“Sequel Mania”: Why Hollywood is Getting Increasingly Interested in Sequels？

By

Tao Zhu

Dr. Sue Mialon
Adviser

Department of Economics

Dr. Sue Mialon
Adviser

Dr. Jong Kim
Committee Member

Dr. Ryan Cook
Committee Member

2016

“Sequel Mania”: Why Hollywood is Getting Increasingly Interested in Sequels？

By

Tao Zhu

Dr. Sue Mialon
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2016

Abstract

"Sequel Mania": Why Hollywood is Getting Increasingly Interested in Sequels？

By Tao Zhu

The Paper aims to understand why Hollywood studios choose to make more sequels in recent years. I examine the box office performance of sequels in both nominal terms (revenue) and real terms (moviegoers) based on 9 variables and the interaction terms. A total number of 396 sequels, ranging in a fifty-year period, are analyzed under two OLS regression models. The main findings include that a return of two times of the budget input can be expected from making sequels. The performance of previous films in the movie franchise is positively related with the performance of the sequel. The critics rating, as a proxy of the quality of the sequel, plays a significant role in affecting audience's decision, which indicates that the audience are rational in making decisions. One noteworthy finding is that sequels made after two years of its predecessor cannot attract as many viewers as those made within two years. This is attributed to the momentum of audience's memories. Finally, my work reveals that a sequel rated R can negatively effect the impact of budget on box office.

"Sequel Mania": Why Hollywood is Getting Increasingly Interested in Sequels?

By

Tao Zhu

Dr. Sue Mialon
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2016

Acknowledgements

I would like to show my appreciation to my honors committee, as well as others who help me with this research paper. In particular, I want to express my gratitude to my main advisor, Prof. Sue Mialon, for sharing her time and providing countless valuable suggestions throughout the past year. I also would like to thank Prof. Jong Kim and Prof. Ryan Cook for their useful guidance. The suggestions and help of Prof. Matthew Bernstein and Prof. Andrew Francis-Tan make a noticeable difference in the production of this thesis. Finally, it's important to recognize Dr. James Steffen and Dr. Rob O'Reilly for their guidance on the building of my dataset.

Table of Contents

I. Introduction

Back in the golden age of the film industry in 1950s, movies were the third-largest retail business in the United States, only after grocery stores and cars[1]. Film at that time was one of the most profitable businesses. But in the next 20 years, the advent of television industry and the policy of “Trust-busting” broke up the studios and scattered audience attention. While the global audience of movies increase dramatically, the film industry is on the wane from 1990s. As a result, the annual number of tickets a typical American buys reduces from twenty to four within twenty years. Accordingly, studios make fewer movies than they used to, and they have to spend more money marketing them since they've lost their guaranteed weekly audience.

Now that the studios are making fewer, more expensive films, there is much more risk riding on each project than it used to. The reluctance of studios to take risks on original concepts follows a series of disastrous results. 19 out of the 20 biggest box office bombs, after adjusting the nominal losses for inflation in order to make a fair comparison, are original films. It is believed that Hollywood has been mitigating that risk by turning to safer projects like sequels and adaptations that already have a built-in audience. Sequels tend to cost less, can be made faster, and don’t require nearly as much promotion as an original film. This paper aims to provide some empirical evidence of these intuitive speculations.

The first question we need to answer, though seems quite easy at first glance, is that what exactly is a sequel? By definition, sequels movies are those that follow the story of, or expand upon previous movies in the same series. But rare cases like *Avengers* (2012)

---

[1] Epstein, Edward Jay. "The big picture." *The New Logic of Money and Power in* (2005).

are super-sequels with three current movie franchises funneling into a single sequel. For the purposes of my study, I treat them as a sequel of all three predecessors. Likewise, reboots, which means a new film discarding all continuity in an established series in order to recreate its characters, timeline and backstory from the beginning, are hard to be categorized. For example, *Batman Begins* (2005) is the first of the Christopher Nolan's *Dark Knight* trilogy, but it's certainly not an original film. In my study, I treat all these reboots as sequels in that the stories are well known by the audiences due to their predecessors.

Figure 1 shows the prevalence of sequels in top-ten grossing films over the past decade, and shows a slightly upward trend of sequels' dominance in box office performance. 64% of the past two-decade's annual top-ten grossing films were sequels, and those sequels on average brought home $33 million more than the originals. The high point is the year 2011, when 9 out of 10 films are sequels, and 19 of the top 25. But even when sequels comprise less of the market, they are worth, on average, $23 million more than their original counterparts[2]. Multiple reports and articles on the film industry have also concluded that Hollywood is entering the time of "Sequel Mania"[3]. By one count, a record of 37 sequels, not including reboots of the previous hits, are expected to hit screens this year (2016) as Hollywood executives steer clear of original and potentially costly ideas.

On the other hand, Revenue isn't the same as profit. For instance, $1 revenue at the box office for *Avengers* (2012) was more profitable to the production Studio, Marvel, than $1 at the box office for *The Amazing Spider-Man* (2012) to the production studio,

2 Jake Lehrhoff, "Hollywood's Sequel Problem", Fortune.com
3 Originally used by Forbes in 2011.

FOX, due to the difference in copyright expenses. After ticket grosses are shared with theatres, often, studios pay a predetermined fraction of a movie's profits to its stars, in addition to the costs of producing, marketing and releasing a film. Although the profit seems a better measurement of the performance of films, in my paper I use the revenue of the sequels in both nominal and real terms because the formula of profit calculation differs in a case-by-case study. Also, it's hard to get the internal data regarding the marketing expense and distribution expense.

This paper empirically estimates how the expected revenue and moviegoers of sequels are influenced by different factors, including exterior factors such as the performance of previous films in the franchise and interior factors such as its own quality, on the condition that the previous film performed well enough for the studio to license a sequel.

I mainly follow models provided by the book *Hollywood Economics* (2004), where De Vany (2004) compiles his groundbreaking and standard-setting research in the economic study of nearly every side of the U.S. motion-pictures industry. My work presents three main contributions relative to prior literatures. First, I create two models in nominal and real terms. With different ways of measurement, I want to locate the significant factors in two respective models. Second, I control the effect of stars for films in the same series by assuming it is kept constant throughout the franchise. This makes sense in that in order to reduce the cost of sequels, studios tend to keep the original staff members and main characters in the story. Due to the nature of artistic production, many artistic outputs, including films, are influenced by too many vague and subtle factors that are hard to be quantified and so never appear in the dataset, thus potentially biasing the results. My dataset would eliminate such problem. Third, most previous works focus on the qualitative rather than the quantitative part of question. My work fills in the hole by adopting econometric approach, the OLS regression, to find the causality between each possible variable.

I employ public-use data from *The Numbers*, which is the largest online provider of motion-pictures industry data and research services. Other sources include Rotten Tomatoes and MPAA's Annual Theatrical Market Statistics.

My main findings include: 1) the performance of previous films has a positive effect on the sequels; budget is significant in predicting nominal return and number of moviegoers. The nominal return of a sequel is two times of its budget; 2) whether the leading stars returns in the sequels wouldn't significantly affect the performance of 3) sequels produced within two years of the release of its predecessor tend to perform better than those produced later; 4) R-rated Films negatively affect the return of certain budget.

The rest of the paper is organized as follows. Section II provides a brief overview of the literatures regarding the motion-pictures industry itself form the studio's perspective, as well a review of the researches on sequels. Section III shows the dataset I create by combining multiple sources, as well as the variables I use in my model. Section IV summarizes two empirical methods, tests on interaction terms, as well as the results. Section V discusses the limitations and suggestions for further studies. Conclusion is given in Section VI.

II. Literature Review

The most famous hypothesis in the motion-pictures industry is called the “nobody knows anything” hypothesis, which is described and documented by David Walls and Arthur De Vany in a series of articles and his famous book *Hollywood Economics* (2004). This hypothesis argues that “prediction is impossible; one can only say that the expected revenue of a movie is X plus or minus infinity.” (De vany, 2004) For example, if a producer is choosing to invest some money between two potential film projects, one original story A and one sequel B, which is the third film of a successful trilogy, the producer should not use the past to help choose what to do or to predict future revenues. Even though the previous films in the franchise are successful and the characters are well known by the audiences, according to the “nobody knows anything” hypothesis, these facts have little impact on the likelihood that a this movie will be a hit and, therefore, those facts should not impact studio’s decision to invest. The producer should simply assess both projects and decide which to invest in by objective evaluation of quality.

The “Nobody knows anything” hypothesis appears to negate what could be considered the force of artistic decisions and the film producer’s role. In other words, any formulas created by studios with a combination of stars, budget or releasing date are unrelated with the box office performance of movies.

Caves (2000) points out in his book that there may exist some clues as to how we might challenge the “nobody knows anything” hypothesis. Caves writes of the toy industry, where toy developers may not know what makes a toy a hit but they know what fails. This process then, suggests that just because what remains is unpredictable, does not mean that the decisions producers make are meaningless, nor that what they do, using formula and past experience with other films to predict future financial success, is totally futile. Perhaps they are eliminating failures, failures that do not show up in the traditional industry standard data. And that leads to the hypothesis of “some people know some things”.

Research has shown that the brand serves as the major power in anticipating the sales of new products. Because the brand has reputation among the customers, thus reducing the risk of new products. Accordingly, some film series can be thought of as brand as well. The Disneyland and Universal Studio are examples of amusement parks that features strong film brands. Likewise, people will pay for the sequels because of the excellent quality of the original film as well as their interest in a complete story. Sood (2006) examines movie sequels as brand extensions of experiential goods and finds the consistency of storyline will affect opening-week box-office performance because the moviegoers who purchase a ticket on the opening week have little reference of the sequel’s quality. They watch the sequel because they are interested in how the story is going on. In contrast, moviegoers who purchase a ticket on subsequent weeks will have a chance of being exposed to reviews and, thus, might have an opinion about the movie that is based on the movie content.

In terms of the performance of sequels, a recent study by Suman (2007) uses a random sample of 167 films released between 1991 and 1993, and finds that the box office revenue of sequels are considerably different from that of the original films. Sequels do better than their contemporaneous non-sequels, more so when they are released sooner after their parents, and when more intervening sequels come before them. Ravid (1999) analyzes the subsets by setting sequel as a dummy variable and finds that sequels are not necessarily more successful than the original movies; in fact, they generally follow the box office performance of the original film, which means they are positively related with its predecessor's performance. They do, however, perform much better than the median original movies.

In sum, the two concepts, decisions and choices in the arts and creative industries are as much about eliminating failures as they are about insuring hits and the idea that some qualities in sequels make it more predictable than original films, have led to suggest contrary evidence to the "nobody knows anything" hypothesis, both framing the development of the structure and dataset for this paper. In light of that, I try to combine a more inclusive dataset with more available variables. Suman (2007) and Ravid (1999) provide empirical models that treat sequel as a dummy variable. With more available data and more variables, my paper shifts the focus entirely to sequels. Also, sequels not shown in theater and those with missing data, both of which are included in Ravid's (1999) work, are eliminated from my dataset on the condition that studios have taken into account the elimination of possible failures.

III. Dataset

I use the dataset mainly from *The Numbers*, which serves over 1,000 clients in the film industry. The dataset consists of the budget, revenue and other data related to theatrical release. Some historical data about budget and opening weekend revenue are missing for old films, thus could potentially biasing the result. As a result, I exclude some old sequels from my dataset to prevent such problem. At the same time, Rotten Tomatoes, which has a group of verified critics, provides the data of "critics rating" and "audience rating" that calculates the percentage of users who rates the film positively.

Critic reviews generally use 4-star ratings and are often qualitative. The users' score is more detailed, because users rate the movie on a scale of 0–10. The final numbers of both critics rating and audience rating are on a scale of 0-100 calculated by Rotten Tomatoes itself. Finally, I get the average ticket price of each year from MPAA’s annual *Theatrical Market Statistics*. By dividing the nominal revenue by the corresponding ticket price of that year, I calculate an approximate number of moviegoers, which is the real term of a film’s performance regardless of the inflation.

Basically the two models with either nominal term or real term measure the same thing, because the estimate tickets are calculated by dividing the total revenue by average ticket price. But through different coefficient, the two functions provide two different measures of the performance of sequels; how much revenue or how many tickets to be expected from a certain amount of input in each independent variable.

The data for this study consists of 152 Hollywood movie series, 396 sequels in total, from a fifty-year period. Table 1 presents summary statistics for all independent variables used in the analysis.

The two dependent variables I want to measure in the models are *Revenue* and *Viewer. Revenue* represents the box office performance of the sequel in nominal dollar value. It is the most commonly used data regarding a film’s performance. *Viewer* is the number of moviegoers, which is calculated by dividing the revenue of each film by the average ticket price of that year. Since the prices vary each year and differ by states, the nominal value of revenue may not be a good measure for the success of movies because of inflation. Thus, we use a proxy that measures the success in real terms, the number of moviegoers.

Budget denotes the budget for the sequel. Budget is not the actual expense on the film, but serves as a proxy for the expense of the film as well as its expected value from the perspective of studios. *PreAveRev* is calculated by averaging the revenue of all the previous films in the same movie series, while *PreAveTic* is calculated by averaging the number of tickets sold (moviegoers) of all the previous films. *Critics* refers to the critics rating of each sequel, thus can be viewed as a measurement of the overall quality of the sequels. On the other hand, *Audience* denotes the audiences' rating of the previous film, as an indicator of the expectation of the sequel. *ROMA (Return of Main Artists)* is a dummy variable equal to 1 if any of the three leading stars return in the sequel, and thus is an indicator of whether there's any change in the main cast. Here the leading star refers to the three main artists whose names appear on each film's homepage on IMDB. *Rfilm* is a dummy variable equal to 1 if the sequel is a R-rated film, and 0 otherwise. *Source* is a dummy variable measures whether the movie franchise is based on original screenplay, or is adaptation of novels, comics or TV. The value of Source equals 1 if the story is an original screenplay and 0 if not. *Year* is a dummy variable equal to 1 if the sequel is produced after 2000, and 0 otherwise. I choose year 2000 because in this year the number of sequels on the list of annual top-10 grossing films jumped from 1 to 5, and gradually became the mainstream afterwards. *Timespan* is also a dummy variable equal to 1 if the time span between the release date of the sequel and its predecessor is longer than two years. Two year is basically the duration enough for combining all the time need for pre-production, production and post-production of a sequel. This dummy variable is designed to test if the studio decide when to shoot the sequel will affect the outcome.

Table 2 and Table 3 each show the pairwise correlation between different variables in nominal and real terms, which measure the relative strength of the linear relationship between them. No strong correlation is observed in two tables, which means I can disregard the possibility of multicollinearity.

IV. Empirical Strategy and Results

The main model in the following estimates the performance of sequels with regard to each independent variable discussed above:

(1) $Revenue_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveRev_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + u_i$

(2)$Viewer_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveTic_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + u_i$

While model (1) estimates the performance of sequels in nominal terms, model (2) estimates in real terms. β_i is the coefficient (multiplier) of the variable X_i. u_i is the error term, which is estimated with robust methods.

1) A comparison between nominal and real terms (Table 2)

Table 4 presents the results of the two mentioned regressions in nominal and real terms. The nominal model measures how much money the studio may expect to earn with 1 unit of input in each regressor, while the real terms measure how many audiences may watch the film with regard to the change in regressor.

The budget is significant at 1 percent level and indicates that for every \$1 input in budget, the studio can expect \$2 in revenue and around 0.25 movie tickets. De Vany

(2004) analyzes the profit and returns of film from 1982-1996 in his book, where he concludes that "The median movie lost about 3.8 million dollars and a film had to reach all the way up into the 78^{th} percentile of the gross profit distribution before it broke even in its theatrical run". In this sense, the sequels are considerably more profitable in the context of all theatrically released films.

The performance of previous films is significant both in nominal and real terms. If the previous films have higher box office revenue (or number of viewers), the expected revenue (or expected tickets sold) will increase accordingly. For the nominal term, the coefficient equals 0.542. So 1 million dollars' increase in previous average revenue will lead to 0.542 million dollars increase in the sequel. Likewise, the ticket numbers sold (with coefficient 0.346) will also increase with a better box office performance.

The critics rating and audience rating both serve as the quality and popularity of the sequels, but actually the effect of ratings varies significantly between critics and audience. The audience rating of the last film, as an indication of people's expectation for the sequel, is not significant enough to explain the performance of the current sequel. On the other hand, the critics rating, as a proxy for the sequel's quality, are available on magazines, newspapers or online after premiere, which can influence audience's decision as whether to watch the film or not. In both models, the critics rating is significant to predict a positive relation with the box office. This means that the objective quality of the film still serves as the major force for a sustainably strong performance in box office. A sequel from a big name series may perform well in the first weekend. After audience's enthusiasm decreases over the weeks, it is its reputation, namely the objective quality of the sequel that keeps the film from flopping.

In terms of whether the leading star returns in the sequel, the coefficient seems to be relatively unimportant to the overall performance of the sequel. Since around 66% stars return in the sequel, this result may be biased by other more important factors. This finding of the uncertainty of stars' influence is consistent with De Vany's conclusion in his Book (2004), where he writes, "Stars make the distribution less skewed. However, movies with stars do not stochastically dominate movies without stars in terms of gross return."

The effect of rating and source are not significant in both models. This implies that if the sequel is rated R or if it is based on original screenplay cannot effectively predict its box office performance.

The dummy variable "Timespan" is significant in real terms. In real terms, because the number of tickets is measured in unit of million, the slope coefficient of -8.47 indicates that the making sequels within two years after its predecessor can attract 8.47 million more audiences than otherwise. This result indicates that the interest and enthusiasm of the audience decrease as time passes by. If the duration between two films is too long, some audiences who have watched the predecessor won't choose to watch the sequel because their memory of the previous story is blurred. This finding is consistent with that of Suman (2007).

The coefficient of "Year" variable shows that sequels made after 2000 are expected to earn 40 million dollars more in revenue than films made before 2000. The main reason can be attributed to the increase in ticket prices since there is no significant result observed on the number of moviegoers. If the price increase is only due to inflation, this implies that the sequels made after 2000 are not necessarily more profitable than previous sequels.

2) Interaction Analysis on the Regression Model

a. ROMA

It would be interesting to see whether the impact of budget was more significant when the leading stars returns in the sequel. To see this effect, we added an interaction term to the model. One may expect that when the leading star returns in the sequel (ROMA=1), the expected return of the same budget, compared with a sequel whose ROMA is 0, tends to be higher because the risk is smaller. The equation is given as follows:

$$Revenue_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveRev_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * ROMA + u_i$$

$$Viewer_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveTic_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * ROMA + u_i$$

A new variable Budget*ROMA is added to the right side of both original equations with the coefficient β_{10}. Table 5 presents the result of the new regression model. The new interaction term isn't significant in this case, which means that there is no significant difference in having stars for the same budget power for the model.

b. Rfilm

Similarly, we can test whether a sequel rated R will have a differential impact on box office for the same amount of budget:

$$Revenue_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveRev_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * Rfilm + u_i$$

$$Viewer_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveTic_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * Rfilm + u_i$$

An interaction term called $Budget * Rfilm$ is added to the right side of two original regression models. The result is given through Tables 6. The interaction term is significant in nominal terms.

The regression outcome implies that,

$$Revenue_i = -138.2 + 2.037 Budget_i - 0.694 Budget * Rfilm + u_i$$

If the film is rated R (*Rfilm*=1), the effect of budget is 2.037-0.694=1.343, while that of a non-R-rated film (*Rfilm*=0) equals 2.037. The presence of a significant interaction indicates that for R-rated movies, the effect of budget on the revenue is smaller than non-R rated movies.

c. Source

The following equation tests whether a sequel whose script is an original screenplay will influence the effect of budget on box office:

$$Revenue_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveRev_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * Source + u_i$$

$$Viewer_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveTic_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * Source + u_i$$

A new variable $Budget * Source$ is added to the right side of both original equations with the coefficient β_{10}. Table 7 presents the result of the new regression model. The new interaction term isn't significant, thus having no effect of having original score on the effect of budget on the performance of the movie.

d. Year

The following equation tests whether a sequel produced before or after 2000 will influence the effect of budget on box office:

$$Revenue_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveRev_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * Year + u_i$$

$$Viewer_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveTic_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * Year + u_i$$

A new variable $Budget * Year$ is added to the right side of both original equations with the coefficient β_{10}. Table 8 presents the result of the new regression model. The new interaction term isn't significant in this case, which means the production year has no significant effect on the return of budget.

e. Timespan

The following equation tests whether a sequel produced within two years of the release date of its predecessor or after two years will influence the effect of budget on box office:

$$Revenue_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveRev_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * Timespan + u_i$$

$$Viewer_i = \beta_0 + \beta_1 Budget_i + \beta_2 PreAveTic_i + \beta_3 Critics_i + \beta_4 Audience_i + \beta_5 ROMA + \beta_6 Rfilm + \beta_7 Source + \beta_8 Year + \beta_9 Timespan + \beta_{10} Budget * Timespan + u_i$$

A new variable $budget * Timespan$ is added to the right side of both original equations with the coefficient β_{10}. Table 9 presents the result of the new regression model. The new interaction term isn't significant in this case due to its large standard error, which means it has no prediction power for the model.

V. Discussion and Extension

The dataset I create doesn't incorporate all the available sequels ever produced in history. First of all, some data regarding the production budget and revenue were not released by studios, thus leaving some of my dataset blank. Second, as Jansen (2005) points out, "Most previous econometric studies on the determinants of motion picture success only consider successful films." Data of those sequels with very weak attendance are missing on the raw dataset from *The Numbers*. It also excludes sequels released in film festivals, sequels that are not picked up and are never released, sequels that end up going straight to DVD, or sequel projects that are initiated but never carried out. What is so critical about these data is that they represent the very sequels or films that don't make revenues. I do observe many sequels on The Numbers are only recorded with their titles and basic information, with all the numerical data missing. If these sequels are systematically different then the ones that are released, then perhaps there is some suggestion that there are some key factors existing just because the elimination of these sequels, while whose data are only known by studios. My results have already shown that the decisions made by studios on making more sequels are somewhat effective. One possible problem is that, since some sequels don't make it to the theaters or are abandoned halfway, the studios have to make more profits through the already released sequels as well as other movies in order to cover the loss. The extension underlined is whether the current evidence of profit is enough to make up for all the errors. Thus, a more inclusive dataset that includes absolute failures is needed for more thorough studies.

Another possible extension is that more tests on interaction terms can be adopted to gain further knowledge of the model. For example, the critics rating can possibly affect the return of budget. With a higher rating, which means the sequel has a better quality, the return of the sequel can be expected to be higher.

Also, it will be interesting to record the box office revenue on a daily or weekly basis as time series data. Since my model only examine the total revenue, the change rate in revenue over time can also reveal some problems with regard to the sustainability of sequels.

VI. Conclusion

This paper, in trying to interpret the phenomenon of “Sequel Mania”, gives the reader an overview of how the several different variables affect the box office performance of sequels.

The results show that sequels are lucrative in both nominal and real terms. A return of two times of the budget input can be expected from making sequels. The performance of previous films in the movie franchise is positively related with the performance of the sequel. The critics rating plays a significant role in affecting audience’s decision. These ratings represent the objective quality of the sequels, thus showing that audiences are rational in making decisions.

In terms of dummy variables, the effect of the return of leading stars seems unclear due to its large standard error. The sequels made after 2000 are more lucrative in nominal terms. But after the revenue is adjusted for inflation, no significance is observed, thus implying that although in recent years the nominal revenue of sequels is higher, they are not necessarily more lucrative historically speaking. One noteworthy finding is that sequels made after two years of its predecessor cannot attract as many viewers as those made within two years. This is attributed to the momentum of audience's memories.

More research can be done to build on the work of this paper. One possible problem of my dataset is that we only observe the films already been released and recorded with detailed data. There also exist some sequels that don't have a chance to be shown in theaters, or recorded with missing data. Future works can build on my results to test how these "successful" sequels cover the losses in the film industry.

Also, my work suggests that adding interaction terms to a regression model can expand understanding of the relationships among the variables in the model. My work reveals that an R rating can negatively effect the impact of budget on box office. So future works can test more hypotheses with other interaction terms.

VII. References:

1. Albert, S. (1998) "Movie Stars and the distribution of financially Successful Films in the Motion Picture Industry" (1998) *Journal of Cultural Economics*, 22(4) 249-270
2. Bagella M. and L. Becchetti (1999) "The determinants of motion picture box office performance: Evidence from the movies produced in Italy." *Journal of Cultural economics*, 23(4) 237-256
3. Caves R. (2000). "Creative industries, Contracts Between Art and Commerce", Harvard University Press: Cambridge.
4. Chisholm D. (2004) "Two-part share contracts, risk, and the life cycle of stars: Some empirical results from motion picture contracts" *Journal of Cultural economics*, 22(4), 37-56.
5. Cuadrado M. and Marta Frasquet (1999) "Segmentation of Cinema Audiences: An Exploratory Study Applied to young Consumers" *Journal of Cultural economics*, 23(4) 257-267
6. Deuchert, E. Kossi Adjamah and Florian Pauly (2005) "For Oscar Glory or Oscar Money? Academy awards and Movie Success." *Journal of Cultural Economics*, Volume 29, No. 3, pp 177-190.
7. DeVany A. (2004) *Hollywood Economics*, Routledge:NewYork.
8. DeVany A. and Walls (1999), "Uncertainty in the Movies: can star power reduce the terror of the box office?" *Journal of Cultural Economics*, 23(4) 285-318.
9. Ginsburgh V. and Weyers S. (1999) "On the perceived quality of movies" *Journal of Cultural economics*, 23(4) 269-283

10. Hadida A.L. (2004) “Reputation resources, commitment and performance of film projects in the USA and Canada (1988-1997)”, University of Cambridge working paper – The Judge Iinstitute of Management
11. Hennig-Thurau, T.; Houston M., and G. J. Walsh (2003) “Determinants of motion picture box office and profitability: an interrelationship approach” Working paper No. 4, September 2003. Center of research on motion picture success and marketing – Bauhaus-University of Weimar
12. Jansen C. (2005) “The performance of German motion pictures, profits and subsidies: some empirical evidence” *Journal of Cultural economics*, 29(3) 191 - 212.
13. Ralf dewanter & Michael Westermann (2005) “Cinema demand in Germany” *Journal of Cultural economics*, 29(3) 191 - 212.
14. Ravid, S. A. (1999) “Information, blockbusters, and Stars: a study of the film industry” *The Journal of Business*, 72, 4 pp 463-492.
15. Elberse, Anita. (2007) "The power of stars: Do star actors drive the success of movies?" *Journal of Marketing* 71.4: 102-120.
16. Agrawal, J., & Kamakura, W. A. (1995). “The economic worth of celebrity endorsers: An event study analysis.” The Journal of Marketing, 56-62.
17. Epstein, Edward Jay. (2005) "Hollywood's Profits, Demystified." Slate.
18. Claycamp, Henry J., and Lucien E. Liddy. (1969) "Prediction of new product performance: An analytical approach." Journal of Marketing Research: 414-420.
19. Herbig, Paul, John Milewicz, and Jim Golden. (1994) "A model of reputation building and destruction." Journal of Business Research 31.1: 23-31.

20. Sood, Sanjay, and Xavier Drèze. (2006) "Brand extensions of experiential goods: Movie sequel evaluations." Journal of Consumer Research 33.3: 352-360.
21. Keller, Kevin Lane, and David A. Aaker. (1992) "The effects of sequential introduction of brand extensions." *Journal of marketing research*: 35-50.
22. Keller, Kevin Lane. "Conceptualizing, measuring, and managing customer-based brand equity." *the Journal of Marketing* (1993): 1-22.

VIII. Appendix

Figure 1: Prevalence of Sequels in Top-ten Grossing Films, 1996-2015

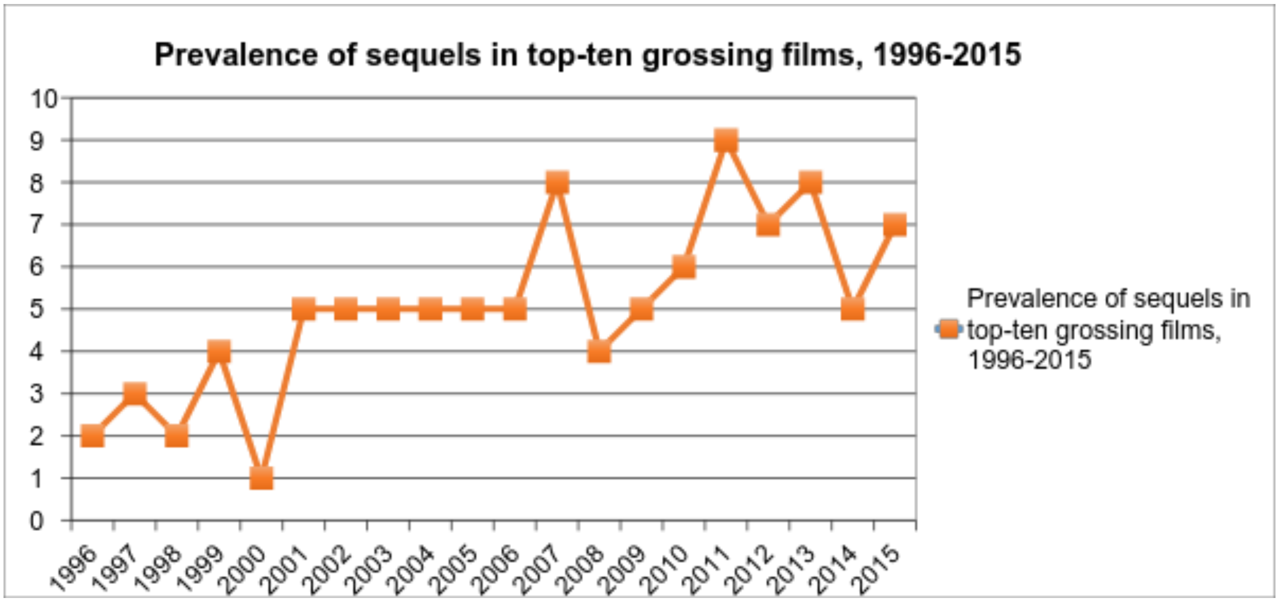

Table 1: Descriptive Statistics of All the Variables

Descriptive Statistics of All the Variables					
Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Budget	341	79.25572	67.63646	1.25	300
PreAveRev	396	261.968	246.8053	0.0020093	1519.479
PreAveTic	396	58.77403	51.69693	0.490247	352.7066
Audience	396	64.95202	19.62652	13	98
Critics	396	49.48232	26.48596	2	100
ROMA	396	0.6666667	0.4720009	0	1
Rfilm	396	0.3106061	0.4633271	0	1
Source	396	0.5782828	0.4944585	0	1
Year	396	0.6262626	0.4844072	0	1
Timespan	396	0.6893939	0.4633271	0	1

Table 2: Correlation Matrix in Nominal Terms

	preave~v	budget	Critics	Audien~s	ROMA	R	Source
preaverev	1.0000						
budget	0.6584	1.0000					
Critics	0.2442	0.2593	1.0000				
Audiences	0.2606	0.1241	0.2920	1.0000			
ROMA	0.2888	0.2136	0.2409	0.3194	1.0000		
R	-0.3287	-0.4237	-0.2293	-0.0938	-0.1968	1.0000	
Source	-0.2684	-0.3168	-0.2734	-0.1698	-0.1591	0.3853	1.0000
Year	0.3437	0.4494	-0.0100	-0.0671	-0.0037	-0.0680	-0.1524
Timespan	0.0166	0.1579	0.0102	-0.0434	-0.0463	0.0024	0.0346

	Year	Timespan
Year	1.0000	
Timespan	0.1244	1.0000

Table 3: Correlation Matrix in Real Terms

	preave~s	budget	Critics	Audien~s	ROMA	R	Source
preavetick~s	1.0000						
budget	0.3996	1.0000					
Critics	0.3120	0.2593	1.0000				
Audiences	0.2855	0.1241	0.2920	1.0000			
ROMA	0.1984	0.2136	0.2409	0.3194	1.0000		
R	-0.3121	-0.4237	-0.2293	-0.0938	-0.1968	1.0000	
Source	-0.2289	-0.3168	-0.2734	-0.1698	-0.1591	0.3853	1.0000
Year	0.0214	0.4494	-0.0100	-0.0671	-0.0037	-0.0680	-0.1524
Timespan	0.0748	0.1579	0.0102	-0.0434	-0.0463	0.0024	0.0346

	Year	Timespan
Year	1.0000	
Timespan	0.1244	1.0000

Table 4: Impact on Box Office Performance

Impact on Box Office Performance

	(1)	(2)
VARIABLES	Viewer	Revenue
Budget	0.264***	1.948***
	(0.0396)	(0.312)
PreAveTic	0.346***	
	(0.0753)	
Critics	0.365***	2.214***
	(0.0604)	(0.352)
ROMA	6.275	2.493
	(3.830)	(21.85)
Audience	0.0812	0.196
	(0.0732)	(0.547)
Rfilm	-6.353*	-15.64
	(3.481)	(17.46)
Source	4.205	18.40
	(3.427)	(19.95)
Year	-4.035	40.86**
	(4.141)	(17.05)
Timespan	-8.469***	-26.32
	(3.023)	(22.08)
PreAveRev		0.542***
		(0.0886)
Constant	-9.451	-120.7***
	(6.127)	(37.41)
Observations	341	341
R-squared	0.674	0.753

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, *p<0.1

Table 5: Interaction Effect on Box Office (ROMA)

Interaction Effect on Box Office (ROMA)

	(1)	(2)
VARIABLES	Revenue	Viewer
Budget	1.823***	0.251***
	(0.422)	(0.0557)
PreAveRev	0.541***	
	(0.0893)	
Critics	2.222***	0.365***
	(0.351)	(0.0604)
ROMA	-9.066	5.101
	(21.27)	(4.234)
Audience	0.160	0.0771
	(0.527)	(0.0740)
Rfilm	-16.63	-6.432*
	(17.80)	(3.497)
Source	17.54	4.111
	(19.88)	(3.432)
Year	40.00**	-4.109
	(16.70)	(4.137)
Timespan	-24.35	-8.267***
	(21.57)	(3.075)
Budget*ROMA	0.174	0.0175
	(0.437)	(0.0527)
PreAveTic		0.347***
		(0.0752)
Constant	-111.2***	-8.505
	(35.60)	(6.453)
Observations	341	341
R-squared	0.753	0.674

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Interaction Effect on Box Office (Rfilm)

Interaction Effect on Box Office (Rfilm)		
	(1)	(2)
VARIABLES	Revenue	Viewer
Budget	2.037***	0.262***
	(0.333)	(0.0402)
PreAveRev	0.535***	
	(0.0895)	
Critics	2.240***	0.364***
	(0.355)	(0.0601)
ROMA	6.835	6.178
	(22.14)	(3.844)
Audience	0.250	0.0803
	(0.551)	(0.0736)
Rfilm	15.77	-7.185
	(21.66)	(4.380)
Source	19.75	4.179
	(19.96)	(3.429)
Year	41.90**	-4.060
	(17.00)	(4.152)
Timespan	-23.69	-8.549***
	(21.73)	(3.067)
Budget*Rfilm	-0.694**	0.0180
	(0.335)	(0.0595)
PreAveTic		0.346***
		(0.0755)
Constant	-138.2***	-9.004
	(38.90)	(6.472)
Observations	341	341
R-squared	0.755	0.674

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Interaction Effect on Box Office (Source)

Interaction Effect on Box Office (Source)

	(5)	(6)
VARIABLES	Revenue	Viewer
Budget	2.146***	0.269***
	(0.386)	(0.0452)
PreAveRev	0.534***	
	(0.0896)	
Critics	2.204***	0.364***
	(0.350)	(0.0606)
ROMA	6.475	6.363
	(22.19)	(3.866)
Audience	0.287	0.0832
	(0.545)	(0.0730)
Rfilm	-24.23	-6.561*
	(17.29)	(3.510)
Source	59.45**	5.270
	(27.30)	(4.875)
Year	40.58**	-4.048
	(17.08)	(4.148)
Timespan	-23.45	-8.381***
	(21.43)	(3.053)
PreAveTic		0.346***
		(0.0754)
Budget*Source	-0.485	-0.0127
	(0.326)	(0.0445)
Constant	-147.7***	-10.15
	(40.47)	(6.535)
Observations	341	341
R-squared	0.755	0.674

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Interaction Effect on Box Office (Year)

Interaction Effect on Box Office (Year)

	(1)	(2)
VARIABLES	Revenue	Viewer
Budget	2.255***	0.383***
	(0.461)	(0.0907)
PreAveRev	0.544***	
	(0.0893)	
Critics	2.239***	0.374***
	(0.351)	(0.0605)
ROMA	2.468	6.329*
	(21.78)	(3.789)
Audience	0.158	0.0687
	(0.543)	(0.0725)
Rfilm	-15.64	-6.432*
	(17.45)	(3.440)
Source	18.34	4.216
	(19.91)	(3.412)
Year	54.30**	1.090
	(21.62)	(4.959)
Timespan	-28.13	-9.202***
	(21.93)	(3.055)
Budget*Year	-0.343	-0.131
	(0.464)	(0.0854)
PreAveTic		0.345***
		(0.0748)
Constant	-128.6***	-12.51*
	(39.38)	(6.557)
Observations	341	341
R-squared	0.753	0.677

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Interaction Effect on Box Office (Timespan)

Interaction Effect on Box Office (Timespan)

	(1)	(2)
VARIABLES	Revenue	Viewer
Budget	2.019***	0.258***
	(0.520)	(0.0544)
PreAveRev	0.540***	
	(0.0926)	
Critics	2.224***	0.364***
	(0.359)	(0.0609)
ROMA	1.905	6.330*
	(21.36)	(3.809)
Audience	0.184	0.0823
	(0.528)	(0.0729)
Rfilm	-16.04	-6.319*
	(17.84)	(3.489)
Source	18.79	4.177
	(20.32)	(3.446)
Year	40.55**	-3.995
	(16.80)	(4.158)
Timespan	-19.79	-9.006**
	(19.22)	(3.703)
budget*Timespan	-0.0962	0.00778
	(0.417)	(0.0465)
PreAveTic		0.347***
		(0.0760)
Constant	-123.7***	-9.214
	(41.94)	(6.329)
Observations	341	341
R-squared	0.753	0.674

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1