Are Hollywood Stars Worth the Price Tag?

TJ Cole and Chris Foote

Advised by: Dr. Michelle Connolly

Honors thesis submitted in partial fulfillment of the requirements for Graduation with Distinction in Economics in Trinity College of Duke University

Duke University

Durham, North Carolina

2018

Acknowledgements

We would like to thank Dr. Connolly for her unwavering patience throughout this process. We would not have been able to complete this without her time and hugely helpful insights. We would also like to thank our thesis classes from both semesters for their support and creative input.

Abstract

We investigate the effect of a lead actor's popularity on the profitability of films. Google search data is used as a proxy for actor popularity. We then investigate if lead actor's popularity has a different effect on movies that are not part of a sequel or franchise, and those that belong to specific genres. The most profitable movies are franchises and sequels. Movies are more profitable when they are action movies rated G or PG, although in certain circumstances a small number of horror movies and musicals can be hugely profitable. We find that across all groups of movies our proxy for lead actor popularity has no significant effect on a film's profitability.

Table of Contents

1. Introduction:

Movies are risky investments. Even for industry experts, it is difficult to predict which films will be blockbuster hits, and which will flop with little revenue. The most profitable movie to date, *Paranormal Activity*, cost only $15,000 to make and earned $193 million for Paramount Studios (BoxOfficeMojo.com). In contrast, big-budget films can incur fantastic losses. For example, *Cowboys & Aliens*, starring famous actors Harrison Ford and Daniel Craig, cost $163 million to make, but after estimating the advertising costs, the box office revenues fell short of total costs by $75 million (BoxOfficeMojo.com). This study aims to investigate the profit provided by the status of a film's lead actor or actors.

The film market is an incredibly valuable in the United States. The industry is dominated by six major studios that together hold approximately 75% of box office market share (The Numbers, 2017). In 2016, films generated $11.4 billion in box office revenues alone. Total inflation-adjusted box office revenue has been relatively stagnant over the past 10 years with rising ticket prices offset by falling ticket sales (The Numbers, 2017; MPAA, 2017). The rise of in-home streaming through Netflix and other platforms is a close substitute contributing to falling ticket sales. USA Today reported that spending on streaming services will surpass domestic box office revenue by $2 billion by the end of 2017.

The process of making a movie is multifaceted, and therefore presents a great deal of uncertainty to filmmakers. Whether it is an original idea, a particular script, or the search for content that draws a filmmaker to create a film, this inception has the potential to make or break the film. Once the film is chosen, the process of producing a film incorporates six main phases: financing the project, hiring the various personnel required, filming, post-production (editing, dubbing, scoring), distribution to theaters both domestic and abroad, and advertisement and promotion (Faulkner & Anderson, 1987). Each facet of this operation presents risk to the filmmakers, and an error in a single phase can lead to unexpected changes in profits. With the availability of a nearly limitless number of substitutes in the forms of other movies and entertainment, the demand for any particular movie is necessarily elastic. This makes the process of creating a film extraordinarily difficult and requires producers to optimize every aspect of a movie.

One unavoidable component of a film is the lead actor or actors. Often prominent on the film's poster, in advertisements, and in trailers, even unknown leads are inherently tied to the production and its public image. Popular lead actors receive significant amounts of money for their work, with well-known actors earning as much as $68 million in a single year (Forbes, 2017). Films are willing to spend a large share of their budgets on the cast. Do such actors generate enough revenue to justify their massive price tags? This study hypothesizes that more popular actors have a positive effect on Box Office returns.

2. Literature Review:

The nature of the film industry lends itself to variability. Faulkner and Anderson (1987) explain that this is due to the artistic component of the content itself, constant changes in capital flows, and stochastic demand. The authors show ways in which film contractors (directors, producers, etc.) reduce inherent risk in the motion picture industry. Contractors are able to minimize risk by utilizing short-term (per-project) contracts for capital and labor and mutuality of choice. Mutuality of choice describes how a contractor's accumulated successes make him more or less valuable, allowing him to choose the more attractive prospects while the more attractive prospects actively seek out more successful contractors.

Due to this variability, multiple studies attempted to investigate the drivers of box office revenue. These factors can be segmented into market specific traits and film specific traits.

2.1 Market Specific Traits

To account for changes over time, studies control for the year of the film's release. This encompasses overall market climate, changes in substitutes for movie theaters such as DVD's and streaming services, and changes in consumer preferences consumer preferences. Such factors are constantly changing and have the ability to affect a films profits.

Timing within a year also affects motion picture profits. Attendance to theaters varies based on seasonality, thus making the release date of a movie important to control. For example, a movie released in the holiday, when demand for films is high, may receive more revenues than if it were released in February. Many studies control for year (Wallace et al.) or for season of release (Basuroy et al., 2003; Litman & Kohl, 1989), but rarely for both.

2.2 Film Specific Traits

While there are countless aspects of a film, past literature reveals that certain traits tend to say more about a given movie's profits. These include whether or not the film is a sequel, a measure of critics' reviews, and the popularity of the lead actor or actors.

Sequels have been shown to have a positive correlation to a movie's revenue (Nelson & Goltfelty, 2009; Litman & Kohl, 1987; Ravid, 1999). Albert (1998) hypothesizes that consumers are more likely to see a movie that fits a familiar, comfortable type, which he labels a "mark." Albert (1998) argues that a star actor, a sequel, or an adapted movie (from a book, TV show, or past movie) can mark a movie, making it more appealing to its audience and translating to increased revenue. Albert (1998) does not account for the difference in demand for a franchise, in a series of three or more films, versus simply a sequel.

Critic reviews are largely considered a significant factor to a film's revenue (Litman & Kohl, 1989; Basuroy et al., 2003). The degree to which critiques are a determining factor for the popularity of the film is unclear. However, evidence from Basuroy et al. (2003) suggests a negativity bias towards critic reviews, meaning negative reviews hurt a movie to a larger degree than positive reviews help it.

The most controversial aspect of a film's revenue is the effect of a star actor. Some argue that a star actor boosts revenue (Nelson & Goltfelty, 2009; Elberse, 2007). Ravid (1999) claims that a star actor does not affect the magnitude of a film's financial success. Basuroy et al. (2003) find that stars have no effect given positive reviews, yet do increase revenues given negative reviews. This last argument supports an idea presented by Albert (1998); that stars are not only hired to increase revenue, but also to act as a hedge against underperformance.

2.3 Methodologies

In relevant literature, box office revenue is often used as the dependent variable to measure a film's success. However, different authors estimate box office revenue in different ways. Litman (1983) uses movie rentals as a proxy for box office revenue due to data availability. DeVany & Walls (1999) employ cumulative box office revenue up to a certain time period. Basuroy et al. (2003) utilize weekly box office revenues. Nelson & Glotfelty (2009) use cumulative box office revenues broken down by country. These papers do not incorporate budget into their dependent variable, nor do they look at return on investment. The current study will include the budget in the dependent variable in order to explain variation of profit.

There is no agreed upon standard for measuring star power, and many studies use inadequate metrics. Authors typically use a binary variable for star power determined by whether or not the top actor starred in a box office hit in previous years (Litman, 1983, Litman & Kohl, 1989; Albert, 1998) or was nominated for an Oscar in the years leading up to the movie (Basuroy et. al., 2003; Ravid, 1999). Nelson and Goltfelty (2009) incorporate a continuous variable for star power. The authors use IMDB, a website that tracks movies, actors, directors etc. and their work. IMDB has a "score" generated by a proprietary algorithm that takes multiple attributes into account, focusing on page views (IMDB). The current study aims to use a more accurate variable to measure star power and its subsequent effect on a film's profit.

3. Theoretical Framework

Although films are a form of art, they are also investments. Outside of exceptions such as passion projects that do not expect to make money, these investments are made by firms seeking to maximize profit. The investment has many components including production and marketing. Depending on the notoriety of the actors, casting can also be a big expense. If the decision is made to hire a star actor, his or her salary is often a large proportion of the budget.

Revenue alone does not show the success of a movie. Certain films are able to capture more revenue due to budget. In terms of dollar returns on movies, a film with a $1 million budget earning $10 million in box office revenue would be much more successful than a film with a $9 million budget and the same revenue. Instead of observing sheer revenue, a metric for profit makes the most sense. Profit displays the money made from a film after accounting for both the production and advertising expenditures.

It is important to note the methods of measuring revenue and budget of a film. Revenue can be broken into many different streams, including, but not limited to, box office, TV rights, streaming, licensing, and auxiliary revenue. Different films' profit is affected by each stream differently. For example, a franchise with a young fan base, such as the *Star Wars* series, earns a sizable amount of revenue from selling toys, whereas a film such as *Silence of the Lambs* does not. Production budget measures the cost of producing a film. This includes all costs pre-production, the filming, and post-production and includes costs such as salaries, set design, sound editing, and so on. This is separate from the print and advertising (P&A) budget, the cost of advertising the film before and after the movie airs on all forms of media, as well as the cost of printing the physical film that the studio sends to movie theaters. For major studio releases, the average P&A budget was just over half of average production budgets (The Numbers, 2017). The P&A costs incur before, and after the film's release date, making it difficult to discern whether this budget is influenced by revenue or not. Due to this potential endogeneity with the film's revenue and the lack of granular data on P&A budgets, we chose not to include P&A spending in our budget estimation.

The most accurate measure of profit would include all streams of revenue and total budget (P&A and production). However, time presents an issue in this calculation. More time since the release date translates to increased revenue in certain streams, such as auxiliary, streaming, and TV right revenues, due to the annual nature of these goods. Limiting revenue to what a film accrues in theaters and the budget to simply production budget removes this issue. So long as the films are measured by the same criteria, this reduced metric will still provide a consistent measure of a film's financial success relative to peers. We also operate under the assumption that the ratio of P&A to production budget is relatively constant across the film industry. That is, that many firms do not spend double the production budget on advertising while others spend nothing on advertising. If this were the case, then the profit metric would be inaccurate by omitting P&A expenditures. Similarly, we assume that production budget is defined the same way across the film industry,.

The variance in profit also depends on the the motives for making a film. Many actors, directors, and writers partake in so-called "passion projects," or films that they make for personal reasons. However, given the focus on return on investment, it is crucial to focus on films that aim to make money. This lends itself to limiting the films to those made only by large studios, so we can be sure that each film is fully profit seeking and aspiring to maximize return on investment.

To accurately measure the lead actor or actors impact on profit, it is necessary to control for other variables that might have an impact on a film's revenue. For one, critic's reviews of the film can alter its ticket sales. The film's genre is a distinguishing characteristic that has potential to affect its profit. Likewise, a film's MPAA rating can legally preclude viewers under 17 years old from seeing a film in theaters, and affect a parent's decision on whether the movie is appropriate for children in different age groups (Ravid, 1999). Both factors affect a film's profit. Another attribute of a film is its translation style in other countries. Because foreign revenue accounts for over half of total box office revenue, it is critical to distinguish whether the film is dubbed or has subtitles for different languages. Season of release may determine a film's profit. Movies released during holidays and summer often generate more revenue than other films released in other seasons (Ravid, 1999). This distinction is important in driving a film's profit. Furthermore, the movie industry, and, consequently, ticket sales, have shifted significantly over time. Therefore, year of release is also important.

The distinction between a sequel and a franchise is critical. A franchise film, which has been classified as a series of films of three or more, has a distinctly different fan base than a sequel. Often, franchises are established films with dedicated viewers and star actors. This gives them a Box Office advantage over a sequel. Due to this, it is important to clarify the difference and control for both sequel and franchise.

Google search data serve a proxy for lead actor popularity. An actor's name is Googled for a variety of reasons, whether it be to see news, learn of a new movie, or see an actor's past performances. Magnitude of Google searches show overall interest in a given person. It follows that the more an actor's name is searched, the more people are likely to see his or her new film. Since Google is not a film-specific website and a ubiquitous search engine, results are not limited to cinephiles and frequent movie-goers. The data should give an accurate representation of total interest in a given actor for certain period of time. This study employs Google searches as a more accurate proxy than other methods of defining an actor's popularity.

In order to pinpoint exactly where an actor may have the most impact on profit, this study regresses different segments of its film sample. After attempting to explain variance in the entire sample of films, it will break the data down into genre specific regressions as well as look into other scenarios that might make an actor's popularity a significant factor in profit. This aims to test if, for example, popular actors increase revenue in dramas while they do not in action films.

4. Empirical Specifications

4.1 Regression Components

This study employs an OLS regression in order to explain variance in returns of investments using a multitude of variables listed below. In this model, the dependent variable is a firm's profit for a given film. To construct this value a film's total cost is subtracted from its gross revenue. We define total gross revenue as revenue from domestic and foreign box-office ticket sales. Foreign markets account for over fifty percent of a film's total revenue on average (MPAA, 2017). It does not include any merchandise associated with the film, any licensing agreements, rentals, or streaming sales. Total cost of the film is measured as the sum of the production budget.[1]

$$\begin{aligned}(Y_i - Z_i) = \alpha &+ \beta_1 Year_i + \beta_2 Season_i + \beta_3 Rating_i + \beta_4 Genre_i + \beta_5 Studio_i + \beta_6 Sequel_i \\ &+ \beta_7 Franchise_i + \beta_8 Source_i + \beta_9 Runtime_i + \beta_{10} AudienceScore_i \\ &+ \beta_{11} Votes_i + \beta_{12} LeadActorPopularity_i + \beta_{13} OscarNominations_i \\ &+ \beta_{14} BestPicture_i + \varepsilon_i\end{aligned}$$

A continuous variable of year will control for market fluctuations on an annual basis, as well as for inflation. Season will indicate whether profit is affected by the season of release.

[1] In the equation below, Y_i is worldwide gross revenue and Z_i is production budget.

Rating and Genre are a list of indicators that aim to control for each film's MPAA rating (G, PG, etc.) and genre. Studio indicates whether or not a major studio produced the given film. Sequel and franchise control for film's that have predecessors or are a part of a film franchise. Source indicates whether a screenplay is original or adapted. Runtime is length of the full film, while Audience Score and Votes are metrics from IMDB. Lead Actor Popularity, our variable of interest, is the multiple of times a given actor's name is searched in the year prior to the given film's release. The amount of major Oscar nominations a film receives is controlled, as well as whether the film won best picture at the Academy Awards.

4.2 Omissions

We chose to omit some notable variables including streaming revenue from popular services such as Netflix. Licensing agreement between content creators, such as Disney, and streaming platforms like Netflix are done in bundled packages. That is, streaming services pay a lump sum for the rights to stream all content from a particular company. This reality means it is difficult to measure the impact of individual films on overall streaming revenue.

Other previous methodologies have included an independent variable for the number of screens showing a particular film (Nelson & Glotfelty, 2009). This is problematic, because the variable is endogenous. Better performing movies will be shown on more screens, and retained longer by theaters to meet demand.

5. Data

5.1 Sources

We use multiple datasets to build a comprehensive view of films. IMDB provides titles, alternative titles, principal cast, and director for approximately 8 million movies. IMDB also provides a popularity rating for each film, constructed using votes from users of the IMDB website. In addition, IMDB provides the amount of users that voted on the aforementioned popularity score. IMDB also records length of a film in minutes.

Revenue and budget data comes from Opus Data, scraped from The-Numbers.com. This provides production budget, domestic gross revenue and worldwide gross revenue. Gross revenue is purely ticket sales at theaters and does not include the cost of distributing the film to movie theaters domestically and overseas. This includes the money a film makes from being shown in theaters. The lifetime of this revenue is generally around four weeks after release, but can be as short as two weeks and as long as eight weeks (The Numbers, 2017). Production budget describes the costs of making the movie, including cast and crew salaries, filming costs, editing costs, and so on. It does not incorporate advertising and promotion costs. Observations that do not include both revenue and budget data are omitted. Opus Data, via the-numbers.com, additionally provides data on release date, title, rating, genre, sequel, source, and studio. This study constructs a franchise variable for series of three or more films, as well as a foreign share variable, the share of total revenue that comes from abroad.

As a proxy for the popularity of an actor, we use Google search data to generate a score for each actor. Google provides the multiple of times that a keyword is searched compared to another keyword, serving as a baseline, for a specified time period. This study uses “Google website” as a comparison against all actor names as it provides a relatively constant baseline for which to compare search results from all actors, recorded from 1 year leading up to the release of each film on the number of searches the star actor receives.

5.2 Characteristics

The IMDB rating system gives each title a score of 0 to 10 based on votes on the IMDB website. Votes are cast by everyday users of IMDB and compiled by IMDB into a single metric by IMDB. The number of votes per title is also recorded.

The revenue and budget data is listed in dollars with no adjustment for inflation. Given the short amount of time a film spends in theaters, controlling for year of release effectively adjusts for inflation. The-numbers.com provides domestic and worldwide revenue[2], and from that we calculate international revenue. Profit for a given film i, P_i, is calculated using the formula below:

$$P_i = \ Revenue_i - Budget_i$$

It also lists U.S. release date by day, which is transformed to season and year. Month of release is classified as either holiday (November through January), summer (May through August), or other.

MPAA rating values are PG-13, R, PG, G, NC-17, Not Rated (NR), and various earlier forms of these ratings which have been changed to reflect current ratings. Dummy variables for each rating are created. Similarly, dummy variables are created for genre values. These include Action, Comedy, Drama, Thriller/Suspense, Horror, Documentary, and Musical/Concert. The genres Adventure and Western are added to Action, and Romantic Comedy and Black

---

[2] Worldwide revenue emcompasses both domestic and international revenues.

Comedy are added to Comedy. A dummy variable is created to reflect whether a film is a sequel or not.

Original screenplays are distinguished from adapted screenplays. Adapted screenplays can be adapted from any number of different sources, such as books, graphic novels, prior films, or plays. Original screenplays cannot be based on any prior story. Another variable is created to indicate whether or not a film is produced by a top ten studio by market share. The top ten studios are as follows: Warner Bros., Walt Disney, Universal, Sony Pictures, 20th Century Fox, Paramount Pictures, Lionsgate, MGM, New Line, Miramax.

Runtime gives the length of a film in minutes. Major Oscar Nominations list the number of nominations that a film receives for the following awards: Best Picture, Best Director, Best Leading Actor, and Best Leading Actress. Additionally, an indicator for the movies that won Best Picture at the Oscars is included.

Google searches are taken as a trailing one-year period ending on the release date of the film. The search magnitude of each actor's full name is recorded as a multiple of searches relative to the term "Google website" for each film. Below is a table showing characteristics of these data, as well as selected summary statistics.

Table 1. Variable Codes

Variable Name	Type	Code	Source
Production Budget	Number		Opus Data
Domestic Gross	Number		Opus Data
Worldwide Gross	Number		Opus Data
Profit	Number		Constructed
Foreign Share	Number		Constructed
Year	Number		Opus Data
Season	Number	0 - Other 1 - Holiday (11-1) 2 - Summer (5-8)	Opus Data
Rating	Number	0 - G 1 - PG 2 - PG-13 3 - R 4 - NC-17 5 - NR	Opus Data
Genre	Number	0 - Action/Adventure/Western 1 - Comedy/Romantic Comedy/Black Comedy 2 - Concert Performance/Musical 3 - Documentary 4 - Drama 5 - Horror 6 - Thriller/Suspense	Opus Data
Sequel	Number	0 - Non-sequel 1 - Sequel	Opus Data
Franchise	Number	0 - Non-Franchise 1 - Franchise	Created
Source	Number	0 - Original Screenplay 1 - Adapted Screenplay	Opus Data
Studio	Number	0 - Not top ten studio 1 - Top ten studio	Opus Data
Runtime	Number		IMDb
Average Rating	Number		IMDb
Number of Votes	Number		IMDb
Google Search Ratio	Number		Google Trends
Major Oscar Nominations	Number		Google
Best Picture Wins	Number	0 - Film did not win Best Picture at the Oscars 1 - Film did win Best Picture	Google

5.3 Summary Statistics

Table 2. Summary Statistics

Variable	Mean	Median	Std. Dev.	Min	Max	Count
Production Budget (mil)	$ 48.65	$ 29.00	55.18	$ 0.01	$ 425.00	1241
Worldwide Gross (mil)	$ 149.56	$ 67.87	231.74	$ 0.10	$ 2,783.92	1241
Domestic Gross (mil)	$ 61.99	$ 35.06	83.29	$ -	$ 936.66	1241
Profit (mil)	$ 100.91	$ 33.29	189.40	$ (110.45)	$ 2,358.92	1241
Foreign Share (%)	0.47	0.50	0.26	0.00	1.00	1241
Year	2011	2011	3.59	2005	2017	1241
Season						
Holiday	0.26	0	0	0	1	328
Summer	0.32	0	0.47	0	1	394
Other	0.42	0	0.49	0	1	519
Rating						
G	0.02	0	0.14	0	1	24
PG	0.16	0	0.37	0	1	200
PG-13	0.41	0	0.49	0	1	504
R	0.39	0	0.49	0	1	486
NC-17	0.00	0	0.03	0	1	1
NR	0.02	0	0.14	0	1	26
Genre						
Action	0.30	0	0.46	0	1	370
Comedy	0.24	0	0.43	0	1	296
Musical	0.01	0	0.12	0	1	18
Drama	0.25	0	0.44	0	1	316
Horror	0.07	0	0.26	0	1	88
Thriller	0.12	0	0.33	0	1	153
Sequel						
Sequel	0.13	0	0.34	0	1	160
Non-Sequel	0.87	1	0.34	0	1	1081
Franchise						
Franchise	0.06	0	0.24	0	1	73
Non-Franchise	0.94	1	0.24	0	1	1168
Source						
Adapted Screenplay	0.45	0	0.50	0	1	563
Original Screenplay	0.55	1	0.50	0	1	678
Studio						
Major Studio	0.79	1	0.41	0	1	980
Other	0.21	0	0.41	0	1	261
Runtime (minutes)	109.06	106.00	18.75	69	334	1241
Average Rating	6.40	6.50	1.02	2	9	1241
Number of Votes (thousands)	128.99	73.75	163.78	0	1690	1241
Actor Google Search Ratio	0.75	0.32	1.61	0	22	1241
Major Oscar Nominations	0.15	0.00	0.55	0	3	1241
Oscar Best Picture						
Win	0.01	0.00	0.08	0	1	9
Did Not Win	0.99	1.00	0.08	0	1	1232

5.4 Limitations

Opus Data often uses estimated budget data. Many values are an even factor of one million or half a million dollars. Exact budget data can be very hard to obtain for a film, and a round number is often reported. This can lead to issues of inaccuracy of rounding or overestimating or underestimating a budget. Bias may exist, however, in which films report revenue and budget data.

It is possible that Google searches render inaccurate results. Google Trends reports searches of the exact search terms inputted. This may provide misleading results if, for example, an actor shares a name with another celebrity. Google search data are entered manually and incorrect data is likely removed, however it is possible that some results can be misleading. It also presents the issue that an actor may be googled for completely different reasons than an interest in his or her film. This may give an actor a high popularity score that does not translate into revenue for the film.

Lastly, IMDB ratings are calculated by everyday voters. It is possible that there is some bias to who votes on IMDB webpages. Additionally, films with fewer ratings are affected more by a single vote. This may lead to skewed IMDB ratings because of the small sample size on the website.

6. Results

The table below demonstrates results of an OLS regression with profit as the dependent variable against the aforementioned independent variables.

Table 3. Regression 1 Results (All Movies in Data Set)

	Dependent variable:
	Worldwide Gross Revenue - Production Budget
Year	**4,688,788***** (977,140)
Season (Baseline = Other)	
Holiday	**17,142,010*** (8,741,114)
Summer	**14,541,143*** (8,156,586)
Foreign Share	**34,494,547**** (15,035,289)
MPAA Rating (Baseline=G)	
PG	-26,083,198 (25,873,869)
PG-13	**-70,383,928***** (25,789,584)
R	**-117,551,550***** (25,972,808)
NC-17	-186,942,902 (121,534,593)
NR	**-96,766,167***** (36,258,336)
Genre (Baseline=Action)	
Comedy	-14,644,338 (11,056,624)
Musical	**73,531,432**** (29,220,644)
Drama	**-21,600,847*** (11,434,369)
Horror	3,767,881 (15,534,136)
Thriller	**-36,168,390***** (12,571,043)
Major Studio	10,349,894 (9,271,223)
Sequel	**63,715,090***** (13,573,060)
Franchise	**160,023,341***** (19,730,741)
Adapted Screenplay	-11,838,351 (7,460,542)
Runtime (minutes)	364,843 (231,318)
Audience Rating	-4,689,089 (4,381,746)
Number of Votes	**671***** (29.424)
Lead Actor Popularity	1,130,962 (2,169,735)
Major Oscar Nominations	-7,832,417 (7,209,300)
Best Picture	**-112,647,051***** (42,204,895)
Constant	**-9,374,602,373***** (1,960,832,720)
Observations	1,241
R^2	0.617

Adjusted R^2	0.610
Residual Std. Error	118,313,444.000 (df = 1216)
F Statistic	81.741^{***} (df = 24; 1216)

Note: Standard errors in parentheses. ***, **, *, p < 0.01, 0.05, and 0.10 respectively.

As seen in the regression results above *year* has a positive and statistically significant effect on a film's profit. Holding other variables fixed, a film released in a given year is expected to earn $4.69 million more than a film released in the previous year. This coefficient is probably, at least in part, capturing nominal changes due to inflation, as well as annual changed in demand for motion pictures.

The categorical variables differentiating *season* are also positive and significant. Movies opening during the holiday season are predicted to earn over $17 million more than those released in an off-season, while movies opening during the summer are predicted to earn over $14 million more than those released in an off-season. There is potential two-way causality here. In all likelihood, studios strategically release films that they believe will be more profitable during these two times of the year when box-office ticket sales are higher.

Compared to other ratings, the 24 G-rated movies in the data set are the most profitable. This group's profit is being skewed upwards by massive hits such as Pixar's *Toy Story 3, Ratatouille*, and *Monsters University*. Each other *MPAA rating* (PG, PG-13, R, NC-17, and NR) has a negative coefficient. MPAA ratings of PG-13, R, and NR are also statistically significant. G movies expect to earn over $70 million more than PG-13 movies, about $117 million more than R movies, and about $97 million more than NR movies.

The baseline genre of action movies is not as profitable as the 18 musicals in the data set. The coefficient for musicals is over $73 million and highly significant. This number is skewed upwards by box office monsters, such as *Beauty and the Beast*, *Hairspray*, and *Mamma Mia!* The coefficient for horror movies is small, positive, and insignificant. The coefficients for the other genres are negative. The coefficients for dramas and thrillers, -$21.6 million and -$36.2 million respectively, are both statistically significant.

Two of the largest positive and statistically significant coefficients in the regression are *franchise* and *sequel*. A film released as a sequel was expected to earn over $63 million more in profit compared to a film that is not. A film released as part of a franchise was expected to earn over $160 million more in profit compared to a film that stands on its own. Here there is potential two-way causality. Unprofitable films are not turned into sequels or franchises. Therefore, the *sequel* and *franchise* variables are also partially capturing the effect of a successful first film in the series and not just the effect of a film having established predecessors.

Based on the regression results, it appears the dominant strategy for studios is to produce action movies, market them within a sequel or franchise structure, make sure the films have a lenient G or PG rating, and then release those films during a peak holiday or summer season.

Separately, the coefficient for *number of votes* is positive and highly significant. With each extra vote a movie receives, that film expected to earn $671 more in profit. However, there is a potential endogeneity problem here. It may be the case that more profitable movies receive more votes because more people have been to the theaters to see them and now want to voice their opinion. The estimate for *best picture* is large and negative. Films that did not win best picture are predicted to profit about $112 million dollars more than films that did, all else equal. This makes intuitive sense, as movies deemed the best film of the year by the Academy of Motion Pictures are often made with more of an artistic intention, as opposed to one that is profit-seeking.

The regression has an adjusted R^2 of 0.61, explaining 61% of the variation of movie profit with the variation of explanatory variables. However, *lead actor popularity* is not statistically significant and had a very small positive coefficient of $1.13 million. This is contrary to our initial hypothesis that more popular lead actors would have a significant and positive effect on a film's profitability. It is possible that in this first regression the large coefficients for *franchise* and *sequel* were dwarfing the potential significance of other explanatory variables, namely *lead actor popularity*. To test this hypothesis, we run a regression excluding movies that were sequels and franchises.

Table 4. Regression 2 Results (Non-Sequels, Non-Franchises)

	Dependent variable:
	Worldwide Gross Revenue - Production Budget
Year	**3,620,225***** (936,655)
Season (Baseline=Other)	
Holiday	9,209,403 (8,296,274)
Summer	**13,509,889*** (7,797,027)
Foreign Share	**33,640,562**** (13,985,522)
MPAA Rating (Baseline=G)	
PG	9,002,548 (25,938,549)
PG-13	-34,979,793 (25,736,270)

R	**-70,912,410***** (25,900,537)
NC-17	-143,014,190 (109,033,623)
NR	-46,576,863 (34,793,420)
Genre (Baseline=Action)	
Comedy	-14,192,020 (10,379,400)
Musical	**85,855,011***** (26,217,603)
Drama	-16,578,201 (10,668,449)
Horror	**26,270,297*** (15,567,257)
Thriller	**-33,305,766***** (11,738,331)
Major Studio	**15,980,631*** (8,447,712)
Adapted Screenplay	-11,663,953 (7,095,452)
Runtime (minutes)	21,603 (221,093)
Audience Rating	-5,965,883 (4,166,117)
Number of Votes	**644***** (28.207)
Lead Actor Popularity	669,006 (2,017,190)
Major Oscar Nominations	123,793 (6,555,977)
Best Picture	**-117,328,784***** (37,761,458)
Constant	-7,223,364,419*** (1,879,734,492)
Observations	1,080
R^2	0.504
Adjusted R^2	0.494
Residual Std. Error	105,504,146.000 (df = 1057)
F Statistic	48.789*** (df = 22; 1057)

Note: Standard errors in parentheses. ***, **, *, p < 0.01, 0.05, and 0.10 respectively.

We observe slight differences between the regression including franchise and sequel films and the one excluding them. In this regression, the indicator *holiday* is not significant, while the indicator *major studio* is positive and significant. All else equal, films that are not franchises or sequels are predicted to earn $15.98 million more if they were produced by a major studio. In addition, the coefficient estimate for PG-13 is no longer significant, and the coefficient for horror movies is now significant. Interestingly, the dominant strategy of producing action film with G or PG ratings and marketing them during peak seasons still holds. However, in this regression there are a few highly profitable horror films like *It* and *Get Out* skewing the genre's coefficient upwards.

The regression has a lower adjusted R^2 of 0.494, explaining about 49% of the variation in film profit with the variation of the independent variables. Still, the coefficient estimate for *lead actor popularity* is not statistically significant and even smaller compared to that of the previous regression. It is possible that *lead actor popularity* is more important in certain genres than others. To test this hypothesis, we run regressions on each individual genre with 200 or more observations. The three genres are drama, comedy, and action.

Table 6. Regression 3-5 Results (Genre Specific)

	Dependent variable:		
	Worldwide Gross Revenue - Production Budget		
	(3) Drama	(4) Comedy	(5) Action
Year	**1,700,481***	158,960	**6,879,724*****
	(885,787)	(886,279)	(2,599,604)
Season (Baseline=Other)			
Holiday	**12,778,632***	10,523,101	29,567,263
	(7,584,821)	(7,825,434)	(23,250,023)
Summer	8,303,746	**16,161,800****	-730,701
	(8,039,602)	(6,272,763)	(22,332,140)
Foreign Share	**26,127,213****	**45,455,078*****	**83,213,357***
	(11,657,872)	(13,113,972)	(47,434,225)

MPAA Rating (Baseline=G)			
PG	1,056,209 (39,777,464)	16,600,146 (48,326,937)	-16,986,036 (40,206,238)
PG-13	-28,862,909 (39,206,679)	7,654,423 (47,609,996)	**-96,117,531**** (43,513,629)
R	-39,358,833 (39,250,165)	-14,566,124 (47,856,229)	**-191,386,285***** (43,775,656)
NC-17	-92,528,113 (67,417,102)	-	-
NR	-24,129,615 (44,237,545)	-13,554,783 (52,364,257)	**-166,228,011**** (79,654,931)
Major Studio	**20,611,186***** (7,148,473)	**18,560,074**** (7,761,123)	20,595,700 (33,658,763)
Sequel	38,583,853 (25,311,602)	**43,584,306***** (10,732,946)	**88,673,508***** (28,236,298)
Franchise	52,723,677 (60,194,392)	**276,926,499***** (49,412,593)	**109,288,855***** (37,418,644)
Adapted Screenplay	8,115,034 (6,688,309)	4,901,090 (6,895,721)	**-46,472,338**** (18,744,954)
Runtime (minutes)	-98,092 (165,264)	**593,946**** (260,919.800)	941,365 (633,604)
Audience Rating	**-18,539,331***** (4,416,203)	**-5,991,789*** (3,308,348)	-708,634 (11,849,723)
Number of Votes	**409***** (34.901)	**414***** (36.93)	**878***** (67.633)
Lead Actor Popularity	-648,852	1,245,802	-1,873,265

	(1,572,667)	(1,574,956)	(8,515,119)
Major Oscar Nominations	7,125,082	**-19,171,563****	18,732,130
	(4,528,623)	(9,926,946)	(29,390,039)
Won Best Picture	-13,285,560	-57,781,446	**-375,692,763****
	(25,191,389)	(53,825,425)	(181,487,880)
Constant	**-3,290,671,338***	-372,005,666	**-13,898,240,569*****
	(1,779,823,188)	(1,774,337,734)	(5,209,473,001)
Observations	316	296	370
R^2	0.512	0.619	0.657
Adjusted R^2	0.481	0.594	0.640
Residual Std. Error	54,139,908 (df = 296)	47,039,669 (df = 277)	167,601,779 (df = 351)
F Statistic	16.357*** (df = 19; 296)	24.950*** (df = 18; 277)	37.405*** (df = 18; 351)

Note: Standard errors in parentheses. ***, **, *, p < 0.01, 0.05, and 0.10 respectively.

The adjusted R^2 for dramas is 0.481 and the lowest of the three genres, indicating that those have the greatest variation in profit when controlling for these explanatory variables. The adjusted R^2 for comedies and action movies are 0.594 and 0.640 respectively. More of the variation of these genres' profit is explained by the variation in the independent variables measured here. There are other notable differences between the genre-specific regressions.

Year is significant and positive for both action movies and dramas, but not for comedies, indicating the market for comedies might be stagnant compared to a positive trend for action movies and dramas. *Foreign share* is positive and statistically significant for all three genres, about $26 million for drama, $45 million for comedy, and $83 million for action. As discussed in an earlier section of the paper, the foreign market is becoming increasingly important for box-office profits, and the results show movies with a higher proportion of revenues overseas tend to be more profitable. *MPAA rating* is insignificant for comedy and drama, but negative and statistically significant for action movies with PG-13, R, or NR ratings. Intuitively, dramas are not often marketed in the G or PG range, so the insignificance of the MPAA rating is logical.

Sequel and *franchise* are positive and significant for action movies and comedies, but insignificant for dramas. This also makes sense as dramas are not often part of a sequel or franchise in the first place. *Audience rating* from IMDb is negative and significant for both dramas and comedies, but insignificant for action movies. In addition, *number of votes* is significant and positive across all three genres, with the highest coefficient estimate, 878, for action movies.

Lead actor popularity is insignificant across all three genres, and the coefficient is negative for action movies and dramas.

7. Summary of Results

In films released between 2005 and mid-2017 our proxy for lead actor popularity, the multiple of google searches for the actor's name compared to a baseline, has no significant effect on that film's profit. This is true across all three major genres, comedy, drama, and action, as well as sequels, franchises, and films that had no predecessor. These findings are inconsistent with our original hypothesis. There are multiple reasons why we may have observed these results, and we put forth the most likely of them below.

1. More popular lead actors do not make films more profitable. This would mean that other factors are more dominant in driving a film's profit such as the quality of the script, special effects, genre, marketing, or any number of other variables.
2. Movie-goers are not using google as a source for information on Hollywood stars.
3. More popular lead actors and their high salaries might inflate production budgets, negating any positive effect the actors might have on box-office revenue.
4. Having a well-known lead actor in a film might reduce the necessary marketing expenditures. This would mean that an actor's popularity is positively effecting profit, but we are unable to see the effect in our regressions because we do not have marketing data.
5. Having a popular lead actor might boost other sources of a films revenue such as toy and merchandise sales, or warrant a higher price for the movie when being sold as part of a bundle to a streaming service.

8. Suggestions for Further Research

Using a similar regression structure, constructing a more precise proxy for actor popularity will likely yield better results. This can potentially be achieved by creating a weighted average of the actor's google searches and the actor's IMDb STARMeter ranking available with an IMDb Pro account. The IMDb STARMeter ranking system has historical information available that shows an actor's STARMeter rank on a given day, or his highest rank during a time frame.

In addition, a few more explanatory variables may help control for some of the unexplained variation in a given film's profit. One is a proxy for the critical reception of the film, via Rotten Tomatoes or another aggregator. Another is a measure of the director's popularity at the time of release, which can be proxied through google search data or potentially through IMDb STARMeter. A third category to control for is the supporting cast. To determine a method for indicating an ensemble cast, one that is comprised of many actors, usually well-known, sharing some equivalent screen time, could explain some variation in profit. Further, proxies for the popularity of the second or third leading actors may also explain some of the variation in profit. Finally, although the information released by studios is sparse, the marketing budgets for a film plays a huge role in its bottom-line.

Another area for further research, depending on the availability of data, is the bargaining process between streaming platforms like Netflix and content creators such as Miramax. Content creators and streaming platforms probably construct estimates on how often a film would be streamed, and how much that number of streams is worth. If, in the future, there is granular data detailing these estimations for each film, it would be interesting to see if more popular actors end up in movies sold at higher price points within the bundle.

<u>References</u>:

Albert, S. (1998). Movie stars and the distribution of financially successful films in the motion picture industry. *Journal of Cultural Economics*, 22(December), 249–270.

Austin, B. (1989). *Immediate seating: A look at movie audiences*. Belmont, CA: Wadsworth.

Basuroy, S., Chatterjee, S., & Ravid, A. (2003). How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing*, 67(October), 103–117.

Boxofficemojo.com. *Cowboys & Aliens (2011) - Box Office Mojo*. [online] Available at: http://www.boxofficemojo.com/movies/?id=cowboysandaliens.htm [Accessed 11 Oct. 2017].

Boxofficemojo.com. *Paranormal Activity (2009) - Box Office Mojo*. [online] Available at: http://www.boxofficemojo.com/movies/?id=paranormalactivity.htm [Accessed 11 Oct. 2017].

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

Faulkner, R., & Anderson, A. (1987). Short-term projects and emergent careers: Evidence from hollywood. *American Journal of Sociology*, 92(January), 879–909.

Google. (2018). *Google Trends*. [dataset] Retrieved from: https://trends.google.com/trends/ [Accessed Apr. 2018].

Help.imdb.com. (2018). *IMDb | Help*. [online] Available at: https://help.imdb.com/article/imdbpro/industry-research/faq-for-starmetermoviemeter-and-companymeter/GSPB7HDNPKVT5VHC?ref_=helpsect_pro_6_1# [Accessed 18 Apr. 2018].

IMDb. (2018). *IMDb Interfaces*. [dataset] Retrieved from: https://datasets.imdbws.com/ [Accessed Apr. 2018].

Litman, B. (1983). Predicting success of theatrical movies: An empirical study. *Journal of Popular Culture*, 16(Spring), 159–175.

Litman, B., & Kohl, L. (1989). Predicting financial success of motion pictures: The 80's experience. *Journal of Media Economics*, 2(Fall), 35–50.

MPAA. (2017). Theatrical Market Statistics: 2016. *MPAA.* Accessed at https://www.mpaa.org/wp-content/uploads/2017/03/MPAA-Theatrical-Market-Statistics-2016_Final-1.pdf [Accessed October 30, 2017].

MPAA. (2017). The Economic Contribution of the Motion Picture & Television Industry to the United States. *MPAA.* Accessed at: https://www.mpaa.org/wp-content/uploads/2017/02/Economic-Contribution-of-the-Motion-Picture-Television-Industry-to-the-United-States-2017.pdf [Accessed October 30, 2017].

Nash Information Services. (2018). All Time Domestic Sequel Box Office. *Nash Information Services via the-numbers.com.* [dataset] Retrieved from: https://www.the-numbers.com/box-office-records/domestic/all-movies/cumulative/sequel [Accessed April 15, 2018].

Nash Information Services. (2018). Domestic Genres Box Office Records. *Nash Information Services via the-numbers.com.* [dataset] Retrieved from: https://www.the-numbers.com/box-office-records/domestic/all-movies/genres/ [Accessed April 15, 2018].

Nash Information Services. (2018). Domestic MPAA Box Office Records. *Nash Information Services via the-numbers.com.* [dataset] Retrieved from: https://www.the-numbers.com/box-office-records/domestic/all-movies/mpaa-ratings/ [Accessed April 15, 2018].

Nash Information Services. (2018). Domestic Sources Box Office Records. *Nash Information Services via the-numbers.com.* [dataset] Retrieved from: https://www.the-numbers.com/box-office-records/domestic/all-movies/sources/ [Accessed April 15, 2018].

Nash Information Services. (2018). Domestic Theatrical Distributors Box Office Records. *Nash Information Services via the-numbers.com.* [dataset] Retrieved from: https://www.the-numbers.com/box-office-records/domestic/all-movies/theatrical-distributors/ [Accessed April 15, 2018].

Nash Information Services. (2018). Movie Budgets. *Nash Information Services via the-numbers.com.* [dataset] Retrieved from: https://www.the-numbers.com/movie/budgets/all [Accessed April 15, 2018].

Nelson, Randy A., and Robert Glotfelty. “Movie Stars and Box Office Revenues: an Empirical Analysis.” *Journal of Cultural Economics*, vol. 36, no. 2, 2012, pp. 141–166. JSTOR, JSTOR, www.jstor.org/stable/43549820.

The Numbers. (2017). Domestic Movie Theatrical Market Summary 1995 to 2017. Accessed at: http://www.the-numbers.com/market/ [Accessed October 30, 2017].

Ravid, A. (1999). Information, blockbusters, and stars: A study of the film industry. *Journal of Business*, 72(October), 463–492.

Robehmed, Natalie. Forbes.com. (2017). *The World's Highest-Paid Actors 2017: Mark Wahlberg Leads With $68 Million.* [online] Available at: https://www.forbes.com/sites/natalierobehmed/2017/08/22/the-worlds-highest-paid-actors-2017-mark-wahlberg-leads-with-68-million/#7fa21fde3f92 [Accessed 18 April 2018].

Snider, M. (2017). Theaters taking hits from movie viewing at home, weak box office. *USA TODAY*. Accessed at: https://www.usatoday.com/story/money/business/2017/08/02/movie-theaters-getting-pinched-sluggish-box-office-movies-viewing-home/532256001/ [Accessed October 30, 2017]

Wallace, W. Timothy, Seigerman, Alan, and Holbrook, Morris B. The Role of Actors and Actresses in the Success of Films: How Much is a Movie Star Worth? *Columbia University*.

<u>A. Appendix</u>

A1.

The ordinary least-squared regression with all observations and explanatory variables as Regression (1), with the addition of *Production Budget* as an explanatory variable, yielded an adjusted R^2 of 0.6597. The coefficient for *lead actor popularity* was -$166,700 and statistically insignificant.

A2.

The ordinary least-squared regression only including the movies below the median IMDb *audience rating*, and including all explanatory variables other than *audience rating*, yielded an adjusted R^2 of 0.73. The coefficient for *lead actor popularity* was -$1,132,146 and statistically insignificant.