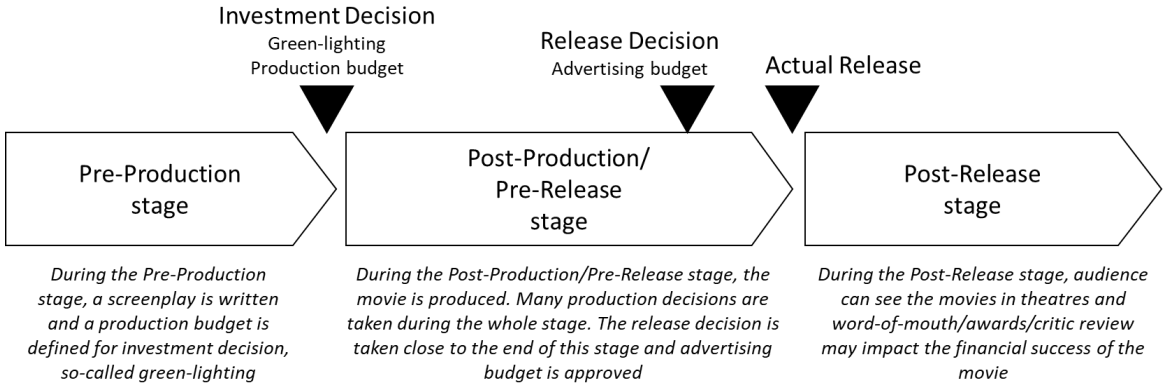
Since Litman’s first publication in 1983, three main types of forecasting models have emerged has most able to predict a movie’s success (Sharda & Delen, 2006; Ghiassi et al., 2015):

A. Econometric models exploring predictive value of different variables for box-office revenue. These models are mostly based on multivariate regression analyses (Litman 1983; Litman and Kohl 1989; Litman and Ahn 1998; Ravid 1999; De Vany and Walls; 1999; Simonoff and Sparrow 2000; Walls, 2005; Pangarker & Smit, 2013). The current study will be using this type of model as it is based on the Pangarker & Smit (2013) study.
B. Behavioural models focusing on the individual’s decision-making process when choosing a movie among all other entertainment alternatives (Anast, 1967; Sawhney & Eliashberg, 1996; De Silva, 1998; Eliashberg et al., 2000).
C. Artificial Neural Networks, already used as an effective forecasting method in retail sales (Alon, Qi, & Sadowski, 2001) and taking advantage of the great development in computational power and increased access to movie-related data and social-media data (Sharda and Delen (2006); Eliashberg, Hui, and Zhang (2007); Ghiassi, et al. (2015)). Several studies using machine learning techniques have produced prediction models with increasing levels of accuracy (Sharda and Delen (2006); Eliashberg, Hui, and Zhang (2007); Ghiassi et al. (2015)).

As presented by Ghiassi et al. (2015), the time frame of the decision-making process is a key dimension for the predictive modelling of a movie’s success, as it dictates the availability of data. They define three main stages where important business decisions have to be made and where

available data differ noticeably: (i) pre-production (i.e., before the decision to produce the movie has been taken, called green-lighting), (ii) post-production but pre-release (i.e. after the movie has been produced but before it is released to the main public) and (iii) post-release (i.e. after the movie has been released in theatres).

Figure 3: Description of the different stages of the decision making-process



Note: Source: made by author

Forecasting models falling into the category (iii) of 'post release' model show more accurate forecasting results as there are more explanatory variables available, including critic reviews, first-week sales and word-of-mouth effects. Several studies have shown a high level of accuracy for the total box-office revenue after the first week of box office sales are determined. Using a behavioural model, Sawhney & Eliashberg (1996) found that "cumulative box-office revenues can be predicted with reasonable accuracy (often within 10% of the actual) using as little as two or three data points [i.e. weeks of sales]" but that accuracy felt considerably when the first weeks of sales were not available. Similarly, several studies have shown correlations between consumer-generated word-of-mouth volume (but less so for word-of-mouth valence, i.e. positive or negative sentiment) and box-office revenues (Duan et al., 2008; Liu et al., 2010, Kim et al., 2017).

However, research has shown that predicting an accurate estimate of a movie's box-office revenue at stage (i) or (ii) , i.e. before its release, is much more difficult to obtain. Following Litman's initial work in 1983, several subsequent studies were conducted to include more data points (Litman and Kohl; 1989; Litman & Ahn, 1998; Ravid 1999). However, in 1999, De Vany and Walls reached the conclusion that it was impossible to attribute the success of a movie to individual causal factors. They showed that movie revenues were Levy-distributed with extreme skew and argued that box-office revenue outcomes diverged over all value. They concluded that studios should drive a strategy of portfolio of films to minimise risk rather than selecting individual film projects, a very relevant

conclusion for the early decision stages but still yet to be researched on (Sacco & Teti, 2020). Richard Caves (2000) expressed these difficulties through the nobody knows principle: “That is, producers and executives know a great deal about what has succeeded commercially in the past and constantly seek to extrapolate that knowledge to new projects. But their ability to predict at an early stage the commercial success of a new film project is almost nonexistent”. Since then, focus has been put on identifying the most prevalent variables for pre-release prediction. In 2005, Walls demonstrated that even though “nobody knows anything” when predicting the financial success of a movie, much is known about the attributes of movies that have been financially successful and that the systematic component of box-office revenue can be quantified by mean of these most prevalent variables.

Most prevalent variables for pre-release/post-production success prediction

According to Pangarker and Smit (2013), the following pre-release variables have been shown by the major preceding studies as having the largest impact on box-office success:

1) Film genres
2) MPAA ratings
3) Production costs
4) Major studio involvement in movie’s release
5) Academy Award nominations or awards
6) Timing of release
7) Critics’ reviews before release

The movie’s cinematographic genre is shown as having an influence on box office revenue by several studies (Anast, 1967; Simonoff and Sparrow, 2000; Vany and Walls, 2002; W. David Walls, 2005) but these studies reached different conclusions: according to Anast (1967), action-adventure has a negative impact on revenue and eroticism a positive one, while Litman (1983) finds that only the science-fiction genre has a significant impact, both studies performed for US movies. Neelamegham and Chinatagunta (1999) find that across countries the thriller genre is the most popular, while romance genre was the least popular. Eliashberg, Hui, and Zhang (2014) shows that romance and thriller are among the five most important features of their box-office performance model. Lash and Zhao (2016), one of the few studies looking at ROI instead of Revenue, showed that Drama had a negative effect on a movie’s ROI. It is important to note that these studies used sample data over various time periods and various geographies and that this can imply a changing relationship between genre and revenue, as audience tastes would have changed over the years and the geographies.

The movie’s MPAA rating G (General Audiences), PG (Parental Guidance Suggested), PG-13 (Parents Strongly Cautioned), R (Restricted), NC-17 (Adults Only)) or other content rating are also identified as having an impact on box-office revenue. Even if Litman (1983), Austin (1984) and Austin & Gordon (1987) find that movie ratings are not a significant predictor of financial success,

Ravid (1999) provides evidence using a linear regression model that G and PG rated films have a positive impact on the financial success of a film. De Vany & Walls (2002) and Walls (2005) show that PG-13 and R ratings show lower returns than G rated movies. In the same manner, Sochay (1994) and Sawhney & Eliashberg (1996) conclude that R-rated movies show lesser box office revenue than the other ratings.

The movie's production costs are strongly correlated with its box office revenue according to several studies (Prag & Casavant, 1994; Terry et al., 2005; Simonton, 2005: Deniz & Hasbrouck, 2012), even if Litman (1983) and De Vany & Walls (1999) argue that high productions costs may not directly lead to bigger profit and can also mean excessive salaries, production delays or inefficient management, inflating the total movie cost with no effect on the quality of the result. Ghiassi et al. (2015) observed that the accuracy of their forecasting model increased after having introduced production costs as a predictive variable. It is important to note that production costs have only recently been made available (Simonton 2009) and that the available production costs does not include marketing and advertising expenses and that these expenses are not communicated by movie. According to Prag & Casavant (1994) and Hennig-Thurau et al. (2007), this is acceptable as these expenses are strongly correlated to production costs.

The market power of the studio involved in the movie's release: Litman (1983) showed that movies released by a major studio company performed better at the box office than movies released by smaller studios or independent distributors.

Nominations or awards received by the actors, the director or the movie have been shown as having an impact on the movie's success, and have even quantified this impact. Nelson, Donahue, Waldman and Wheaton (2001) estimated that being nominated to an Academy Award would increase the movie's box-office revenue by $4.8 million and winning an Academy Award by $12 million. Similarly Litman (1983) estimated a best actor or best picture nomination to an $7.3 million increase in box-office revenue and that a major category win would mean a $16 million increase whereas Dodds and Holbrook (1988) estimated a nomination for best actor to $6.5 million, best actress to $7 million and best picture to $7.9 million. Finally, Deuchert, Adjamah and Pauly (2005) showed that while the awards have a positive effect, the main effect is through nominations.

The date, timing, or season of the film's release have been shown as having a positive correlation with revenue by Litman (1983) for the Christman period and by Sochay (1994) for the summer months (in the northern hemisphere). Chang and Ki (2005) showed that release periods (Summer and Easter) significantly related to total box office performance. Interestingly, Terry et al., 2005 did not find any significant correlation between a holiday release and a movie's revenue and explained it by distributors releasing their movies several weeks before the holidays. Ghiassi et al., (2015) observed

that the accuracy of their forecasting model increased after having introduced seasonality as a predictive variable. Einav (2007) showed that the positive correlation between season and box-office revenue is also explained by the fact that studios choose the peak period to release their better quality and higher production costs movies. Noticable is that this variable will highly change between countries and continent (Einav, 2007; Eliashberg & Elberse, 2003).

The quantity and/or quality of reviews by film critics were studied by Eliashberg and Shugan (1997) and Ravid, Basuroy & Chatterjee (2003) and both research identified the role of a critic as having an impact on box-office revenue, even if Ravid, Basuroy & Chatterjee (2003) concluded that negative reviews hurt performance more than positive reviews help performance. Derrick, Williams, and Scott (2014) found that only quantity of pre-release movie critics was correlated with revenue (and not valence/quality of the reviews). According to their study, only some weeks after release (post-release) would the valence (positive critics) of the professional reviews correlate with box-office revenue. Reinstein and Snyder (2000) demonstrated that only a few critics had the power to influence consumer demand and thereby box-office revenue.

The review of the most prevalent pre-release variables as identified by Pangarker and Smit (2013) shows that conclusions are not always consistent between studies. This can be partly explained by the very heterogeneous set of data used by the research in this field, both in terms of data set and in terms of set of rules to define each variable. The majority of the above research has been focused on the pre-release/post-production stage of the decision process. However, there is an even greater value in being able to predict success at the pre-production stage, before the investment decision is taken.

Predicting success at the pre-production stage - The green-lighting decision

In the film industry, green-lighting is the key moment when the critical decision of producing and financing a movie is formally taken. This decision is most usually based on a screenplay, also called a script, which is the blueprint of the movie, expressing the movements, the actions and the dialogues occurring throughout the story. As it is the main piece of information decision makers have at their disposal pre-production and even if the green-lighting of a movie leads to large amounts of money being invested, this decision is therefore “largely a guesswork based on experts’ experience and intuitions” (Eliashberg, J., Hui, S. K., & Zhang, Z. (2007)), mainly based on the reading of the screenplay.

According to Ghiassi et al. (2015): “To date, literature has not offered a viable method which successfully predicts these patterns without utilising post-release or post-production data.” Predicting box-office revenue only based on the pre-production information remains a critical issue for the film industry (Sharda & Delen, 2006).

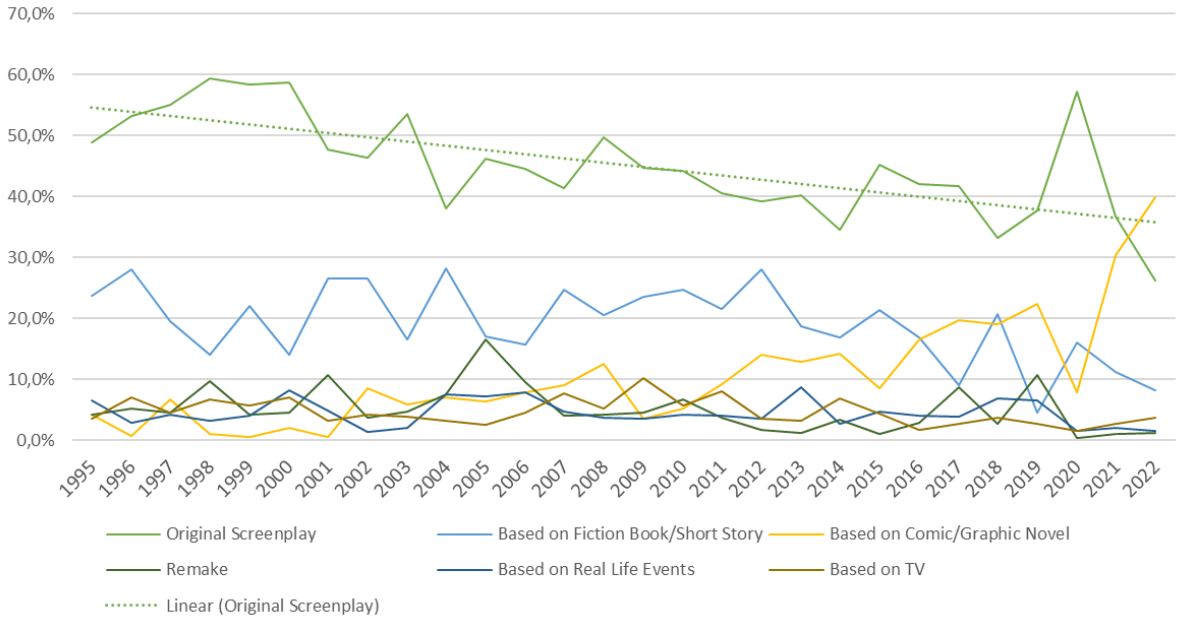
Why this decision point is so important has been described by Caves (2001). Caves explains that information on a movie's quality and chance of success is revealed gradually during the entire production time. This means that costs are sunk progressively and that decision makers can carry movies to completion without ever questioning the green-lighting. This escalation of commitment despite negative outcome has been shown in the psychology field (Juliusson et al., 2003; Karlsson et al., 2005) While increasing the success rate of the green-lighting decision is very complex, little science usually goes into the process (Eliashberg & Leenders, 2006). Even a marginal increase in this success rate would bring high financial value to the decision makers as investment and costs keep increasing in modern movies (Eliashberg et al., 2007; Hunter, Smith, & Singh, 2015).

In addition to the previous discussed pre-production variables "Film genres", "MPAA ratings", "Production costs" and "Studio market power", Simonton (2008) identifies the following features of a screenplay as possible predictors for the success of a movie at the "green-lighting" stage: (i) Sequels or prequels, i.e. a script taking place before or after an already existing movie (ii) Remakes, i.e. new versions of earlier movies (iii) Adaptations from comic books, plays, novels and other medias (iv) True Stories and (v) Original story. These five predictors can be grouped under a common variable "the origin of the story". If one would like to be able to predict the success of a movie at the green-lighting stage, the origin of the story would be one of the most important pre-production variables to take into account.

As Barry Gunter put it in his book *Predicting Movie Success at the Box Office* (2018): "The story is in fact the critical variable that tends to determine whether a movie idea gets the green light and enters production at all. Movie studio executives make presumptions about the potential of a story to deliver a successful movie. Often the decision-making at this stage is guided by intuition and ad hoc experience rather than on systematic scientific enquiry. Producers can be attracted by stories that are already well known because they have been told as novels or as real-life events".

What Barry Gunter points out as an attraction by the producer for already known stories is confirmed by the graph 3 below showing a trend since 2000 with less market share of Original Screenplays. However, this trend may well still largely be driven by "a guesswork based on experts' experience and intuitions" (Eliashberg, J., Hui, S. K., & Zhang, Z. (2007)).

Figure 4: Number of Movies made per "Origin of the Story" (6 largest), for North America and period 1995-2022



Note: Source: https://www.the-numbers.com/market/sources

Table 2 shows box office and number of movies per "origin of the story" for North America and for the period 1995-2022. As seen in this table, the box-office per movie is very different depending on the origin of the story which may show that the origin of the story matters. However, this relation may only be apparent.

Table 2: Per "Origin of the Story" (10 largest), for North America and period 1095-2022: Number of Movies, Total box-office, Box office per movie, Share of box-office

Origin of the Story	Movies release in North America 1995-2002	Total Box Office North America 1995-2022	Box Office per movie North America 1995-2022	Share Box Office North America 1995-2022
Original Screenplay	8,057	$107,143,227,619	$13 298 154	44.53%
Based on Fiction Book/Short Story	2,171	$47,172,756,058	$21 728 584	19.61%
Based on Comic/Graphic Novel	259	$24,693,320,169	$95 341 005	10.26%
Remake	333	$12,883,843,587	$38 690 221	5.35%
Based on Real Life Events	3,257	$11,440,858,176	$3 512 698	4.75%
Based on TV	232	$11,365,324,528	$48 988 468	4.72%
Based on Factual Book/Article	306	$7,535,891,206	$24 627 095	3.13%
Spin-Off	43	$3,905,808,380	$90 832 753	1.62%
Based on Game	55	$2,196,458,088	$39 935 602	0.91%
Based on Play	274	$2,112,882,468	$7 711 250	0.88%

Note: Source: https://www.the-numbers.com

Research focusing on predicting success at the green-lighting/pre-production stage is scarce (Joshi & Mao, 2010; Gunter, 2018) and can be grouped in two kind of study: (a) identifying how explanatory variables known at this stage impact a movie's success and (b) using text mining to "take apart" the screenplay and find key concepts or variables that can be related to movie success.

As seen earlier in this review, genre and rating have been part of research through regression analyses. Taking a broader approach, Michael T. Lash & Kang Zhao (2016) proposed a decision support system to aid movie investment decisions at the early stage of movie production, albeit not specifically at pre-production. They integrated in their model a "what" feature to gather "pre-production" variable and that covered genre, rating, a selection of 30 topics related to the movie's screenplay and finally whether the movie's screenplay was adapted from a comic, a true story, a book or a novel (the origin of the story). Their results showed that the top negative explanatory coefficients for success were all of the 'What' features, including genre (drama and foreign), 'R' rating, and plot topics related to wars and music.

The impact of being a sequel for a movie's success has also been included in several research but with different conclusions. While Prag & Casavant (1994), Ravid (1999) and Simonov & Sparrow (2000) showed a positive impact on total box-office revenue, other showed that the positive effect mostly applies to the opening weeks of the theatrical run (De Vany & Walls, 1999; Litman & Kohl, 1989; Simonton, 2005). Sood and Drèze (2004) showed that brand-extension from the first movie to the

sequel could negatively impact movie performance due to satiation and decrease consumer enjoyment.

Joshi & Mao (2010) focused specifically their study on movies based on best-selling books, asserting that "while book adaptation is an often used strategy in the motion picture industry, it has received little academic attention". Based on brand extension theory they showed that being based on a best-selling book impacts movies' opening box-office weekend but that this impact fades over time while stronger movie-related predictors such as reviews come into play.

Movies scripts have been analysed through textual analysis and text-mining data-processing to support green-lightning decisions. Eliashberg, Hui and Zang (2007) combined natural-language processing techniques, human input and statistical learning methods to forecast a movie's return on investment based only on movie scripts. Eliashberg, Hui and Zang (2014) improved their previous and applied a kernel-based approach to assess box-office performance, showing an improvement in accuracy of box-office revenue prediction compared to their previous works and finding that genre and content variables (such as "Early Exposition" or "Strong Nemesis") were the strongest predictors of box office revenues while text-level and semantic variables were less so. Hunter, Smith & Singh (2016) followed up in the previous research and created a pre-production model where the main factors were derived from the textual and content analysis of the screenplays of these films, determined through the application of network text analysis. Their results confirmed the results from Eliashberg, Hui and Zang (2014).

Research Gap and Contribution

As per the literature review, research to understand how to predict a movie's success has been growing both in breadth and in complexity during the last decades. It has come a long way from 1978 when the president of MPAA said "no one, absolutely no one, can tell you what a movie is going to do in the marketplace" to today's advanced artificial neural network able to predict movie revenues with increasing levels of accuracy.

However, most of the research focuses on variables available on the later stage of the movie's lifecycle, either post-production or post-release. There is an obvious gap in the current research in understanding how the known variables at pre-production will affect the success of a movie. It is one of the intent of this study to analyse one of these variables, the "origin of the story".

Even if post-production and post-release variables are better predictors of a movie success, a better understanding of the impact of pre-production variables would be of great value in a portfolio and risk diversification approach at green-lighting. Understanding better how a portfolio strategy at pre-production stage could be put in place by better understanding the impact of known

pre-production variables is another intent of this study. In the long run there are critical issues to be addressed at the pre-production stage in balancing content innovation and original creation with “safe bets”.

Current literature is mostly based on US domestic data as the US movie industry is the predominant one and only few studies have focused on other markets (India, Korea, China). This study will focus on France to give a different perspective and to avoid a “blockbuster” bias when focusing only on the US domestic market. France is among the five largest country when looking at number of movies produced and French audience may also have a different perception of creative content.

Finally, this study will focus on Return-of-investment (ROI) and not on box-office revenue to value the financial performance of a movie. Using the ROI makes more financial sense than using revenue as ROI assesses the profitability of a movie. The lack of focus on ROI by previous research came from the lack of access to reliable production cost data, which is not the case anymore today.

III. Theoretical framework

Defining the success of a movie

As described by Simonton (2009), the success of a movie can be described by the following triad: (1) Critical Evaluations, (2) Financial Performance and (3) Movie Awards. However, as seen in the literature review, financial performance of a movie has been shown as dependent on critical evaluations and movie awards. Therefore this study will focus on the Financial Performance of a movie, and use Critical Evaluations and Movie Award as explanatory variables. In this study, financial performance will be the measure of a movie's success.

Financial performance can be measured in several ways. The most common ways used in previous research has been total box-office revenue or even first week's box-office revenue. However, recent research has been focusing more on profitability as a better measure of a movie's financial performance.. Therefore the author of this study chose to consider the financial performance of a movie in terms of return on investment (ROI). Using ROI, a movie with high box-office revenues but even higher production costs will not be considered financially successful. It has been shown that revenue and ROI are not correlated in the movie industry (Lash & Zhao, 2016). This metric is most likely the most relevant from the perspective of investors and producers, even more so as the notion of portfolio risk management has started to emerge.

The following formula will be used to define ROI:

$$ROI = \frac{revenue - production\ costs}{production\ costs}$$

Two important assumptions have been applied to this formula:

- In this study, revenue will be limited to theatrical revenue. Revenue from other streams like video-on-demand and streaming platforms based on subscription will not be included. This is due to the fact that revenue for digital channels are not reported by movie and follows different business models, from a movie being sold to a streaming platform to a studio driving its own premium subscription service. The period chosen for this study (2017-2019) being pre-pandemic and focusing on French production with less digital reach, the lack of this share of revenue in the ROI calculation is considered acceptable.

- Concerning production costs, the values gathered for this study do not include advertising expenses, as this data is protected information and sources are often proprietary. However, several studies have shown that advertising expenditures do not offer additional predictive value as these costs are strongly correlated to production cost (Elberse & Anand, 2007;

Ghiassi et al., 2015). Therefore this simplification of not including the advertising expenditures in the production costs is considered acceptable.

Why the origin of the story could impact a movie's success

As De Vany and Walls (1999) wrote, "the audience makes a movie a hit" and therefore the success of a movie all comes down to how an audience chooses which movie to see. The question is therefore if the origin of the story can have an impact on how the public decides on the next movie they will watch in a theatre.

Several factors can influence the viewer's decision-making. Past experience (Juliusson, Karlsson, & Gärling, 2005) and cognitive biases (Stanovich & West, 2008) are among the most important. When something positive results from a decision, people are more likely to decide in a similar way. Cognitive biases such as belief bias implies the over dependence on prior knowledge in arriving at decisions. These factors influencing the decision-making of the audience would therefore tend to imply that people who have enjoyed reading a novel or watching a play in the past will more likely choose to watch a movie telling the same story. As Joshi & Mao (2010) described concerning books as the origin of the story, "the retrieved positive attitude toward the book will provide a favourable context for judgement, increasing consumers' expectation of the movie's performance and in turn their intention to watch the film".

Feldman and Lynch (1988) proposed a decision-making framework that also supports why audience already knowing the story of a new movie could influence their decision to watch the movie. Their study showed that any information will more likely be used by consumers in their decision-making if it is perceived as relevant to the decision and accessible at time of the decision and that other information is less accessible. This would imply that already knowing the story could influence the audience to select the movie, although mostly during the early weeks and before more information (such as critical review or word-of-mouth) is available.

Apart from the psychological dimension on an individual level, other aspects of the origin of the story may impact a movie's success:

- An important community of fans already existing before the movie is launched would also potentially induce higher box-office revenue. However, it is a risk as well if the fans are disappointed by the movie and give negative reviews after having seen it.
- The fact that a story has already been validated by the public and that the public already has shown a large interest in it, being a fictional or a true story, suggests that the movie should be of interest for the public as well.

- True stories may have a stronger dramatic appeal than fictional stories as they are based on real life events and have actually happened. This would potentially create a stronger emotional connection between the audience and the story.

IV. Data

Scope of data

The study is based on the previous research done by Pangarker and Smit (2013) and therefore the scope of data in terms of selected variables is very similar. In addition to the variables used by Pangarker and Smit, this study adds the variables that describe the origin of the story for each movie included in the dataset.

However, one of the objectives of the study is to analyse movies that are not produced in the US in order to get a different perspective compared to the majority of the research on the subject. France was therefore chosen as the scope of this study because it is among the five largest countries in terms of movie production and several reliable movie databases exist for the French movie industry. The movies selected for analysis in the dataset will therefore have been produced in France, or primarily produced in France in case of international collaborations.

The chosen period for the scope of this study is 2017-2019, meaning that movies in the dataset have been released in France between the 1st of January 2017 and the 31st of December 2019. Due to the significant impact of the Covid-19 pandemic on the movie industry, it was decided not to include movies released later than 2019 as they would have been impacted both on the characteristics and financial performance. The choice to select three consecutive years was made in order to keep a certain consistency in the taste and trends of the movie public as the focus will be on the audience's preferences in terms of story.

As the financial performance of the movies is measured with ROI, the movies lacking information about revenue or production cost have been removed from the scope of analysis. From an initial list of 825 movies, the final dataset contains 272 movies with complete data.

Data collection

The main database sources for collecting the data were the following:

- **CBO Box-office** (https://www.cbo-boxoffice.com/ - membership required): Title, Release Date, Distribution Company, Production Country, Production Costs (EUR or USD).
- **JP's Box-Office** (http://jpbox-office.com/): Title, Release Date, Production Costs (EUR), Revenue Worldwide (USD), Genre.
- **Allociné** (https://www.allocine.fr/): Title, Release Date, Genre, Production Country, Review from Professional Critics.
- **Wikipedia** (https://sv.wikipedia.org/): Awards nominations, Origin of the Story, Sequel.

The data was collected with a combination of automated and manual gathering. The Origin of the Story was meticulously and systematically controlled.

The Title and Release Date have been collected from each database as they form the "key" for identifying the correct movie in each database (several movies can have the same title but have been released at different times).

Even if France has several movie databases of good quality, an insight from the data collection process was that it is much more manual than it could have been with databases for US-movies that offer ready-made data extracts and data collection services.

Data preparation

The following variables have been used in the study, each requiring a different level of preparation.

Origin of the story

The different categories of origin of the story used in the study were defined during the data gathering. When a movie with a new category of origin of story appeared, that category was added as a variable in the dataset. This resulted in the 6 following dummy variables:

- **BOOK**: 1 if the movie is based on a Fictional Book or Short Novel, 0 if not.
- **COMIC**: 1 if the movie is based on a Comic/Graphic Novel, 0 if not.
- **MOVIE**: 1 if the movie is based on a Previous Movie, such as a Remake, 0 if not.
- **ORIGINAL**: 1 if the movie is based on an Original Screenplay, 0 if not.
- **PLAY**: 1 if the movie is based on a Play, 0 if not.
- **TRUE**: 1 if the movie is based on Real Life Events, 0 if not.

Production costs and Return on investment (ROI)

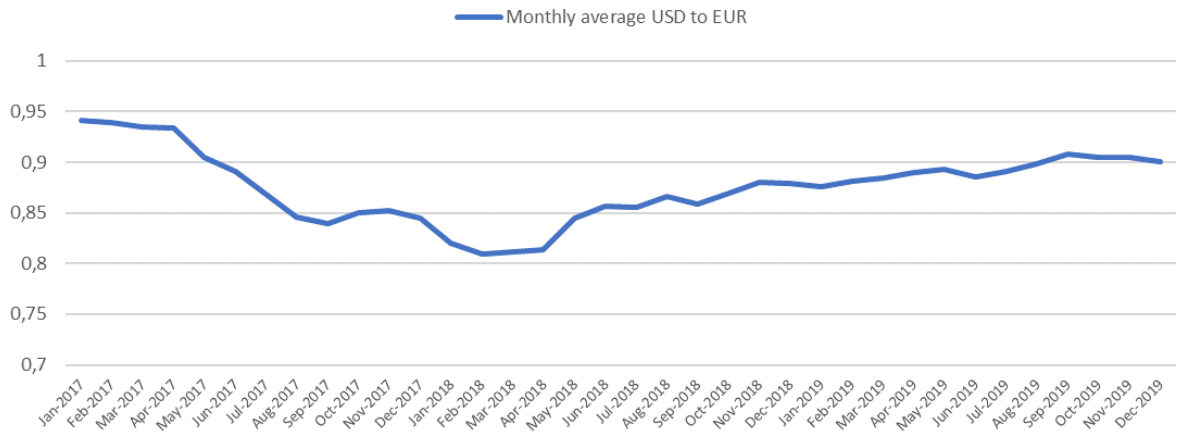
The ROI calculation requires the production cost and the worldwide revenue of each movie. Production costs for French movies were collected in EUR and are used in EUR in this study.

The Worldwide revenue of each movie was needed in order to calculate each movie's ROI. Production costs for French movies are given in Euros whereas worldwide box-office revenues are presented in US Dollars as this is a consolidation of several markets' revenue. This required a currency conversion from USD to EUR in order to be able to calculate a correct ROI.

To minimise the impact of the currency conversion, the monthly average for USD to EUR corresponding to each movie's release month was used. As most of a movie's revenue happens during the first weeks after release, this method was judged preferable compared to a yearly USD to EUR

currency rate or even to a EUR to USD conversion of the production cost, as it is not known when exactly the production cost was spent.

Figure 5: Monthly average currency rate from USD to EUR - 2017 to 2019



Note: Source X-Rates - https://www.x-rates.com/

Awards nomination

Following the same principles as Pangarker and Smit, the number of award nominations was extracted for each movie, including both Oscar nominations and César nominations by the French Academy Awards. In the dataset, this variable ranged from 0 (no Oscar or Awards nominations) to 12 for the movie “J'accuse", released in 2019.

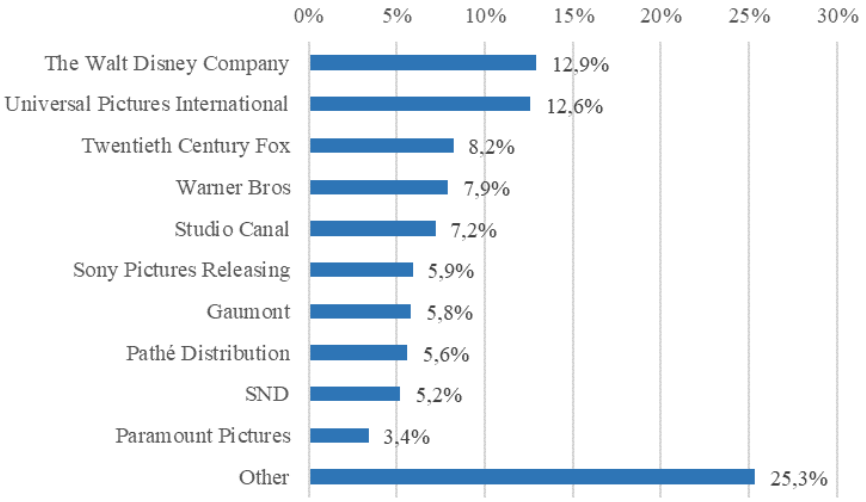
Professional critic reviews

Rating from professional critic reviews was extracted from the Allociné database. For each movie, Allociné gathers the reviews from up to 40 magazines specialised in movie and entertainment and normalises these reviews to a rating ranging from 0 to 5, with one decimal. Each movie receives a final rating from the average of all the individual ratings. In our dataset, the lowest rating is 1,4 for the movie “Les Municipaux (trop c’est trop)” (2019) and the highest is 4,6 for the movie “120 battements par minute” (2017).

Major distributor

Major distribution companies were identified as being among the 10 companies with the highest revenue market-share in France during 2017. They represent 74,7% of the whole market. In the dataset 99 out of 272 (36,4%) of the movies were released by a major distributor. Such movies received 1 for this variable, while the other received 0.

Figure 6: Market Share of 10 major movie distributor in France - 2017



Note: Source: CBO Box Office

Released during Holiday season

In France, holiday seasons are considered to be during school holidays. Each movie was therefore assigned a dummy binary variable with the value of 1 if released during school holidays, and 0 if not.

Genre

The genre of each movie was retrieved from the Allociné database. Same assumption was made as in the Pangarker and Smit's study and Action/Adventure and Drama were selected as the two most predictive genres. Each movie was assigned two dummy binary variables: one with the value of 1 if its genre was Action /Adventure and 0 if not and another with the value of 1 if its genre was Drama and 0 if not. Some movies of the dataset belonged to both genres, and some to neither of them.

Sequel

Information about a movie being a sequel or not was gathered manually. Each movie was assigned a dummy variable with the value of 1 if being the sequel to a previously released movie, and 0 if not.

Description of the dataset

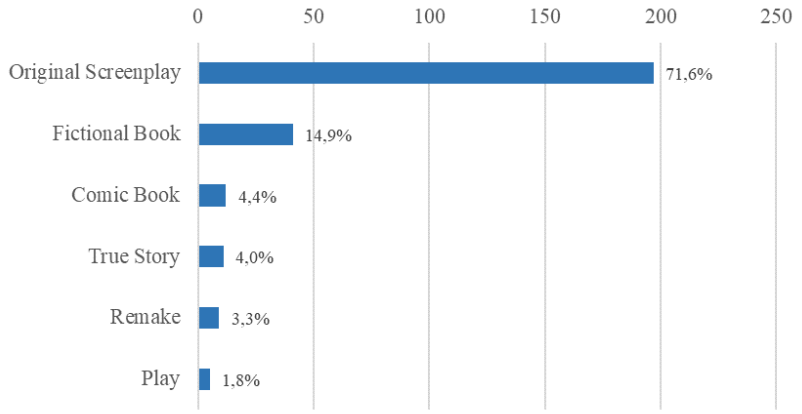
The dataset used in this study includes 272 movies produced in France and released between 2017 and 2019. The following tables and graphs describe the main aspects of this data set.

Table 3: Per year number of released movies produced in France in the study's dataset, their Average production costs (EUR) and Average worldwide revenue (EUR)

Year	Released movies produced in France	Average production costs per movie (EUR)	Average Worldwide Revenue per movie (EUR)
2017	110	9 484 754 EUR	5 279 680 EUR
2018	89	7 395 033 EUR	5 164 890 EUR
2019	73	7 518 427 EUR	5 723 049 EUR

*Note: S*ource: JP's Box-office

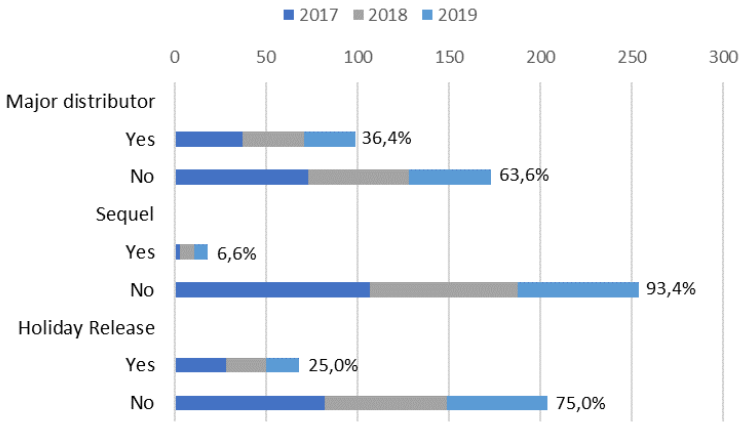
Figure 7: Number of movie from the dataset (2017-2019) per origin of the story



*Note: S*ource: Wikipedia

In the study's dataset we find that 71,6% of the French movies' stories are based on an original screenplay (see graph 6), whereas 44.5% of the US-produced movies were based on an original screenplay in 2019 (from nearly 60% in 2000). This difference shows that the French movie industry is still using a large share of original content when producing movies while the US movie industry has seen a decline in this type of origin of the story. It may show a stronger willingness in France than in the US to tell stories that have never been told before. It may also show the lack of existing content from French literature or other cultural sources that can be transformed into good movie material.

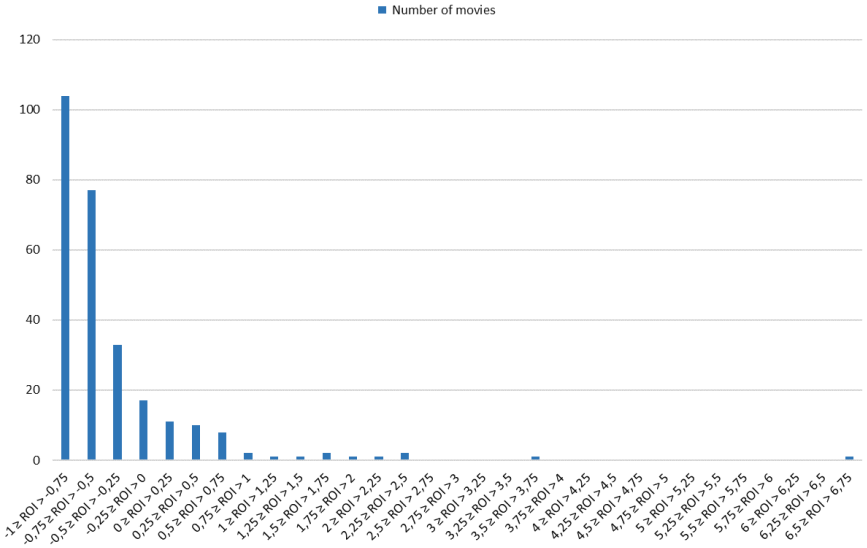
Figure 8: Number of movies per year from the dataset per "Major Distributor (Yes/No)", "Sequel (Yes/No)" and "Holiday Release (Yes/No)"



*Note: S*ource: JP's Box-office, Allociné

Other interesting aspects of the dataset are the explanatory variable. Graph 7 shows that 36,4% of the dataset's movies have been released by major distributors, 6,6% are sequels and 25,0% have been released during a holiday period.

Figure 9: Number of movies from the dataset per ROI



*Note: S*ource: JP's Box-office, CBO Box-office

Finally, an interesting observation from the dataset concerns the movies' ROI, where the data appears highly skewed. Another observation is the low ROI of French movies. In the study's dataset, only 42 of 272 (i.e. 15,4%) of the movies have a positive ROI and made a profit. This is in line with earlier study (CNC, 2004; Bosel & Chamaret, 2008; BMFTV; 2014) showing that roughly 10% of the movies produced in France were profitable. A characteristic of the French movie industry is that it is

highly publicly funded by the ministry of the culture and other institutions, making it possible to survive as a business despite high losses.

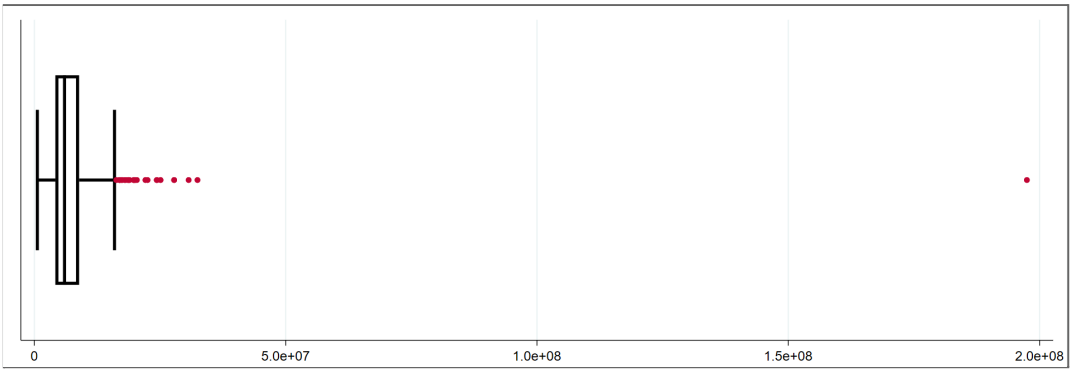
Data verification

Outliers in the data

In order to control for outliers in the dataset, a box plot was constructed for all variables that were not dummy variables. These box plots can be found in the appendix (C.f. Appendix 1).

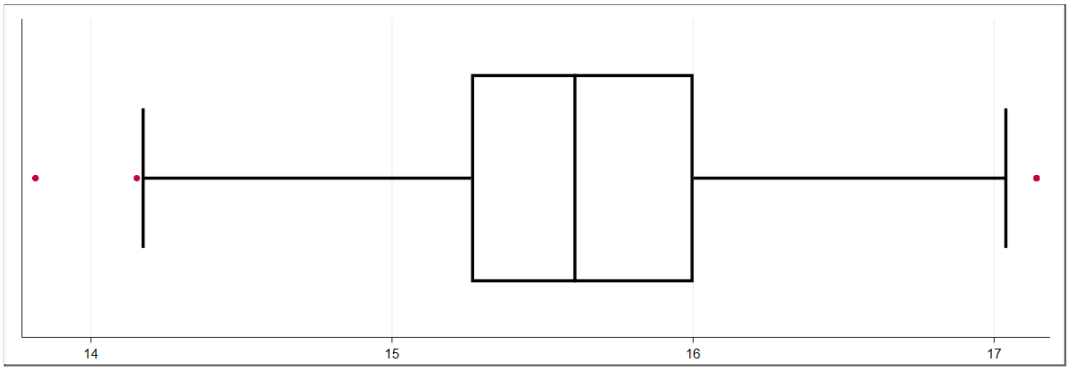
On figure 2, the box plot for the variable BUDGET shows multiple outliers, and one that is particularly extreme. The author proceeded to use the log function on the variable, followed by the winsor function for three extreme values on both sides. On figure 3 we can see the box plot for the new variable BUDGET, after having been logged and winsorized.

Figure 10: Box plot of the variable BUDGET, before log and winsor (in euros)



*Note: S*ource: Author rendering of data from CBO Box-office, JP's Box-office

Figure 11: Box plot of the variable BUDGET, after log and winsor



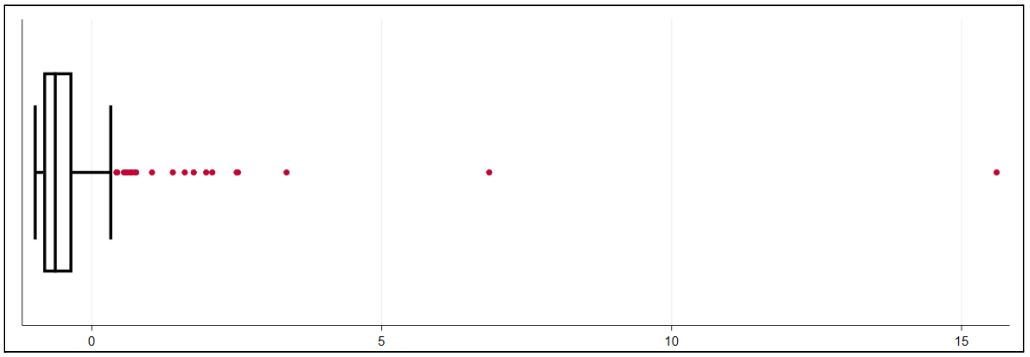
*Note: S*ource: Author rendering of data from CBO Box-office, JP's Box-office

The box plot for the dependent variable ROI also shows multiple outliers, as can be seen on figure 4. There are especially two (2) extreme outliers. After inspecting the data, it was revealed that the movies in question were "Les Misérables" and "Les Invisibles". Both worldwide revenues and

production costs were verified for these movies. The data on “Les Misérables” was correct, however “Les Invisibles”, which is the most extreme outlier on the boxplot, had wrong production costs, the one from another movie with the same exact title but released in 2012. This shows the value of analysing the outliers before proceeding to the regression analysis, as well as the difficulties to work with several databases that are not harmonised.

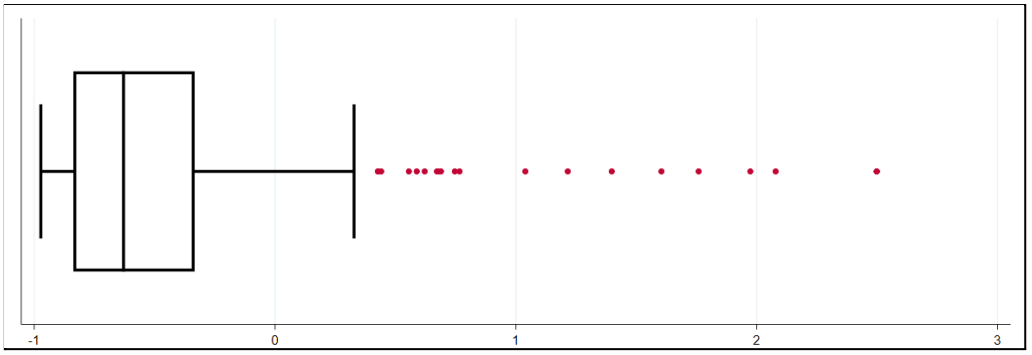
The production costs were corrected in the dataset, and the ROI was thereafter winsorized with 3 values on each side. The new box plot of the ROI after correction and after winsor can be seen below on figure 5.

Figure 12: Box plot of the dependent variable ROI, before winsor



*Note: S*ource: Author rendering of data from CBO Box-office, JP’s Box-office

Figure 13: Box plot of the dependent variable ROI, after correction in data and winsor



*Note: S*ource: Author rendering of data from CBO Box-office, JP’s Box-office

Finally, the extreme values in the variable AWARDS (C.f. figure A3 in the Appendix) were not considered to be outliers, as it is normal that most movies have 0 awards, and some of them have many.

Multicollinearity

If there is a strong correlation between multiple independent variables, it is said that there is multicollinearity. Multicollinearity can increase the variances of the parameter estimates, thereby

possibly causing insignificant predictors even though the overall model is significant (Joshi 2012). However, multicollinearity can be detected by examining a correlation matrix (Belsley et al. 1980).

Table 4: correlation matrix with all variables of the final model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
ROI (1)	1.000														
BUDGET (2)	**.121***	1.000													
ACTION (3)	.071	**.210***	1.000												
DRAMA (4)	-.091	***.138**	-.098	1.000											
AWARDS (5)	**.271***	.059	-.050	**.255***	1.000										
CRITIC (6)	.099	***.270**	-.084	**.367***	**.472***	1.000									
MAJOR (7)	**.216***	**.433***	.116	-.108	.059	-.105	1.000								
HOLIDAY (8)	.047	.048	-.012	-.082	.063	-.054	-.049	1.000							
SEQUEL (9)	**.181***	**.195***	.033	-.106	-.030	***.276**	.014	-.017	1.000						
BOOK (10)	-.088	.101	-.078	**.197***	**.132***	.105	.066	-.101	-0.11	1.000					
COMIC (11)	-.007	**.181***	**.160***	-.114	-.081	-.083	.098	.041	.087	-.091	1.000				
MOVIE (12)	-.029	.103*	-.036	-.010	-.036	-.074	.055	-.023	.027	.027	-.042	1.000			
ORIGINAL (13)	.076	***.230**	-.116	***.128**	-.044	-.030	***.149**	.090	.065	***.683**	***.348**	***.317**	1.000		
PLAY (14)	-.006	.027	-.025	-.007	.014	-.050	.067	-.016	-.036	-.058	-.029	-.027	***.222**	1.000	
TRUE (15)	.057	**.131***	**.275***	**.161***	.048	.099	.116	-.032	-.055	.018	-.044	.059	***.333**	-.028	1.000

Note: values with a * have a significance of p-value < 0.05 ; Source: Author rendering of data from CBO Box-office, JP's Box-office, Allociné, Wikipedia

As we can see on the correlation matrix in table 4, only a few of the correlations are statistically significant. This is problematic as we can not guarantee that there is not any multicollinearity. For the significant correlations, only two variables have a large enough correlation between each other for it to potentially be an issue. These are the variables BOOK and ORIGINAL, with a correlation of -0.683 on a significance level of 0.05. This makes sense, as movies with original stories are the most prevalent in dataset (72.4% according to table 6), followed by movies based on books, which represents 15.1% of the dataset according to table 6. Therefore, when a movie is not based on a book, the likelihood of that movie being based on an original story is extremely large. In the same way, if a movie is not based on an original story, it is likely that it is based on a book. The variable ORIGINAL was thus deleted from the regression, as the effect of this variable is captured by all the other story origins. If all other story origin variables are 0, it means that the movie is based on an original story.

V. Method

Descriptive statistics

Control variables

This study seeks to understand the effect of the origin of a story on a movie's profitability. However, in order to isolate the effect of the origin of the story, it is necessary to control for other variables. The control variables need to be limited in order to not overfit the regression model. The sample size of a dataset limits how many variables can be included in the model without overfitting the model. For this study there are 272 complete data points and therefore control variables will need to be limited. An overfit model can cause the regression coefficients, p-values, and R-squared to be misleading, which is an important factor to consider.

Regardless of the number of control variables, it is important to include the most important variables, as omitted variables can cause the coefficient of one or more explanatory variables in the model to be biassed. In order to both limit the number of control variables and include the most important ones, the choice of control variables of this study will be based on the same variables as the one used by Pangarker and Smit (2013). In their study, Pangarker and Smit identified the explanatory variables which have been thoroughly studied in the literature and are considered to be the most relevant variables to include in a regression model for the financial performance of a movie.

A summary of the descriptive statistics of the control variables can be found in table 5 below. Omitted variables studied in the literature will be discussed more in detail in the Discussion section.

Table 5: Descriptive statistics of the control variables

Variable	Obs	Mean	Std. dev.	Min	Max
BUDGET	272	15.647	.630	13.816	17.141
ACTION	272	.033	.179	0	1
DRAMA	272	.221	.415	0	1
AWARDS	272	.787	2.081	0	12
CRITIC	272	3.103	.615	1.4	4.6
MAJOR	272	.364	.482	0	1
HOLIDAY	272	.250	.434	0	1
SEQUEL	272	.066	.249	0	1

Note: The dependent variable BUDGET has been logged and winsorized, which is why its values aren't the actual movie production costs; The variables ACTION, DRAMA, MAJOR, HOLIDAY and SEQUEL are dummy variables. Source: Author rendering of data from CBO Box-office, JP's Box-office, Allociné, Wikipedia

Independent variables

The independent variables are different dummy variables all denoting a category of origin of the story. There was originally supposed to be 6 different dummy variables, one for each category of origin of the story: a dummy variable when the origin of the story is from a book (BOOK), from a comic book (COMIC), from a previously made movie (MOVIE), from a play (PLAY), from an original screenplay (ORIGINAL) and from a true story (TRUE). However, as mentioned in the Data section, the origin of the story from an original screenplay (ORIGINAL) has been deleted from the regression as its effect is already captured by all the other 5 categories.

Table 6 below shows the descriptive statistics of the independent variables. It should be noted that all variables are dummy variables that can only be of value 1 or 0. The mean of these different variables can therefore be interpreted as the proportion of which it is found in the dataset. As we have removed the ORIGINAL dummy variable for original screenplay stories (that represented 72.4% of the movies from the dataset), the most prevalent origin of a story is from a fictional book or novel (BOOK), with 15.1% of all movies in the dataset. All other categories are present in less than 5% of the movies in the dataset.

Table 6: Descriptive statistics of the independent variables

Variable	Obs	Mean	Std. dev.	Min	Max
BOOK	272	.151	.358	0	1
COMIC	272	.044	.206	0	1
MOVIE	272	.037	.189	0	1
PLAY	272	.018	.135	0	1
TRUE	272	.040	.197	0	1

Note: all variables of the table are dummy variables. Source: author rendering of data from CBO Box-office, JP's Box-office, Allociné, Wikipedia

Dependent variable

The dependent variable of the regression is the Return-on-Investment (ROI) of the movie:

$$ROI = \frac{revenue - production\ costs}{production\ costs}$$

As presented in the theoretical framework, the decision was made to modify this aspect from Pangarker and Smit's research. The dependent variable in their study is the movie box-office revenue and not the ROI, which is only partially correlated with revenue (our dataset shows a correlation of 0.6781 between Revenue and ROI with a 0.05 significance level). As described in the Data section, the reason for changing the dependent variable to ROI at the cost of less comparability with the work of Pangarker and Smit is that ROI is a more valuable metric to evaluate the success of a movie for an investor. A movie can have an enormous revenue, but an even greater cost. By analysing the ROI, the

objective is to favour good investments over over-costly blockbusters. Changing this variable does not create any fundamental changes in the regression methodology of this study. However, it does prevent the results from being directly compared with the results from the Pangarker and Smit's research.

With the choice of ROI instead of revenue as dependent variable, it became necessary to apply worldwide revenue when calculating the ROI. The production costs used in the ROI calculation cannot be split geographically and therefore the revenue used in the ROI formula must be the worldwide revenue to obtain a correct ROI calculation. As movies produced in France are mostly targeting a French speaking audience, the movies' revenue coming outside of France, and representing 15.6% of the total revenue of our dataset, is assumed to come mostly from French speaking countries such as Belgium, Switzerland and French Canada. The impact of the origin of the story in these countries is assumed to be similar to France.

Table 7: Descriptive statistics of the dependent variable

Variable	Obs	Mean	Std. dev.	Min	Max
ROI	272	−.447	.629	−.973	2.498

Note: ROI is not given in % but in decimal value. Source: Author rendering of data from CBO Box-office, JP's Box-office

Stepwise regression

Stepwise regression is a process of building a model by successively adding or removing variables based on the t-statistics of their estimated coefficients. Properly used, the stepwise regression allows for more information and better analysis than the ordinary regression.

The stepwise procedure can either be forward or backward. If it is a forward stepwise regression, it starts with no variables in the model and proceeds by adding one variable at a time. If it is a backward stepwise regression, it starts with all potential variables in the model and proceeds by removing them one by one. In this study, a forward stepwise regression was used, as the variables used in the final regression have been decided beforehand, thus making the backwards method trivial. The forward method, on the other hand, can still allow for a more in depth analysis of the results, even though the final regression has been decided beforehand.

Thus, a forward stepwise regression was used to more thoroughly analyse the different predictors of the ROI of a movie than simply conducting a direct multivariate regression. The chosen control variables were the same as in the final regression, meaning they were all added during the stepwise procedure. At each step, variables were added based on t-values. The t-values of each variable shown at each step of the stepwise regression can be seen in table A1 of the appendix.

As this study is based on the work of Pangarker and Smit (2013), the final regression uses the same variables as they do. The model is specified as the following, where β denotes the coefficient of the different variables, and i denotes a particular movie:

$$\begin{aligned} ROI \;=\; & \beta_0 + \beta_1 Budget_i + \beta_2 Action_i + \beta_3 Drama_i + \beta_4 Major_i \\ & + \beta_5 Holiday_i + \beta_6 Award_i + \beta_7 Critic_i + \beta_8 Sequel_i \\ & + \beta_9 Book_i + \beta_{10} Comic_i + \beta_{11} Movie_i + \beta_{12} Play_i + \beta_{13} TrueStory_i \\ & + \varepsilon_i \end{aligned}$$

In the first two lines of the equation we find the same regression as Pangarker and Smit (2013), of which the variables serve as control variables in order to understand the effect of the origin of the story. In the third line we find all the variables that represent the origin of the story and that have been added to Pangarker and Smit's regression (2013). They are dummy variables that will respectively take on the value 1 if the origin of story is from a book, a comic, a movie, from an original screenplay, from a play or if it's based on a true story. In the last line of the equation we find epsilon, the error term of the regression.

It was made sure that good scientific measures were taken to ensure the robustness of the regression. There were no variables included in the stepwise regression that would not be added to the final model, i.e. there was no excessive amount of variables included in the stepwise regression that would allow to "fish" for significant coefficients. All variables had a reason to be selected as they were chosen from the Pangarker and Smit study and their choice was also supported by the literature review, showing that all had a reasonable impact on a movie's financial performance.

For every added variable, the value of the adjusted R-squared of the model will be observed, in order to understand how much more variance is explained by the addition of the new variable. By looking at R-squared, it is possible to see how well the model fits the data and if it can explain changes in the dependent variable. R-squared adjusted adjusts for the amount of variables in the model. Since variables are being added one by one, this is a more appropriate metric.

Process for interpreting the final regression

Firstly, the residual plots will be analysed in order to check for unwanted patterns in the residual. In order for the assumptions of the multivariate regression model to hold, no explanatory power should be in the error term. In other words, there should be no observable pattern in the distribution of the residuals. This is the constant variance assumption, or homoscedasticity. This can be confirmed by observing that the residuals are randomly scattered around zero for the entire range of fitted values. Otherwise, the regressions coefficients as well as other results can not be considered valid.

After that, the distribution of the residuals will be verified in order to make sure that they present a normal distribution around the fitted line of the model, as one of the assumptions of a multivariate regression is that the residuals should be approximately normally distributed.

Then, the p-values for each given variable will be analysed, in order to determine whether the observed relationships in the sample also exist in the larger population. For each independent variable, the p-value tests the null hypothesis that the variable has no correlation with the dependent variable. Secondly, these p-values will be interpreted by comparing them to their respective variables' significance level. The chosen significance level in this study will be 0.05. If the p-value is lower than 0.05, the null hypothesis will be rejected and it will be concluded that the independent variable and the dependent variable are correlated. If the p-value is higher than 0.05, we can not conclude that there is a correlation between the independent and the dependent variable.

Finally, the coefficients will be interpreted to reach any conclusions on the effect of the variables.

VI. Results

Stepwise regression analysis

The table 8 below shows the summary of the stepwise regression:

Table 8: Model summary for the stepwise regression

Model [added variable]	R-squared	R-squared adjusted	Std. error of the estimate	Change in R-squared adj.
Regression #1 [AWARDS]	.0736	.0702	.6068	.0702
Regression #2 [MAJOR]	.1135	.1069	.5947	.0367
Regression #3 [SEQUEL]	.1489	.1394	.5838	.0325
Regression #4 [DRAMA]	.1639	.1514	.5797	.0120
Regression #5 [BOOK]	.1733	.1578	.5775	.0064
Regression #6 [CRITIC]	.1813	.1627	.5758	.0049
Regression #7 [TRUE]	.1833	.1616	.5762	-.0011
Regression #8 [COMIC]	.1846	.1598	.5768	-.0018
Regression #9 [ACTION]	.1860	.1581	.5774	-.0017
Regression #10 [HOLIDAY]	.1871	.1559	.5781	-.0022
Regression #11 [MOVIE]	.1878	.1535	.5790	-.0024
Regression #12 [PLAY]	.1881	.1505	.5800	-.0030
Regression #13 [BUDGET]	.1881	.1472	.5811	-.0033

Note: Each model is a step in the stepwise regression. The variable in the square brackets is the variable that was added to the model after the previous regression. Source: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

According to the result of the stepwise regression, the variables which seem to explain the largest portion of the variation in the dependent variable ROI are AWARDS (number of César and Oscar nominations a movie has received), MAJOR (whether or not the movie was distributed by a major distribution company) and SEQUEL (whether or not the movie was a sequel). R-squared adjusted increases once these variables are added to the model: the metric increases by 0.0702 for AWARDS, 0.0367 for MAJOR and 0.0325 for SEQUEL. This means that the variable AWARDS explains 7.02% of the variation in the dependent variable, and that MAJOR and SEQUEL explain 3.67% and 3.25% of the variation in the dependent variable respectively. DRAMA (whether or not a movie is of the drama genre) explains 1.20% of the variation in the dependent variable.

All other variables lead to an increase in R-squared adjusted that is below 1 percentage point. The majority of them even have a negative effect. The variables TRUE (whether or not a movie is based on a true story), COMIC (whether or not a movie is based on a comic book), ACTION (whether or not the movie is in the action genre), HOLIDAY (whether or not the movie was released during a holiday season), MOVIE (whether or not a movie is based on a movie), PLAY (whether or not a

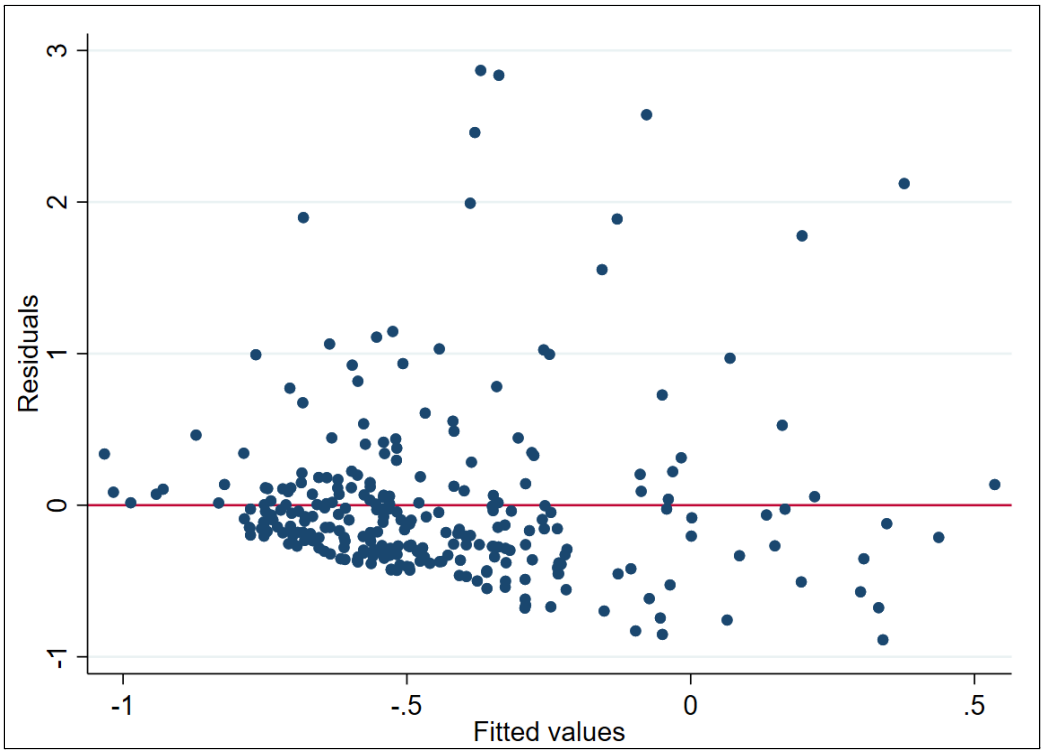
movie is based on a play) and BUDGET (the production costs of the movie) all lead to negative changes in R-squared adjusted. This means that none of these variables improve the model fit by a sufficient amount. Details on each regression of every step can be found in the Appendix 2.

Additionally, the result of the stepwise regression shows that the complete regression (Regression #13) has an R-squared of 0.1881, which means that the model explains 18.81% of the variation in the dependent variable. If we look at R-squared adjusted, which adjusts for the number of variables, the model explains 14,72% of the variation in the dependent variable.

Homoscedasticity

A preliminary regression was conducted in order to analyse the residual plot, which can be seen on figure 6. The residuals appear to be randomly dispersed around 0 once the fitted values go past the value -1. Before -1 however, there seems to be a trend of residuals being above 0, thus not being well dispersed around 0. Overall, there seems to be a trend of residuals having a larger variance when they are positive, while the negative residuals seem to be clustered together. This can be explained by the fact that a movie has a floor for how much it can "flop". However, there is no upper limit for a movie's success. It is therefore expected to see a higher variance on the positive side of the residuals than on the negative one.

Figure 14: Residual plot for the model

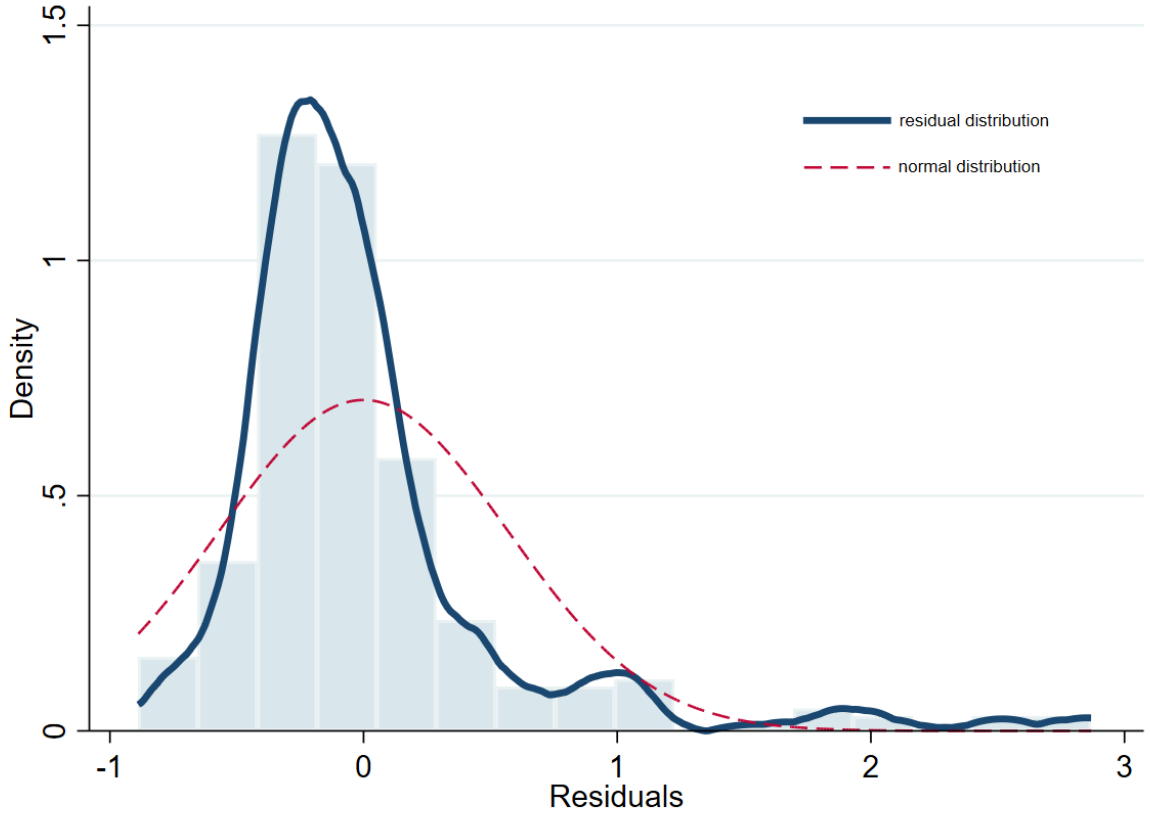


*Note: S*ource: Author rendering of data from CBO Box-office, JP's Box-office, Allociné, Wikipedia

Distribution of residuals

As we can see on figure 7 below, the residuals issued from the multivariate regression have a distribution practically centred around 0 and follow a distribution very close to a normal distribution, albeit with a high spike of density at 0.

Figure 15: Histogram of residuals compared to a normal distribution



*Note: S*ource: Author rendering of data from CBO Box-office, JP's Box-office, Allociné, Wikipedia

Final regression

The table below show the final result of the multivariate regression:

Table 9: Final multivariate regression

ROI	Coefficient	Std. error	t	P > | t |	[95% conf.	interval]
AWARDS	.077	.020	3.81	.000	.037	.117
MAJOR	.263	.083	3.19	.002	.101	.426
SEQUEL	.502	.152	3.30	.001	.203	.801
DRAMA	−.212	.096	−2.22	.027	−.400	−.024
BOOK	−.166	.104	−1.60	.111	−.371	.038
CRITIC	.113	.074	1.51	.132	−.034	.259
TRUE	.110	.194	0.57	.569	−.271	.492
COMIC	−.140	.179	−0.78	.436	−.493	.213
ACTION	.138	.214	0.64	.520	−.283	.558
HOLIDAY	.046	.083	0.55	.580	−.118	.210
MOVIE	−.095	.190	−0.50	.617	−.469	.279
PLAY	−.078	.265	−0.29	.769	−.600	.444
BUDGET	−.003	.069	−0.04	.970	−.138	.133
constant	−.883	1.137	−.78	.438	−3.121	1.356
R-squared	.188					
F-value	4.60					

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

The results of the regression shows that the p-values associated with the estimated coefficients of AWARDS, MAJOR, SEQUEL and DRAMA have respective values of 0.000, 0.002, 0.001 and 0.027. Because the p-values are lower than the significance level 0.05, we can reject the null hypothesis that the independent variables have no correlation with the dependent variable. We can thus conclude that the (1) amount of awards nominations a movie has earned, (2) whether or not the movie was distributed by a major distributor, (3) whether or not the movie is a sequel and (4) whether or not the movie is of the drama genre are all variables that are correlated to the ROI of French movies.

For all other variables of the regression, the p-value is higher than the significance level of 0.05. This means that there is not enough information to reject the null hypothesis, and therefore it can not be concluded that there exists a correlation between these variables and the ROI of French movies.

The variable AWARDS has a positive coefficient, which remains positive in the entire 95% confidence interval. It can be concluded that when a French movie has award nominations, its ROI tends to increase. The coefficient value is 0.077, which suggests that a one-unit change in the number of award nominations of a French movie leads to a mean change in ROI of 0.077, holding all other

variables constant. In other words, one extra award nomination increases ROI with 7.7 percentage points.

The variable MAJOR has a positive coefficient, which remains positive in the entire 95% confidence interval. It can be concluded that when a French movie is made by a major distributor, its ROI tends to increase. The coefficient value is 0.263, which suggests that if a French movie is made by a major distributor, it leads to a mean change in ROI of 0.263, holding all other variables constant. In other words, increasing ROI with 26.3 percentage points.

The variable SEQUEL has a positive coefficient, which remains positive in the entire 95% confidence interval. It can be concluded that when a French movie is a sequel, its ROI tends to increase. The coefficient value is 0.502, which suggests that if a French movie is a sequel, it leads to a mean change in ROI of 0.502, holding all other variables constant. In other words, increasing ROI with 50.2 percentage points.

The variable DRAMA has a negative coefficient, which remains negative in the entire 95% confidence interval. It can be concluded that when a French movie is of the drama genre, its ROI tends to decrease. The coefficient value is -0.212, which suggests that if a French movie is of the drama genre, it leads to a mean change in ROI of -0.212, holding all other variables constant. In other words, decreasing ROI with 21.2 percentage points.

VII. Discussion

Comparison with Pangarker and Smit

One objective of this study is to compare its results with the Pangarker and Smit's study on which it was based. From what they describe as the determinants of Revenue (see table A15 in Appendix 3), this study shows similar findings. Pangarker and Smit show that Major distributor, Award nomination and Sequel all have a positive impact on revenue. According to the results of this study, the same variables have a positive effect on ROI. Although this study does not find any significant results on BUDGETs coefficient (i.e. production costs), a correlation between production costs and ROI can be seen in the correlation matrix in table 4 with a value of 0,121. This is similar to the positive correlation between production costs and revenue shown by Pangarker and Smit, even if the correlation value in this case was much stronger with a value of 0.735. This shows that high production costs lead to high revenue but less so to profitability. Moreover, there exists a possibility that the positive correlation between production costs and ROI is linked to the major distributors, who have a positive impact on ROI and often distribute movies with large production costs (MAJOR and BUDGET show a correlation of 0.433).

One difference with the Pangarker and Smit research is that the drama genre shows a negative impact on the movies financial performance in this study while Pangarker and Smit do not find a significant regression coefficient with revenue, even if they find a negative correlation. It is worth mentioning that Lash and Zhao also found a negative regression coefficient between the genre Drama and the ROI of a movie, in line with the results of this study. These similar results tend to show that the Drama genre albeit an important one from a story-telling perspective is not driving profitability.

Finally, It is noteworthy to see that all correlations that are significant in this study at the 0,05 level (see matrix in table 4) are very similar to the corresponding correlations in the Pangarker and Smit's study. This can be seen as a validation of the results of both the Pangarker and Smit's study and the present study. It also underlines similarities between the reception of French movies and US movies.

The effect of the origin of the story

There are no significant results for the coefficients of variables that refer to the different origins of the story. One hypothesis is that there is not enough data in the dataset to reach significant results. Except for "original screenplay" and "fictional books or novel", all other categories of the origin of the story represent less than 5% of the dataset. As these proportions are most probably representative, a greater sample size seems required in order to reach conclusions on this topic. One of the main challenges in the field of this study is the fact that available data is not normalised or consolidated. Every research uses different data sources and different ways of normalising the data. This is even more true for the

origin of the story, where this data had to be gathered and structured manually. The fact that the data on the origin of the story was gathered manually is also one of the strengths of this study. However, the more normalised, structured and available the data will become, the larger the data sample it will be possible to analyse.

There are nevertheless some interesting takeaways about the origin of the story that can be gathered from this study. The correlation matrix (see table 4) shows significant positive correlation between the variable BOOK and AWARDS, which suggest that a French movie based on a fictional book is more likely to be nominated for an award. This could be explained by the fact that a movie based on a fictional book may have more artistic depth. Could this also mean that movies based on fictional books are more likely to become better movies? This is an area that would be worth doing further studies in.

Another interesting finding is that COMIC (story based on comic books or graphic novels) and TRUE (story based on a real life event) both have a significant positive correlation with BUDGET (production costs) and ACTION (movie from the action genre). This suggests that movies based on comic books and true stories are correlated with higher production costs as well as correlated with being in the action genre. The higher production costs could be explained by the fact that the origin of the story (being a comic book or a real life event) has already been visualised by the audience and that the movie makers have to take this into account and cannot spare on reproducing these already seen images.

Limits

As mentioned above, a notable limit of this study is the lack of data due to the difficulty of access and the lack of normalisation and consolidation. As the majority of movies have original stories, the necessary amount of data to reach significant levels for the variables of origin of story becomes quite substantial.

Another limit in this study is that some variables that are considered important in the literature were not included in the regression as they were not part of Pangarke and Smit's study. Advertising cost is a variable that most likely has an impact on a movie's revenue, the question being if it has an impact on a movie's ROI. As earlier research has shown, advertising costs are strongly correlated to production costs and their effect may already be taken into account by the production costs. This should be confirmed by future research, even if this type of data is rarely disclosed by movie distributors. Another variable not taken into account in this study is "star power", meaning the impact of specific actors or directors on the success and profitability of a movie. This variable has been studied in several earlier research. However, it was decided to disregard star power in this study's

regression, as there is little consensus in the literature on how star power should be measured and each study has used different ways of assessing it.

Finally, a third limitation in this study is the fact that the revenue only accounts for box office in theatres, which despite being the major source of income of movies, is still only a portion of a movie's revenue and does not include incomes from digital channels. This is an issue that can not be solved from the data that is accessible to researchers. This is problematic as it means that the ROI of the movies in this dataset is not the complete ROI of the movies. However, the box office of a movie is most likely correlated with the revenue streams of other distribution channels. Therefore, the coefficients and correlations of this study are still indicative and conclusions can therefore be made.

Future research

The first priority for future research on the origins of the story of movies should be gathering, normalising and consolidating data. Recreating this study with more data would be valuable. An idea could be to use web-scraping in order to gather the data automatically.

Good movies are not always financially successful. However, it can still be valuable to know how to make good movies. This study suggests that there is a positive correlation between movies based on books and number of award nominations. Maybe the origin of the story has a positive effect on the artistic value of a movie and its quality. This could be interesting to study, apart from the financial success of a movie.

VIII. Conclusion

To conclude, this study was conducted in order to gather more knowledge about the factors that affect the revenue of a movie. This with the ultimate goal of predicting if a movie will be successful early in the decision-making process. As there are reasons to believe that the origin of story has an impact on returns, and since origin of story is one of the few factors that are known early in the decision-making process, the author decided to study the effect of origin of story on ROI of movies.

The study contributes to the literature in 3 ways: First of all, its findings are consistent with previous studies focusing on different regions. Indeed, Pangarker and Smit show that Major distributor, Award nomination and Sequel all have a positive impact on revenue in the US. According to the results of this study, the same variables have a positive effect on ROI in France. The second contribution is that study shows that the origin of the story of a movie is correlated with different variables. For example, movies based on books are positively correlated with being nominated for awards, while movies based on comic books are positively correlated with higher budgets. The third contribution is the finding that the origin of the story is correlated with profitability. Movies based on books and comic books have a strong negative correlation with profitability. However, these results aren't significant and thus aren't conclusive.

This is why future studies should focus on gathering more data, in order to find significant results. Other dependent variables could also be studied, such as popularity and quality, in order to understand how the origin of the story affects a movie beyond its profitability.

IX. References

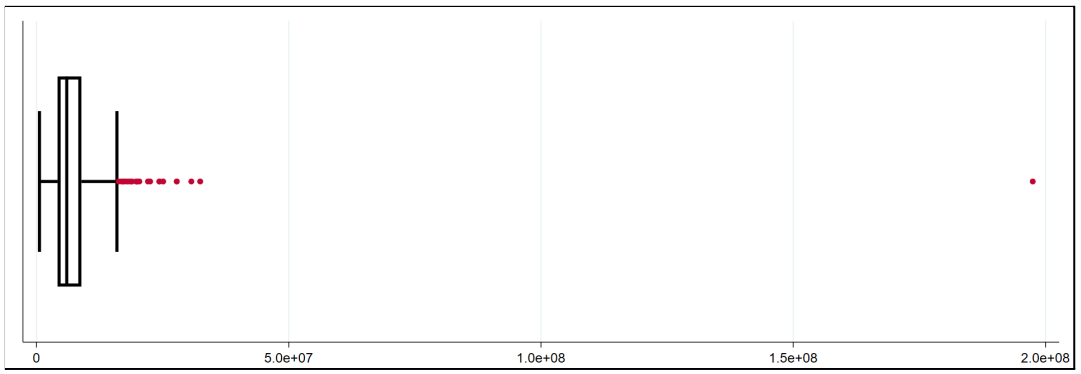
- Anast, P. (1967). 'Differential movie appeals as correlates of attendance', Journalism Quarterly, 44: 86-90.
- Austin, B.A. (1984). 'Portrait of an art film audience', Journal of Communication, 34(winter): 74-87.
- Austin, Bruce A. & Thomas F. Gordon (1987). Movie Genres: Toward a Conceptualized Model and Standardized Definition. In Current Research in Film: Audiences, Economics and the Law, Vol. 3.
- Basuroy, S., Chatterjee, S. and Ravid, S. A. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. Journal of Marketing, 67(4), 103–117.
- Belsley, D. A. (1980). On the efficient computation of the nonlinear full-information maximum-likelihood estimator. Journal of Econometrics, 14(2), 203-225.
- Bomsel, O., & Chamaret, C. (2008). Rentabilité des investissements dans les films français. Paris: Centre d'économie industrielle (CERNA).
- Caves, R. (2000) Creative Industries, Harvard University Press, Mass.
- Caves, R. (2001) Creative Industries: Contracts Between Art and Commerce. Harvard University Press, Cambridge, MA.
- Chakravarty, A., Liu, Y. and Mazumdar, T. (2010). The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation. Forthcoming at Journal of Interactive Marketing.
- Chang, B.-H. and Ki, E.-J. (2005). Devising a Practical Model for Predicting Theatrical Movie Success: Focusing on the Experience Good Property. Journal of Media Economics, 18(4), 247–269.
- Choudhery and Carson K Leung. (2017) Social media mining: prediction of box office revenue. In Proceedings of the 21st International Database Engineering & Applications Symposium, pages 20–29.
- De Silva, I. (1998). Consumer selection of motion pictures. In B. R. Litman, ed. The Motion Picture Mega-Industry.
- De Vani & Walls. (1999). Uncertainty in the movie industry: Does star power reduce the terror of the box office? Journal of Cultural Economics, 23:285–318.
- De Vany, A. & Walls, W.D. (2002). 'Does Hollywood make too many R-rated movies? Risk, stochastic dominance, and the illusion of expectation', The Journal of Business, 75(3): 425-451.
- Deniz, B. and Hasbrouck, R. B. (2012). What Determines Box Office Success of a Movie in the United States? Proceedings for the Northeast Region Decision Sciences Institute, (757), 447.
- Deuchert, E., Adjamah, K. and Pauly, F. (2005), 'For Oscar Glory or Oscar Money?' Journal of Cultural Economics, 29(3), 159-176.
- Dodds, J. C., M. B. Holbrook. (1988). What's an Oscar worth? An empirical estimation of the effects of nominations and awards on movie distribution and revenues. B. A. Austin, ed. Current Research in Film: Audiences, Economics, and Law, Vol. 4. Ablex, Norwood, NJ, 72–88.
- Duan, W., Gu, B. and Whinston, A. (2008). The Dynamics of Online Word-of-Mouth and Product Sales – An Empirical Investigation of the Movie Industry. Forthcoming at Journal of Retailing.
- Einav, L. (2007), 'Seasonality in the U.S. Motion Picture Industry', Rand Journal of Economics, 38(1), 127-145.
- Elberse, A. and Eliashberg, J. (2003), 'Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures', Marketing Science, 22(3), 329-354.
- Eliashberg, J., Elberse, A. and Leenders, M. (2006), 'The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions', Marketing Science, 25(6), 638-661.
- Eliashberg, J., J.-J. Jonker, M. S. Sawhney, B. Wierenga. (2000). MOVIEMOD: An implementable decision-support system for prerelease market evaluation of motion pictures. Marketing Sci. 19(3) 226–243.
- Eliashberg, Jehoshua, Quintus Hegie, Jason Ho, Dennis Huisman, Steven J. Miller, Sanjeev Swami, Charles B. Weinberg, and Berend Wierenga. (2009). Demand-driven scheduling of movies in a multiplex. International Journal of Research in Marketing 26: 75–88
- Eliashberg, S. Hui, and Z. Zhang. (2007). Assessing box office performance using movie scripts: A kernel-based approach. Knowledge and Data Engineering, IEEE Transactions on, 26(11):2639–2648, Nov 2014.
- Eliashberg, S. K. Hui, and Z. J. Zhang. From Story Line to Box Office: A New Approach for Green-Lighting Movie Scripts. Management Science, 53(6):881–893.
- Feldman, J. M., & Lynch, J. G. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. Journal of applied Psychology, 73(3), 421.
- Ghiassi, M., Lio, D. and Moon, B. (2015). Pre-production forecasting of movie revenues with a dynamic artificial neural network. Expert Systems with Applications, 42(6), 3176–3193. https://doi.org/10.1016/j.eswa.2014.11.022
- Hand, C. (2001), 'Increasing returns to Information: Further Evidence from the UK Film Market', Applied Economics Letters, 8, 419-421.
- Hennig-Thurau, T., & Houston, M. B. (2019). Entertainment Science. Data Analytics and Practical Theory for Movies, Games, Books, and Music. Cham: Springer International.
- Hunter, S., Smith, S. and Singh, S. (2016). Predicting Box Office from the Screenplay: An Empirical Model. Journal of Screenwriting, 7(2)
- Joshi, A., & Mao, H. (2012). Adapting to succeed? Leveraging the brand equity of best sellers to succeed at the box office. Journal of the Academy of Marketing Science, 40(4), 558-571.
- Juliusson, E. Á., Karlsson, N., & Gärling, T. (2005). Weighing the past and the future in decision making. European journal of cognitive psychology, 17(4), 561-575.

- Kim, Taegu, Jungsik Hong, and Pilsung Kang. (2017). Box Office Forecasting considering Competitive Environment and Word-of-Mouth in Social Networks: A Case Study of Korean Film Market. Computational Intelligence and Neuroscience 2017: 4315419.
- Lash and Kang Zhao. (2016) Early predictions of movie success: The who, what, and when of profitability. Journal of Management Information Systems, 33(3):874–903.
- Lash, M. T., & Zhao, K. (2016). Early predictions of movie success: The who, what, and when of profitability. Journal of Management Information Systems, 33(3), 874-903.
- Lash, Michael T., and Kang Zhao. (2016). Early Predictions of Movie Success: The Who, What, and When of Profitability. Journal of Management Information Systems 33: 874–903. [CrossRef]
- Lee, Kyung Jae, and Woojin Chang. (2009). Bayesian Belief Network for Box Office Performance: A Case Study of Korean Movies. Expert Systems with Applications 36: 280–91
- Litman, B. R. 1983. Predicting success of theatrical movies: An empirical study. J. Popular Culture 16 159–175.
- Litman, B. R. and Kohl, L. S. (1989). Predicting Financial Success of Motion Pictures: The '80s Experience. Journal of Media Economics, 2(2), 35–50
- Litman, B. R., H. Ahn. (1998). Predicting financial success of motion pictures. B. R. Litman, ed. The Motion Picture Mega-Industry.
- McKenzie, J. (2009), 'Revealed Word of Mouth and Adaptive Supply: Survival of Motion Pictures at the Australian Box Office', Journal of Cultural Economics, forthcoming.
- McKenzie, J. (2012). The economics of movies. A literature survey. Journal of Economic Surveys, 26(1), 42-70.
- MPA. (2019). Theme report 2019
- Neelamegham, R. & Chinatagunta, P. (1999). 'A Bayesian model to forecast new product performance in domestic and international markets', Marketing Science, 18(2): 115-136.
- Nelson, R., Donahue, M., Waldman, D. & Wheaton, C. (2001). 'What's an Oscar worth?' Economic Inquiry, 39(1): 1–6.
- Prag, J. and 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), 217–235.
- Ravid, A., (1999), "Information, Blockbusters, and Stars: A Study of the Film Industry," Journal of Business, 72, 463–492.
- Reinstein, D. A. & Snyder, C.M. (2000). 'The influence of expert reviews on consumer demand for experience goods: A case study of movie critics'. Working Paper, University of California-Berkeley and George Washington University.
- Ruus R., Sharma R. (2020) Predicting Movies' Box Office Result - A Large Scale Study Across Hollywood and Bollywood
- Sawhney, M. and Eliashberg, J. (1996), 'A Parsimonious Model for Forecasting Gross Box Office Revenues of Motion Pictures', Marketing Science, 15(2), 113-131.
- Sharda and D. Delen. (2006). Predicting box-office success of motion pictures with neural networks. Expert Systems with Applications.
- Sharda, R. and Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. Expert Systems with Applications, 30(2), 243–254
- Simonoff, J. S., I. R. Sparrow. (2000). Predicting movie grosses: Winners and losers, blockbusters and sleepers. Chance 13(3) 15–24.
- Simonton, D. K. (2009). Cinematic success criteria and their predictors: The art and business of the film industry. Psychology & marketing, 26(5), 400-420.
- Smit, E., & Pangarker, N. A. (2013). The determinants of box office performance in the film industry revisited. South African Journal of Business Management, 44(3), 47-58.
- Sochay, S. (1994). Predicting the Performance of Motion Pictures. Journal of Media Economics, 7(4), 1–20
- Sood, S., X. Dreze. (2004). Brand extensions of hedonic goods: Movie sequel evaluations. J. Consumer Res.
- Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. Journal of personality and social psychology, 94(4), 672.
- Terry, N., Butler, M. & De'Armond, D. (2005). 'The determinants of domestic box office performance in the motion picture industry', Southwestern Economic Review, 32: 137-148.
- Vany and W. D. Walls. (1999) Uncertainty in the Movie Industry : Does Star Power Reduce the Terror of the Box Office ? Journal of Cultural Economics, 23:285–318.
- Vogel, H. L. (2014). Entertainment industry economics, 9thed., New York: Cambridge UniversityPress.
- Wallace, W. T., A. Seigerman, M. B. Holbrook. (1993). The role of actors and actresses in the success of films: How much is a movie star worth? J. Cultural Econom. 17(1) 1–27.
- Walls, D. (2005), 'Modelling Heavy Tails and Skewness in Film Returns', Applied Financial Economics, 15(17), 1181-1188.
- Zufryden, F. S. (1996). Linking advertising to box office performance of new film releases: A marketing planning model. J. Advertising Res. 36(4) 29–41.

X. Appendix

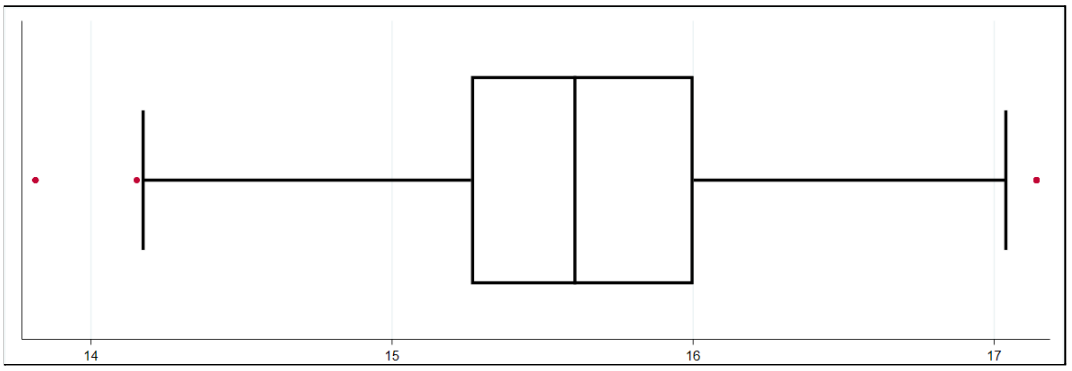
Appendix I: Box Plots

Figure A1: Box plot of the variable BUDGET, before log and winsor (in euros)



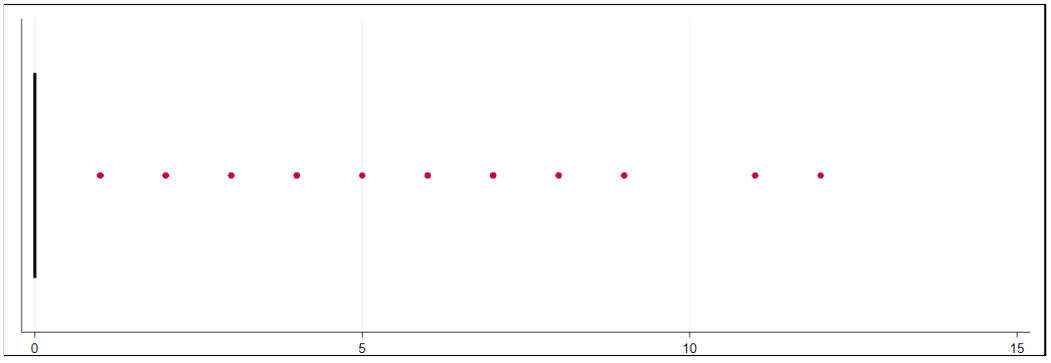
*Note: S*ource: Author rendering of data from CBO Box-office, JP's Box-office

Figure A2: Box plot of the variable BUDGET, after log and winsor



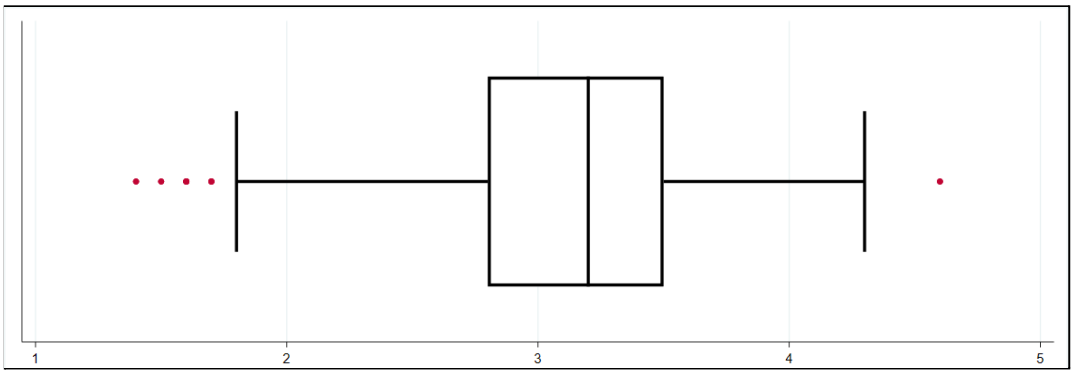
*Note: S*ource: Author rendering of data from CBO Box-office, JP's Box-office

Figure A3: Box plot of the variable AWARDS (in quantity of nominations)



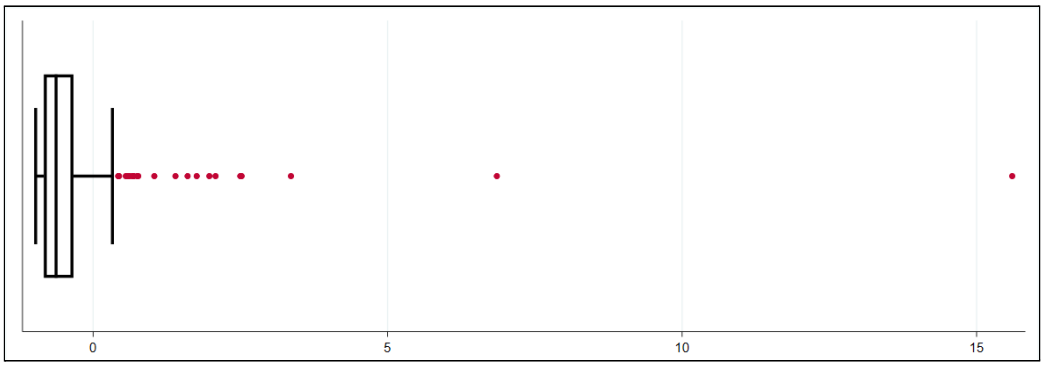
*Note: S*ource: Author rendering of data from Wikipedia

Figure A4: Box plot of the variable CRITIC (in rating from 0 to 5)



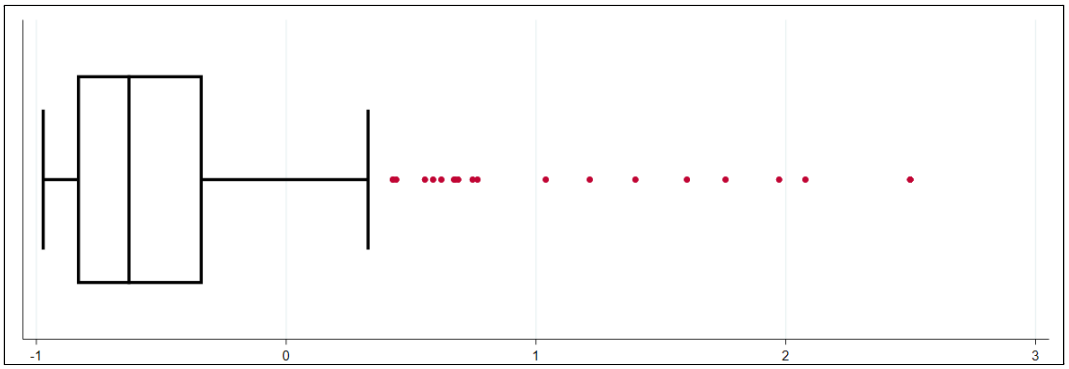
*Note: S*ource: Author rendering of data from Allociné

Figure A5: Box plot of the variable ROI



*Note: S*ource: Author rendering of data from CBO Box-office, JP's Box-office

Figure A6: Box plot of the dependent variable ROI, after correction in data and winsor



*Note: S*ource: Author rendering of data from CBO Box-office, JP's Box-office

Appendix 2: Stepwise regression process

Table A1: t-values of the coefficients of the different variables in each regression in the stepwise regression process

variable \ regression	BUDGET	ACTION	DRAMA	AWARDS	CRITIC	MAJOR	HOLIDAY	SEQUEL	BOOK	COMIC	MOVIE	PLAY	TRUE
#1	2.00	1.17	-1,49	**4.63**	1.64	3.63	0.78	3.06	-1,44	-0,11	-0,47	-0,25	0.93
#2	1.80	1.45	-2,86	X	-0,56	**3.48**	0.51	3.32	-2,14	0.26	-0,32	-0.17	0.74
#3	0.37	1.07	-2.48	X	-0.04	X	0.71	**3.34**	-2.40	-0.09	-0.53	-0.41	0.36
#4	-0.33	0.99	**-2.19**	X	1.00	X	0.77	X	-2.08	-0.37	-0.62	-0.29	0.56
#5	-0.55	0.84	X	X	1.63	X	0.53	X	**-1.74**	-0.55	-0.60	-0.31	0.94
#6	-0.34	0.73	X	X	**1.61**	X	0.38	X	X	-0.67	-0.55	-0.43	0.89
#7	0.05	0.78	X	X	X	X	0.52	X	X	-0.68	-0.47	-0.32	**0.81**
#8	-0.07	0.57	X	X	X	X	0.53	X	X	**-0.65**	-0.52	-0.29	X
#9	0.04	**0.67**	X	X	X	X	0.56	X	X	X	-0.55	-0.31	X
#10	-0.05	X	X	X	X	X	**0.57**	X	X	X	-0.51	-0.29	X
#11	-0.08	X	X	X	X	X	X	X	X	X	**-0.50**	-0.28	X
#12	-0.04	X	X	X	X	X	X	X	X	X	X	**-0.29**	X
#13	-0.04	X	X	X	X	X	X	X	X	X	X	X	X

Note: Source: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A2: Summary of the regression #1 of the stepwise regression

Source	SS	df	MS		
Model	7.89717661	1	7.89717661	Prob > F	0.0000
Residual	99.4045188	270	.368164884	R-squared	0.0736
Total	107.301695	271	.395947215	Adj R-squared	0.0702
				Root MSE	.60677

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0820295	.0177115	4.63	0.000	.0471593	.1168997
_cons	-.5116376	.0393411	-13.01	0.000	-.589092	-.4341831

Note: Source: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A3: Summary of the regression #2 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(2, 269)	=	17.22
Model	12.1784109	2	6.08920544	Prob > F	=	0.0000
Residual	95.1232845	269	.353618158	R-squared	=	0.1135
				Adj R-squared	=	0.1069
Total	107.301695	271	.395947215	Root MSE	=	.59466

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0784439	.0173886	4.51	0.000	.0442087	.112679
MAJOR	.2612115	.0750715	3.48	0.001	.1134091	.4090139
_cons	-.6038898	.0467923	-12.91	0.000	-.6960154	-.5117642

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A4: Summary of the regression #3 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(3, 268)	=	15.63
Model	15.9761622	3	5.32538739	Prob > F	=	0.0000
Residual	91.3255332	268	.340766915	R-squared	=	0.1489
				Adj R-squared	=	0.1394
Total	107.301695	271	.395947215	Root MSE	=	.58375

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.080183	.0170777	4.70	0.000	.0465595	.1138064
MAJOR	.2573787	.0737037	3.49	0.001	.1122669	.4024906
SEQUEL	.4755958	.1424635	3.34	0.001	.1951059	.7560857
_cons	-.6353363	.04689	-13.55	0.000	-.727656	-.5430167

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A5: Summary of the regression #4 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(4, 267)	=	13.09
Model	17.5901567	4	4.39753918	Prob > F	=	0.0000
Residual	89.7115387	267	.335998272	R-squared	=	0.1639
				Adj R-squared	=	0.1514
Total	107.301695	271	.395947215	Root MSE	=	.57965

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0902715	.0175714	5.14	0.000	.0556754	.1248677
MAJOR	.2369684	.0737763	3.21	0.001	.0917112	.3822257
SEQUEL	.4442291	.1421852	3.12	0.002	.1642823	.724176
DRAMA	-.1947015	.0888356	-2.19	0.029	-.3696088	-.0197942
_cons	-.5908203	.0507981	-11.63	0.000	-.6908361	-.4908045

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A6: Summary of the regression #5 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(5, 266)	=	11.15
Model	18.5978641	5	3.71957282	Prob > F	=	0.0000
Residual	88.7038313	266	.33347305	R-squared	=	0.1733
				Adj R-squared	=	0.1578
Total	107.301695	271	.395947215	Root MSE	=	.57747

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0926874	.0175603	5.28	0.000	.0581125	.1272623
MAJOR	.2475635	.0737508	3.36	0.001	.1023539	.3927731
SEQUEL	.4209121	.1422836	2.96	0.003	.1407668	.7010574
DRAMA	-.1680835	.089816	-1.87	0.062	-.3449242	.0087571
BOOK	-.1755486	.1009858	-1.74	0.083	-.3743819	.0232847
_cons	-.5744445	.0514762	-11.16	0.000	-.6757971	-.4730919

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A7: Summary of the regression #6 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(6, 265)	=	9.78
Model	19.4521043	6	3.24201739	Prob > F	=	0.0000
Residual	87.8495911	265	.331507891	R-squared	=	0.1813
				Adj R-squared	=	0.1627
Total	107.301695	271	.395947215	Root MSE	=	.57577

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0787596	.0195404	4.03	0.000	.0402855	.1172338
MAJOR	.2621981	.0740962	3.54	0.000	.116306	.4080902
SEQUEL	.4880905	.1479076	3.30	0.001	.1968668	.7793141
DRAMA	-.2064645	.0926879	-2.23	0.027	-.3889629	-.0239661
BOOK	-.1726706	.1007038	-1.71	0.088	-.370952	.0256108
CRITIC	.1135668	.070747	1.61	0.110	-.025731	.2528646
_cons	-.9175756	.2198307	-4.17	0.000	-1.350413	-.4847386

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A8: Summary of the regression #7 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(7, 264)	=	8.46
Model	19.668545	7	2.80979214	Prob > F	=	0.0000
Residual	87.6331504	264	.331943752	R-squared	=	0.1833
				Adj R-squared	=	0.1616
Total	107.301695	271	.395947215	Root MSE	=	.57615

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0791716	.0195599	4.05	0.000	.0406583	.1176848
MAJOR	.2535017	.0749229	3.38	0.001	.1059792	.4010243
SEQUEL	.4910846	.1480513	3.32	0.001	.1995731	.7825961
DRAMA	-.2178594	.0938162	-2.32	0.021	-.4025826	-.0331363
BOOK	-.1702733	.1008137	-1.69	0.092	-.3687746	.0282279
CRITIC	.1105238	.0708938	1.56	0.120	-.0290654	.250113
TRUE	.1468382	.1818452	0.81	0.420	-.2112133	.5048897
_cons	-.9092776	.2202151	-4.13	0.000	-1.342879	-.4756762

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A9: Summary of the regression #8 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(8, 263)	=	7.44
Model	19.809864	8	2.476233	Prob > F	=	0.0000
Residual	87.4918314	263	.33266856	R-squared	=	0.1846
				Adj R-squared	=	0.1598
Total	107.301695	271	.395947215	Root MSE	=	.57677

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0784545	.0196121	4.00	0.000	.0398377	.1170712
MAJOR	.2585342	.0754011	3.43	0.001	.1100677	.4070008
SEQUEL	.4976567	.1485554	3.35	0.001	.2051473	.790166
DRAMA	-.2214458	.0940796	-2.35	0.019	-.4066908	-.0362007
BOOK	-.1747503	.1011572	-1.73	0.085	-.3739313	.0244308
CRITIC	.1110087	.070975	1.56	0.119	-.0287429	.2507603
TRUE	.1422378	.1821804	0.78	0.436	-.21648	.5009556
COMIC	-.1130368	.1734304	-0.65	0.515	-.4545255	.228452
_cons	-.9058455	.2205182	-4.11	0.000	-1.340051	-.4716396

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A10: Summary of the regression #9 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(9, 262)	=	6.65
Model	19.9615497	9	2.21794997	Prob > F	=	0.0000
Residual	87.3401457	262	.333359335	R-squared	=	0.1860
				Adj R-squared	=	0.1581
Total	107.301695	271	.395947215	Root MSE	=	.57737

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0784438	.0196325	4.00	0.000	.0397863	.1171013
MAJOR	.2556623	.0755993	3.38	0.001	.1068028	.4045218
SEQUEL	.4971043	.1487118	3.34	0.001	.2042818	.7899267
DRAMA	-.2157708	.0945522	-2.28	0.023	-.4019498	-.0295918
BOOK	-.171303	.101391	-1.69	0.092	-.370948	.028342
CRITIC	.1133003	.0711298	1.59	0.112	-.0267586	.2533593
TRUE	.1040971	.1909336	0.55	0.586	-.2718625	.4800568
COMIC	-.1312983	.1757084	-0.75	0.456	-.4772787	.2146821
ACTION	.1415093	.2097824	0.67	0.501	-.2715648	.5545835
_cons	-.9159709	.2212568	-4.14	0.000	-1.351639	-.480303

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A11: Summary of the regression #10 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(10, 261)	=	6.01
Model	20.0712261	10	2.00712261	Prob > F	=	0.0000
Residual	87.2304693	261	.334216357	R-squared	=	0.1871
				Adj R-squared	=	0.1559
Total	107.301695	271	.395947215	Root MSE	=	.57811

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0769444	.0198312	3.88	0.000	.037895	.1159939
MAJOR	.2586935	.0758811	3.41	0.001	.1092764	.4081106
SEQUEL	.5023702	.1491863	3.37	0.001	.2086082	.7961322
DRAMA	-.212202	.0948784	-2.24	0.026	-.3990267	-.0253774
BOOK	-.1658503	.1019665	-1.63	0.105	-.3666321	.0349314
CRITIC	.117039	.0715196	1.64	0.103	-.0237899	.2578679
TRUE	.1042766	.1911791	0.55	0.586	-.2721732	.4807265
COMIC	-.1356795	.1761003	-0.77	0.442	-.4824377	.2110787
ACTION	.1442993	.2101084	0.69	0.493	-.2694239	.5580226
HOLIDAY	.047278	.0825308	0.57	0.567	-.115233	.209789
_cons	-.9411773	.2258685	-4.17	0.000	-1.385934	-.4964208

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A12: Summary of the regression #11 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(11, 260)	=	5.47
Model	20.1546033	11	1.83223666	Prob > F	=	0.0000
Residual	87.1470921	260	.335181123	R-squared	=	0.1878
				Adj R-squared	=	0.1535
Total	107.301695	271	.395947215	Root MSE	=	.57895

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0768229	.0198613	3.87	0.000	.0377135	.1159323
MAJOR	.2604359	.0760708	3.42	0.001	.1106426	.4102293
SEQUEL	.503529	.1494196	3.37	0.001	.2093025	.7977556
DRAMA	-.2122052	.0950153	-2.23	0.026	-.3993026	-.0251077
BOOK	-.1646604	.1021415	-1.61	0.108	-.3657902	.0364695
CRITIC	.1148112	.0717619	1.60	0.111	-.0264973	.2561197
TRUE	.1112203	.1919604	0.58	0.563	-.2667746	.4892153
COMIC	-.1390537	.176484	-0.79	0.431	-.4865737	.2084663
ACTION	.1380944	.2107789	0.66	0.513	-.2769567	.5531454
HOLIDAY	.0465669	.0826622	0.56	0.574	-.1162056	.2093394
MOVIE	-.0939257	.1883219	-0.50	0.618	-.464756	.2769045
_cons	-.9313552	.22705	-4.10	0.000	-1.378446	-.4842643

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A13: Summary of the regression #12 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(12, 259)	=	5.00
Model	20.1838439	12	1.68198699	Prob > F	=	0.0000
Residual	87.1178515	259	.336362361	R-squared	=	0.1881
				Adj R-squared	=	0.1505
Total	107.301695	271	.395947215	Root MSE	=	.57997

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0770892	.0199167	3.87	0.000	.0378699	.1163085
MAJOR	.262042	.0763992	3.43	0.001	.1115994	.4124847
SEQUEL	.5008091	.1499666	3.34	0.001	.2055	.7961183
DRAMA	-.2115647	.0952073	-2.22	0.027	-.3990438	-.0240857
BOOK	-.1669705	.1026208	-1.63	0.105	-.3690478	.0351069
CRITIC	.1132432	.0720847	1.57	0.117	-.0287035	.2551899
TRUE	.109679	.1923694	0.57	0.569	-.2691282	.4884862
COMIC	-.1408655	.1769015	-0.80	0.427	-.4892138	.2074829
ACTION	.136433	.2112251	0.65	0.519	-.2795043	.5523703
HOLIDAY	.0458888	.0828396	0.55	0.580	-.1172361	.2090137
MOVIE	-.0957655	.1887566	-0.51	0.612	-.4674584	.2759275
PLAY	-.0780043	.264563	-0.29	0.768	-.5989727	.4429641
_cons	-.925029	.2284595	-4.05	0.000	-1.374904	-.4751545

*Note: S*ource: Author calculations from datasets CBO Box-office, JP's Box-office, Allociné, Wikipedia

Table A14: Summary of the regression #13 of the stepwise regression

Source	SS	df	MS	Number of obs	=	272
				F(13, 258)	=	4.60
Model	20.1843324	13	1.55264095	Prob > F	=	0.0000
Residual	87.117363	258	.337664198	R-squared	=	0.1881
				Adj R-squared	=	0.1472
Total	107.301695	271	.395947215	Root MSE	=	.58109

winROI	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
AWARDS	.0772222	.0202597	3.81	0.000	.0373269	.1171176
MAJOR	.2632145	.0825206	3.19	0.002	.1007147	.4257142
SEQUEL	.5016369	.1518244	3.30	0.001	.2026641	.8006097
DRAMA	-.2117477	.0955126	-2.22	0.027	-.3998313	-.0236641
BOOK	-.1663753	.1040031	-1.60	0.111	-.3711783	.0384277
CRITIC	.1125568	.0744442	1.51	0.132	-.0340387	.2591524
TRUE	.1104466	.193795	0.57	0.569	-.2711749	.492068
COMIC	-.1398853	.1791072	-0.78	0.436	-.4925834	.2128128
ACTION	.1375156	.2135391	0.64	0.520	-.2829858	.5580171
HOLIDAY	.0460913	.0831703	0.55	0.580	-.1176878	.2098704
MOVIE	-.0951604	.1897894	-0.50	0.617	-.468894	.2785732
PLAY	-.0778461	.2651071	-0.29	0.769	-.5998954	.4442033
winBUDGET	-.0026227	.0689556	-0.04	0.970	-.1384103	.1331648
_cons	-.8826833	1.13663	-0.78	0.438	-3.120936	1.355569

*Note: S*ource: Author calculations

Appendix 3: Pangarker and Smit (2013)

Table A15: Table of the significant results in Pangarker and Smit's study (2013). Directly copied and titled "Determinants of revenue (2009-2010)"

	Coefficients	t –value	P-value
Intercept	1,077	1,023	0,307
LN Production Cost	0,930	14,682	0,000
Major distributor	0,745	5,240	0,000
Award nomination	0,209	5,974	0,000
Sequel to success	0,958	4,603	0,000
R^2	0,643		
F-value	127,711		
Significance level	0,000		

*Note: S*ource: Pangarker and Smit (2013)

Table A16: Correlation matrix in Pangarker and Smit's study (2013). Directly copied and titled "Correlation matrix of independent and dependent variables"

	Ln Revenue	Ln Production Cost	Action	Drama	Major	Holiday	Award	Critic	Sequel
Ln (Revenue)	1,000								
Ln (Production cost)	0,735*	1,000							
Action	0,363*	0,490*	1,000						
Drama	-0,230*	-0,285*	-0,349*	1,000					
Major	0,485*	0,439*	0,194*	-0,160*	1,000				
Holiday	0,132*	0,104	0,041	-0,094	0,152*	1,000			
Award	0,231*	0,049	0,043	0,152*	-0,031	0,087	1,000		
Critic	-0,023	-0,204*	0,010	0,293*	-0,084	0,019	0,470*	1,000	
Sequel	0,337*	0,242	0,182*	-0,111	0,143*	0,031	-0,023	-0,087	1,000

*Significant at the 0.05 level

*Note: S*ource: Pangarker and Smit (2013)