[Major Themes in Economics](https://scholarworks.uni.edu/mtie)

Volume 21 Article 7

Spring 2019

Determinants of Box Office Performance: Return of the Regressions

Griffin Scott
University of Northern Iowa

Determinants of Box Office Performance: Return of the Regressions

Griffin Scott

ABSTRACT. Brewer, Kelley, and Jozefowicz (2009) used data from the late 1990s to the early 2000's in order to determine what made a movie a success at the box office. I replicate their study using data from 2014 to 2018. Ex ante, production budget, critical reviews, horror and comedy movies were all significant at the one percent level. Ex post, the same variables plus number of screens, summer and holiday release, star power, and word-of-mouth were significant at the one percent level.

> "Very difficult to understand American audience, what they like, what they don't like. Some movies I like very much, it doesn't work. Some movies I don't like, it gets big box office. Very difficult."
>
> – *Jackie Chan*

I. Introduction

While Jackie Chan is not a Hollywood executive, he has been in the movie business for over 50 years, more than enough time to know how unpredictable the box office can be. Knowing what makes movies successful would allow filmmakers to use their resources to create as much revenue as possible. Not only would this be good for US GDP, but it would also generate utility for the American consumer. Producers, filmmakers, and economists have all tried to create the formula for box office hits. They've had mixed success, as some movies have made over a billion dollars while others fall flat. I will try to determine what factors are correlated with box office returns using fresh data in a regression analysis.

II. Literature Review

I replicated an empirical study performed by Stephanie M. Brewer, Jason M. Kelley, and James J. Jozefowicz (2009). They used movie data from the late nineties to the early 2000s to determine which variables have an effect on a film's gross box office performance. They ran three separate regressions: one to capture pre-release (ex ante) variables; one to capture post-release (ex post) variables; and an additional one to capture post-release variables to correct for collinearity. The ex post models were designed to reach the same result but separates two independent variables (budget and peak number of screens) to correct for collinearity between the two. They found that the model using the peak number of screens was preferable to the model using budget. In their ex ante regression, they found that production budget, the gross revenues of a prequel, income, favorable critical reviews, summer releases and Thanksgiving and Christmas releases had positive and statistically significant coefficients. In their ex post regression, they found that peak screens, star power squared, film award nominations, and positive word-of-mouth information sharing were all positive and significant determinants of box office performance (Brewer et al 2009, 599).

Smith and Smith (1986) published the first significant study on the performance of motion pictures. The purpose of their paper was to analyze the performance of 600 popular films to see which attributes best fit consumer demand. They found that more data was needed for this to be possible. They also concluded that their belief that consumer preferences shifted from "mass taste" films to more specific genres or niches appeared to be true (Smith and Smith 1986, 502). This paper was also important because they concluded that motion pictures are good examples of differentiated products funded by a large initial investment with unknown revenues, similar to other industries. This paper opened the floodgates for further study on the subject.

Bagella and Bechetti (1999) published a similar study on films produced in Italy. Their study looked at movies from 1985 to 1996 to see what useful information they could find from an empirical analysis. Specifically, they studied whether genre, the popularity of actors and directors, production house, and government assistance affected box office admissions. Their main findings were that the comedy genre most accurately reflected the consumer tastes of moviegoers in Italy and the popularity of actors and directors had a positive coefficient (Bagella and Bechetti 1999, 251).

Deuchert et al (2005) studied the effect of Oscar nominations and wins on a movie's financial success. They used a regression analysis as well as a survival analysis on the movie's screen time. They concluded that Oscar nominations create extra revenue for a movie, but Oscar wins have a neutral effect (Deuchert et al 2005, 165).

Ravid (1999) performed an empirical analysis to find what effect star power had on revenues and return on investment. His study tests other variables such as MPAA rating (a suitability rating from the Motion

Picture Association of America), sequels, and budget using data from films produced in the 1990s. He found that star-studded films had increased revenue. He also found that bigger budgets, sequels, highly visible films, and family-oriented films increased revenues. Ravid found that G and PG-rated films are correlated with a higher return on investment while sequels were only slightly correlated (Ravid 1999, 488).

Ravid (2004) did a similar study a few years later on R-rated films. He says that filmmakers are making project choices with the goal of maximizing revenue and return on investment. He thought that films that featured sex and violence would increase return on investment, but the evidence did not show this. The evidence did show, however, that movies that featured sex and violence did increase revenues, especially internationally (Ravid 2004, S1666).

Simonoff and Sparrow (2000) examined whether it is possible to predict movie grosses from generally available information. They performed a regression analysis on the gross box office revenues using several independent variables including genre, budget, country of origin, star power, movie rating, award nominations, among others. They concluded that a person could predict box office grosses with generally available information, but predictions were more accurate after the first weekend of a movie's release. They also found that award nominations and high advertising budgets tend to boost box office revenues (Simonoff and Sparrow 2000, 24).

De Vany (2004) performed a study on the "blockbuster" strategy that film studios use to create buzz for their movies. This strategy involves advertising strategies that show short clips of the film without exposing the quality of the film. This strategy is bolstered by the involvement of big-name actors and actresses associated with the film. This strategy creates a word-of-mouth transfer among moviegoers that further boosts revenues. He tested the hypotheses with a regression analysis on box office revenues. He found that budget, number of screens in the opening weekend, and presences of major stars are all positive and significant.

Walls (2004) did a smaller study of box office revenues that looked into the crime genre. He used this genre because he believed that it transcended the barriers of language and culture that other genres deal with. His main findings were that the domestic success of a film is indicative of foreign success and that production budgets are highly correlated with revenues.

III. Data

This study uses a cross-section of data from the top 100 grossing films each year from 2014 to 2018. I expected 500 observations but because of some data unavailability, the number of observations decreased to 460. Studies by Smith and Smith, Bagella and Bechetti, and Brewer et al, the paper being replicated, used similar methods in their data selection. The variables that I use in my model are the same ones that Brewer et al used in their study, though the data may be collected differently.

Dependent Variable

The dependent variable, gross domestic box office revenue, is measured in millions of dollars and was pulled from the Internet Movie Database. This is solely revenue from a film's first run in theaters. This does not include DVD rentals, streaming purchases, or additional releases of the film. Several papers used this variable as their dependent variable including Brewer et al, Smith and Smith, Ravid et al, and others.

Ex Ante Independent Variables

The production budget for films was also found on the Internet Movie Database. It is measured in millions of dollars and it is expected to have a positive coefficient. Increasing the budget is expected to increase gross revenues but at a diminishing rate. Movie production budgets make for an interesting variable because they include actors' wages, costs of special effects, and other factors designed to increase the value of the movie. Bagella and Becchiti include production budget for this reason. Ravid et al also include this variable and found it to have a positive coefficient.

The critical reviews data was pulled from Rotten Tomatoes. This is a popular website that aggregates ratings provided by movie critics for a particular movie. The rating is a number between 0% to 100%. This is expected to have a positive coefficient because it reflects critics' reviews of the quality of the film and moviegoers are exposed to them in the media and in advertising. Simonoff and Sparrow included this in their study and found it to have a positive effect on revenue.

The gross revenue of a preceding movie was found using the Internet Movie Database. If a movie was a sequel, then the preceding movie's gross box office performance was recorded. This is expected to have a positive coefficient because past success may indicate future success.

Monthly US personal income data were obtained from the Federal Reserve Bank of St. Louis. The data is in billions of US dollars. This is expected to have a positive coefficient because as income rises, the consumption of normal goods should increase as well. Movie tickets are assumed to be normal goods.

The CPI for movie tickets was also obtained from the Federal Reserve Bank of St. Louis. It is the CPI for Urban Consumers for Admission to movies, theaters, and concerts. The index began in December 1997 at 100 and is seasonally adjusted. This is expected to have a negative coefficient in accordance with the law of demand.

A dummy variable is used to capture an established audience. A value of one is used if the movie is based on a preceding movie, television show, book, comic book, or video game. A lack of an established audience was the omitted condition. These data were manually inputted using the Internet Movie Database as a source. This is expected to have a positive coefficient because the movie already has an audience prior to the release of the film.

The SUMMER and XMAS data were found by using the release date from the Internet Movie Database. For SUMMER, a value of one is used for movies released in May, June, or July. For XMAS, a value of one is used for movies released in November or December. Not being released in May, June, July, November, or December was the omitted condition. These are both expected to be positive because these months are considered peak movie-going months.

The star power data was from Forbes. I compiled a list of the top 10 movie actors and actresses for each year from 2013 through 2018. The data is lagged one year to account for ex ante effects. This is a count variable that keeps track of how many of the stars on the list are in a certain movie. The count is squared for each movie since star power is believed to have a quadratic effect, not a linear one. This variable is expected to be positive because the higher the quality of labor, the higher the quality of the product.

Genre is separated into seven variables: science fiction, comedy-drama, drama, comedy, action adventure, horror, and animation. These are the same categories used in Brewer et al's paper. The genres for each movie were found on the Internet Movie Database and each is represented by a dummy variable. There is a variable for each genre, for example, if a movie is a comedy, then COMEDY will have one and the rest will have zero. A prioiri the signs are uncertain because it is unclear as to what consumer preferences are for movie genres.

MPAA ratings are also from the Internet Movie Database. The ratings are G, PG, PG-13, and R. I used dummy variables to capture MPAA ratings with the R rating as the omitted condition. A prioiri the signs are uncertain.

Finally, a dummy variable is used to capture what year the movie was released. The years are 2014, 2015, 2016, 2017, and 2018 with the year 2014 set as the omitted condition. These data were found on the Internet Movie Database. *A prioiri* the signs are uncertain.

Ex Post Independent Variables

Word-of-mouth transfer of information data are from Cinema Score. This is a company that surveys moviegoers right after they leave a movie. The score is graded on the A to F scale with pluses or minuses included. A dummy variable was used to capture this with A+ being 14 and F- being zero. This is expected to have a positive coefficient.

Films that received awards for best picture, best actor, best actress, best supporting actor, best supporting actress, best director, and best screenplay are expected to have a positive effect on gross revenues. The Academy Awards (The Oscars), Golden Globes, British Academy of Film Awards, and Screen Actors Guild Awards are the four awards from which nominations are obtained from. The categories included are Best Actor, Best Actress, Best Supporting Actor, Best Supporting Actress, Best Director, Best Screenplay, Best Picture, and Best Animated Film. Film nomination data was captured by using a dummy variable to account for each nomination a film received. For example, if a movie was nominated for best picture and best actor, the dummy would be two.

Descriptive Statistics

Variable 1	Obs	Mean	Std. Dev.	Min.	Max.
GROSSBOXOF~E	460	1.07e+08	1.14e+08	1.58e+07	9.37e+08
ORIGINALGR~S	460	4.18e+07	1.00e+08	0	9.37e+08
BUDGET	460	7.24e+07	6.51e+07	100000	3.31e+08
SCREENS	460	3347.352	586.5258	1656	4535
INCOME	460	1.63e+13	9.5e+11	1.46e+13	1.80e+13
CPITICKET	460	173.9573	6.703617	163.637	184.789
CRITIC	460	.5511522	.2708821	.05	1
PREVAUD	460	.6326087	.4826192	0	1
SUMMER	460	.273913	.4464504	0	1
THANKSXMAS	460	.1956522	.3971338	0	1
STARPWER	460	.4326087	.7597025	0	7
CSDUMMY	460	11.11522	1.774237	3	14
FILMNON	460	.6391304	2.124705	0	18
SCIFI	460	.1152174	.3196317	0	1
COMDRA	460	.0652174	.2471779	0	1
DRAMA	460	.1782609	.3831489	0	1
COMEDY	460	.1695652	.3756589	0	1
ACTADV	460	.2543478	.4359681	0	1
HORROR	460	.0978261	.2974028	0	1
ANIM	460	.1195652	.3248057	0	1
G	460	.0065217	.0805812	0	1
PG	460	.1913043	.3937563	0	1
PG13	460	.4869565	.500374	0	1
R	460	.3152174	.4651082	0	1
DUM15	460	.2043478	.4036632	0	1
DUM16	460	.2065217	.40525	0	1
DUM17	460	.1847826	.3885437	0	1
DUM18	460	.2021739	.4020585	0	1

V. Model

I used two regressions in my empirical analysis, both using the logarithm of gross domestic box office as the dependent variable. The first model uses ex ante independent variables. Ex ante variables are intended to capture the determinants of box office success before the film is released in theaters. The second model uses ex post independent variables. Ex post variables are intended to capture the determinants of box office success after the film is released in theaters.

The ex ante model tested the following independent variables: logarithm of the production budget, critical review, logarithm of the gross of a preceding film, logarithm of US personal income for the month of release, the CPI for movie tickets, a dummy for an established audience, whether the movie was released in a peak period, star power, genre, MPAA rating, and a dummy for the year of release.

Model 1 – ex ante model

$$\begin{aligned}\log(GROSS) = {} & \beta_1[\log(BUDGET)] + \beta_2\,(CRITIC) + \beta_3[\log\,(ORIGGROSS)] \\ & + \beta_4[\log(INCOME)] + \beta_5\,(CPITICKET) + \beta_6(PREVAUD) + \beta_7\,(SUMMER) \\ & + \beta_8(THANKSXMAS) + \beta_9\,(STARPOWER^2) + \beta_{10}(ACTADV) + \beta_{11}(ANIM) \\ & + \beta_{12}(COMDRA) + \beta_{13}(COMEDY) + \beta_{14}(HORROR) + \beta_{15}(SCIFI) + \beta_{16}\,(G) \\ & + \beta_{17}(PG) + \beta_{18}(PG13) + \beta_{19}(DUM15) + \beta_{20}(DUM16) + \beta_{21}\,(DUM17) \\ & + \beta_{22}(DUM18) + e\end{aligned}$$

The ex post model tested the same variables as the ex ante model with the addition of word-of-mouth information sharing among consumers and award nominations that the film receives. This model uses the logarithm of the peak number of screens that the film was shown instead of the logarithm of the production budget.

Model 2 – ex post model

$$\begin{aligned}\log(GROSS) = {} & \beta_1[\log(SCREENS)] + \beta_2\,(CRITIC) + \beta_3[\log\,(ORIGGROSS)] \\ & + \beta_4[\log(INCOME)] + \beta_5\,(CPITICKET) + \beta_6(PREVAUD) + \beta_7\,(SUMMER) \\ & + \beta_8(THANKSXMAS) + \beta_9\,(STARPOWER^2) + \beta_{10}(CINEMASCORE) \\ & + \beta_{11}(FILMNOM) + \beta_{12}(ACTADV) + \beta_{13}(ANIM) + \beta_{14}(COMDRA) \\ & + \beta_{15}(COMEDY) + \beta_{16}\,(HORROR) + \beta_{17}(SCIFI) + \beta_{18}(G) + \beta_{19}(PG) \\ & + \beta_{20}(PG13) + \beta_{21}\,(DUM15) + \beta_{22}(DUM16) + \beta_{23}(DUM17) + \beta_{24}\,(DUM18) \\ & + e\end{aligned}$$

VI. Results

Model 1 – ex ante model

The ex ante model had several statistically significant variables. At the one percent significance level, budget, critical review, comedy, and horror were all positive.

Source	SS	df	MS
Model	30.1239942	22	1.36927246
Residual	25.9318167	437	.059340542
Total	56.0558109	459	.12212595

Number of obs = 460
F (22, 437) = 23.07
Prob > F = 0.0000
R-squared = 0.5374
Adj R-squared = 0.5141
Root MSE = .2436

LOGGROSS	Coef.	Std. Err.	t	P>| t |	[95% Conf. Interval]	
LOGORIGGROSS	.006576	.0038496	1.71	0.088	-.0009901	.0141421
LOGBUDGET	.4054208	.0385089	10.53	0.000	.3297352	.4811064
LOGINCOME	-6.326645	2.959654	-2.14	0.033	-12.14357	-.5097195
CPITICKET	.0007363	.0099932	0.07	0.941	-.0189043	.0203769
CRITIC	.4370649	.0457353	9.56	0.000	.3471765	.5269534
PREVAUD	.0545157	.0277653	1.96	0.050	-.0000545	.1090859
SUMMER	.0522755	.0295233	1.77	0.077	-.0057498	.1103008
THANKSXMAS	.1029916	.0382287	2.69	0.007	.0278566	.1781267
STARPOWERSQUARED	.0095829	.0042132	2.27	0.023	.0013023	.0178635
SCIFI	.0198062	.0467553	0.42	0.672	-.072087	.1116994
COMDRA	-.008954	.0526875	-0.17	0.865	-.1125065	.0945984
COMEDY	.1087693	.0406945	2.67	0.008	.0287881	.1887505
ACTADV	.0427428	.0389943	1.10	0.274	-.0338969	.1193825
HORROR	.2712178	.0511908	5.30	0.000	.1706071	.3718286
ANIM	.0506797	.0587198	0.86	0.389	-.0647286	.166088
G	.0118906	.1533971	0.08	0.938	-.2895972	.3133784
PG	.0969118	.045923	2.11	0.035	.0066545	.1871692
PG13	.0512642	.0278875	1.84	0.067	-.003546	.1060745
DUM15	.065108	.0633295	1.03	0.304	-.0593602	.1895762
DUM16	.1260316	.1034178	1.22	0.224	-.0772266	.3292897
DUM17	.2377871	.1520143	1.56	0.118	-.0609828	.5365571
DUM18	.3715905	.1984448	1.87	0.062	-.0184343	.7616154
cons	87.62174	38.2647	2.29	0.023	12.41602	162.8275

This is in line with expectations, the replicated study, and the literature. At the five percent significance level, the logarithm of income was negative; the previous audience variable, the peak seasons variables, and the star power variable were all positive. The negative coefficient on income was unexpected, but since income was relatively flat during this period, it makes some sense. Previous audience, peak seasons, and star power were all similar to the replicated study, but my study showed star power as statistically significant when the original study did not. At the ten percent significance level, the log of the original gross of a film was positive. All other variables were not statistically significant, but they had similar signs to Brewer et al's study. This is in line with the replicated study. This model had a decent fit according to its adjusted R-squared value of .5141. This is higher than the replicated study.

Model 2 – ex post model

The ex post regression that I performed also had a better fit than the replicated study in terms of adjusted R-squared (.6966 vs. .645). At the one percent significance level, log number of screens, critical reviews, the holiday release period, star power, word-of-mouth transfer, film nominations, comedy, and horror were all positive. This is similar to Brewer et al's results. At the five percent level, income was negative. Again, this has the opposite sign as the replicated study, but economic conditions are different. CPI of tickets was not statistically significant but was positive in my study, when it was negative in theirs. All other variables were not significant.

Source	SS	df	MS
Model	39.9384223	24	1.66410093
Residual	16.1173886	435	.037051468
Total	56.0558109	459	.12212595

Number of obs = 460
F (24, 435) = 44.91
Prob > F = 0.0000
R-squared = 0.7125
Adj R-squared = 0.6966
Root MSE = .19249

LOGGROSS	Coef.	Std. Err.	t	P>| t |	[95% Conf. Interval]	
LOGORIGGROSS	.002588	.0030633	0.84	0.399	-.0034328	.0086087
LOGSCREENS	2.678559	.1446336	18.52	0.000	2.394292	2.962827
LOGINCOME	-5.584227	2.344804	-2.38	0.018	-10.19278	-.9756741
CPITICKET	-.0055737	.0079043	-0.71	0.481	-.0211091	.0099616
CRITIC	.2209728	.0410347	5.39	0.000	.1403218	.3016239
PREVAUD	.0232875	.0219684	1.06	0.290	-.0198898	.0664648
SUMMER	.026548	.0233886	1.14	0.257	-.0194207	.0725167
THANKSXMAS	.1439847	.0297821	4.83	0.000	.08545	.2025195
STARPOWERSQUARED	.0093922	.0033102	2.84	0.005	.0028862	.0158983
CSDUMMY	.0556173	.0066914	8.31	0.000	.0424657	.0687688
FILMNOM	.0178891	.004899	3.65	0.000	.0082605	.0275178
SCIFI	.0654245	.0374317	1.75	0.081	-.008145	.138994
COMDRA	.0760949	.0424797	1.79	0.074	-.007396	.1595858
COMEDY	.0881965	.0332175	2.66	0.008	.0229097	.1534833
ACTADV	.0776108	.0307865	2.52	0.012	.0171021	.1381196
HORROR	.2443964	.0410486	5.95	0.000	.1637181	.3250747
ANIM	.0547003	.0465829	1.17	0.241	-.0368552	.1462557
G	-.125481	.1219638	-1.03	0.304	-.3651926	.1142305
PG	-.0410932	.0376863	-1.09	0.276	-.115163	.0329766
PG13	.0243544	.0223934	1.09	0.277	-.0196583	.0683671
DUM15	.0760127	.050048	1.52	0.130	-.0223533	.1743788
DUM16	.1544367	.0817746	1.89	0.060	-.0062859	.3151592
DUM17	.2719262	.1202795	2.26	0.024	.035525	.5083275
DUM18	.4089703	.1570446	2.60	0.010	.1003098	.7176307
cons	72.11782	30.33133	2.38	0.018	12.50364	131.732

VII. Conclusion

The movie industry is a billion-dollar industry, and though film is an art form, it is also a business. Filmmakers and economists alike are interested in what factors can increase profits. The study that I performed is my own try at finding the secret recipe. Production budget is a very telling sign of how a movie may perform. While a big budget does not guarantee big revenues, it can certainly raise the minimum that a film could make. Critical reviews are also important since they indicate the quality of a movie and moviegoers seek them out or are exposed to them through advertising. This is becoming more and more important as the popularity and reliability of Rotten Tomatoes, the film critic compiler website, grows. I frequently consult this site to see if a movie is worth the cost of a ticket. The genre variables were very interesting as well. Horror and comedy films increased box office revenues in my regression. I attribute this to the nature of film audiences. Couples often see horror or comedy movies on date nights. It is also reflective of human nature. People like the adrenaline rush of a horror or a thriller, and the endorphin release that laughter causes.

The star power variable was tricky to get data for. A more reliable and accurate measure would likely increase the coefficient and the significance of the variable because I think it has a larger effect on revenues then my study lets on. The summer and holiday months were consistently positive, and this could be expected because students have breaks during this time and workers tend to take vacation days during these months. In this free time, people can see movies with their families. The previous audience and original gross variables are important, and I was surprised that they were not higher and more significant. There are so many remakes, sequels, movies based on books, and mega film franchises these days that have built-in audiences. In looking at the top 10 grossing films of each year, the majority had a previous audience, and several were part of a huge franchise such as Marvel or Star Wars.

Award nominations and word-of-mouth transfer are similar variables to critical reviews. These both reflect quality and the audience's perception of a film. Audience perception is not something that can really be predicted. There are many films that try to create rapport with audiences, but there are generally only a few rare movies that audiences can genuinely connect with. That is the art aspect of films and no variable can really capture that. Recently, the newest and final installment of Marvel's Avengers franchise was released in theaters. The franchise began in 2008 and audiences all over the world connected with these films on a deep level. After seeing the film many moviegoers left in tears and some were completely distraught after seeing some of their favorite characters for the last time. In 11 days, the film has grossed over $2 billion dollars in total and may become the highest grossing film of all time, passing Titanic. No variable can account for this kind of success.

References

Bagella, M. and Beccheti, L. 1999. "The determinants of motion picture box office performance: evidence from movies produced in Italy." *Journal of Cultural Economics* 23, 237-256.

Brewer, Stephanie M, Jason M. Kelley, and James J. Jozefowicz. 2009. "A blueprint for success in the US film Industry." *Applied Economics* 41, 589-606.

Deuchert, E., Adjamah, K., and Pauly, F. 2005. "For Oscar glory or Oscar money?" *Journal of Cultural Economics* 29, 159-176.

De Vany A. 2004. "Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry" *Hollywood Economics* 6.

Ravid, S. A. 1999. "Information, blockbusters, and stars: a study of the film industry." *The Journal of Business* 4, 463-492.

Ravid, S. S. 2004. "Are they all crazy or just risk averse? Some movie puzzles and possible solutions." *Contributions to Economic Analysis* 260, 33-47.

Simonoff, J. S. and Sparrow, I. R. 2000. "Predicting movie grosses: winners and losers, blockbusters and sleepers." *Chance*, 13, 15-24.

Smith, S. P. and Smith, V. K. 1986. "Successful movies: a preliminary empirical analysis." *Applied Economics*, 18, 501-507.

Walls, W. D. 2004. "Revenues, profitability, and returns: clinical analysis of the market for mobster films." *International Journal of Business and Economics*, 2, 93-106.