The Undermined Power of the Invested Minority

Jo Ozaki

University of Akron

Department of Economics

Senior Project

Spring 2021

Abstract

This study builds upon existing literature to examine the effect expert (critic) reviews relative to electronic word of mouth (user reviews) on film performance. This study looks to provide a modern perspective by analyzing films from 2018, compared to the most recent literature which analyzed films released in 2008. Using a mixed effects model, it was found that user reviews and critic reviews are of similar significance on film performance. All four of the linear regression models are capable of explaining over ninety percent of the variation within the data. I find that critics also play roles as an influencer while their role as a predictor seems to be limited suggesting an change in the critics roles over a decade. User review volume or publicity was also found to be significant. Finally, it was shown that there was statistical and slight economic significance in appealing to the casual user or consumer as opposed to the critic, suggesting that films that are designated to casual consumers may have greater film performance all else held equal.

Table of Contents

I. Introduction

Within any market the concept of consumer heterogeneity or having different types of consumers is a crucial aspect to consider when making any decision. Whether it be politics, people searching for a target market segment, or when designing a product, consumer heterogeneity is generally thought of in terms of left or right of politics or will the person by the product or not. However, on top of this dimension of consumer heterogeneity, I propose that the investment of the consumer is also important when deciding who to market towards. In the marketplace consumers lie on a wide spectrum on another dimension: their level of investment as well. This includes everyone from the extremely invested, who spend hours and days a week thinking, anticipating and using a product, to the casual consumer who purchased and enjoys at a surface level once. Ignoring this spectrum of consumer heterogeneity is common, and generally, products are simply made to appeal to the mass and casual audience, whether that be in politics with the median voter model, or movies in the film industry, and even in the target audience of competitive fighting games.

Although this topic is important, it is hard to quantify and measure and thus the literature concerning level of investment of consumers is extremely limited. One industry however, has explored the literature of invested versus casual consumers which is the film industry and how user and critic reviews affect and influence film performance. In this paper, I look to the film industry with critics and casual consumers, and if it is profitable to mainly appeal to critics as opposed to prioritizing the casual mass audience as seen in the market.

A large reason as to why the median voter model as well as companies typically market their product towards the casual audience is a question of sheer quantity. Since the casual audience market is far bigger, it seems like common sense that it would be more profitable to appeal to the casual audience. This includes exceptions such as brands that look to market to only to those invested within the industry such as expensive clothing brands as well as crowdfunding programs or even flexible pricing methodologies. In this paper we will only be focusing on the traditional scenario where everyone is charged the same amount for each unit of the product purchased, which more similarly represents the theoretical political model where every “voter” has only one vote, or where every individual consumer purchase is worth the same to the firm.

There are a couple of reasons as to why it may be profitable to appeal to the invested rather than the casual audience. Firstly, critics and the more invested will generally expend more time and effort than the casual audience to support or criticize a certain good or service which affect other consumers, whereas the casual audience is less likely to engage in such efforts. Furthermore, critics and the more invested, such as professionals within an industry often have a large influence over the general casual public and their opinions are highly regarded. It is said that more than 1/3 people actively look for opinions of critics when deciding to watch a movie, and there has even been a case where Sony Entertainment Studios created a fake critic to leave positive reviews for their films to boost their image and ultimately the success of their movies. (Basuroy et al., 2003)

This paper finds a couple of results. First is that critic and consumer both have statistically significant effects on film performance, however, the effect of critic reviews are not found to be more important than user reviews contrary to the findings of Basuroy et al. (2021). This could be due to the transition of how consumers utilize, and view critic reviews as opposed to user reviews in modern times, as films in this paper are sourced from 2018, while the most recent literature utilizes and analyzes films from 2008 (Basuroy et al., 2021). It is also seen that critics function as influencers, and not as predictors of film performance building upon the literature of Eliashaberg et al. (1997).

The remainder of this paper is organized as follows. The next section provides a summary and background of previous literature and findings. Section 3 discusses the data and variables analyzed within the paper. Section 4 introduces the theoretical discussion. Section 5 explains the empirical methodology. Section 6 reviews regression results and findings. Section 7 concludes with a discussion of academic and managerial implications of the findings.

II. Literature Review

When considering the consumer base, there exists multiple types of spectrums when it comes to differentiating consumers, those who buy and do not, and there is also the distinction between those who are invested in the product, and casual buyers. This stems from the economic theory of consumer heterogeneity and how consumers are divided within invested/critics and casual/general audience (Shi, 2013). Many studies have been investigated if the influence and motivation of the invested minority can lead to profitability over the casual audience. Critics and casual consumers can influence the potential buyers which is explained by the information integration theory that states that attitudes are formed and modified when individuals receive and interpret experiential information with their own existing beliefs or attitudes (Eliashberg et al., 1997). There has even been evidence that suggest that people place such a significant weight on other people's opinions that they may even ignore their own private information (Banerjee, 1993).

There are three approaches to analyzing a film’s performance that the recent literature on this topic explores, which include the Economic approach, the Information Theory approach, and the Communication Theory approach to measure film performance (Hadida, 2009). The

Economic approach focuses on factors of production, revenues and other quantifiable variables, the Information Theory approach looks at the influence of Non-Expert and Critics reviews, and the Communication Theory approach delves into the context and characters of the film. The majority of research has relied on secondary data, however, there have also been studies of experimental nature where participants read artificially created reviews in an attempt to measure the effect of consumer and critic reviews (Tsao, 2014).

The most prominent approach in the literature is the Economic approach, utilizing Box Office Revenues as the key measure for film performance (Hadida, 2009). Various variables, such as production budget, screenings, or number of theaters a movie aired in a given week, have been explored and concluded to have a statistically and economically significant effect on film performance. Discussion about even non-linear relationships have been proposed in some cases, such as the relationship between lead actors and film performance. (Basuroy, 2003). Additional variables considered in this literature explaining film performance are budget, critical reviews, awards, director's past films performance, star power, and movie rating (Basuroy, 2003). However, the lack of variables and factors from the Information Theory and Communication Theory approaches has led to criticism that the Economic Approach has failed to bridge the gap between the artistic dimension and the commercial aspect. Hadida (2006) reviews 135 academic papers and finds only 6 that have adequately accounted for both the artistic and commercial determinants of a film's performance (Hadida, 2006).

Although few prior papers account for the variables from the Communication and Information Theory approaches adequately, there have been studies that surround consumer review and word of mouth (WOM). One study, found that in a study utilizing Yahoo! Movies and its consumer reviews, the inclusion of a WOM variable, reduced forecasting errors of box office revenue by 23 percent (Liu, 2006). There are also other papers that support the argument that WOM is significant in film performance measurement include Duan et al. (2008) and Chen et al. (2004). The literature, overall, has no consensus in its conclusions with studies suggesting that the information overload of consumer reviews leads to selective processing and biased conclusions since consumers on average read only 10 reviews (Alzate et al., 2021). Moreover, 80 percent of consumers worry about the authenticity of WOM with 25 percent of reviews on Yelp claimed to be suspicious (KRC Research, 2012). Finally, research of the implementation of WOM as a measure of film performance shows that, the number of consumer reviews are found to play a significant role which could indicate that the publicity of a film, good or bad, could be an underlying factor and valance, or actual score of the review of WOM is not significant (Duan et al., 2008). Chen et al. (2004) and Goes and Mayzlin (2004) also have challenged the significance of WOM and there appears to be no clear consensus within the literature.

The influence of critic reviews on film performance has been explored significantly deeper than WOM, as Hadida (2009) finds 28 papers in support with only 5 papers challenging the statistical significance of critics as of 2009. Boor and Myron (1992) find that critics generally tend to agree with each other more than WOM with a 91.9 percent probability that 4 out of 6 critics agree, 41.4 percent probability that 6 out of 6 critics agree, whereas consumer reviews tend to be more random. Another study that was done concluded that during this digital age and information overload, expert identity and critic reviews have a significant impact on website performance for tourism, and overall consumer decision and satisfaction (Zhang et al., 2016). Finally, one of the main concerns and criticism of the inclusion of critic review on film performance is whether critics merely look to predict film performance as an expert or have an impact as influencers. This is important because if critics merely act as a predictor and reflect consumer opinions, this will lead to high multi-collinearity between the WOM variable and bias results. Critics have been found to act as both influencers and predictors by analyzing Box Office Revenues for early weeks from releases where critics possess asymmetric information to consumers and their influence is the greatest (Basuroy, 2003).

This paper plans to contribute to and build upon existing literature analyzing films utilizing IMDB which is ranked second among movie rating sites by the website ranking company, Alexa.[1] We also look to analyze data for films released within 2018, which had the most films released within the film industry history of North America at 873 films (Box Office Mojo, 2021). Compared to recent literature, where the films analyzed by Basuroy et al. (2020), of which the films were taken from 2008, this study looks to provide a more modern accurate depiction of the interaction of WOM and critic reviews in the current age on film performance. This study also looks to account for seasonality, which has been proved to be extremely significant in film revenues based upon time of release (Liran, 2007). This is done using information from Vogel's sophisticated measure of seasonality, which depicts normalized weekly attendance over the year (Vogel, 2001). Finally, further analysis will be done for films with positive critic reviews, which had mixed or negative reviews initially to gauge whether it is profitable to appeal to critics by comparing initial responses, which has not been done within existing literature.

[1] http://www.alexa.com/topsi tes/category/Top/Arts/Movies

III. Data

The data for this paper has been sourced from the widely recognized and approved IMDB, for every variable within the dataset. This decision to use IMDB over Rotten Tomatoes as done by Basuroy et al. (2020) has two justifications. Firstly, the critic rating system utilized by Rotten Tomatoes, operates on a approve or disapprove basis and looks to merely average the number of critics that approve and disapprove each given film. There is a major flaw with this system as it is incapable of capturing the degree of approval for a given film under the guise of taking the average of all critics. This poses a large problem especially for smaller films where less critic reviews are available, and the data becomes generalized. IMDB on the other hand, sources its critic reviews from Metacritic which operates on a numeric 1-to-100-point rating, capturing nuanced preferences and beliefs of a film. Secondly, IMDB through the utilization of IMDBpro has access to a multitude of different metrics of performance and facts about a film, many of which have been used as variables within this study. IMDB also provides side by side analysis of user reviews and critic ratings for direct comparisons as Basuroy et al. (2020) states are a benefit of rottentomatoes and has been stated to be significant by Alzate et al. (2021) for proper assessment of the impact of user and critic reviews.

All the variables used within this paper have been sourced from IMDB and the critic reviews through Metacritic.com which are displayed and featured on IMDB. User reviews, volume, and variance was gathered using Anaconda in conjunction with Python to scrape the user review from a scale of 1 to 10 stars, as well as the date from a static website and applying the appropriate formulas and transformations in excel. Similar averaging and counting functions were used to create the variables, critic rating and volume, after sourcing the rating and date from the Metacritic link available in IMDB. Data gathering for star power was done manually by counting the number of actors within the top 10, 100 and 1,000 for each film, respectively. For this paper, Star Power Top 1000 is used, and is a variable counting the number of actors within the top 1,000 due to the fact more normally distributed and less skewed. On top of this, successful movies still have more actors within the top 1,000 similar to the difference in top 10 or 100 allowing for a more nuanced parameter. The remaining variables were all readily available using IMDB and IMDBpro. To account for the possibility of human error, data quality has been assured with the review of 20 percent of the films that were randomly chosen.

According to the descriptive statistics in Table 2, the descriptive statistics for the box office revenues has an extremely large range due to the inclusion of major blockbuster movies, included Black Panther and the Avengers. *Box Office Revenue*, for example has a mean of 11,031 with a max of 338,333, the regression model will analyze the log transformation of box office revenues for a more appropriate measurement. Similarly, *User Review Volume*, *Theaters*, *Budget*, *Number of Awards and Number of Nominations*, as well as *Star Power Top 1,000* have high maximums in comparisons to their mean values, due to the major hits of the year, which act as outliers within the dataset. We can see from the data that Users on average rate movies higher than critics where the mean for *Weekly User Reviews* is 7.01 out of 10 stars and the mean for *Final Critic Rating* is 65.88 out of 100 points. The variable *User Variance* is included to measure the variance of user reviews from week to week and was borrowed from Basuroy et al. (2020). This variable is defined as:

$$UserVariance_{it} = \frac{\sum_{\tau=1}^{t} weekUserVolume_{i\tau}(weekUserReview_{i\tau} - UserReview_{it})^2}{UserVolume_{it}}$$

IV. Theoretical Discussion

The two main focuses of this paper are the relationship of the dependent variable, *Weekly Box Office Revenues* and the key independent variables, weekly user rating and final critic rating. The relationship between User review and box office revenues is expected to be positive, as the more positive the user reviews for a film are, the higher the box office revenue is expected to be. The same is expected for critic review, as a higher critic review is expected to have higher box office revenue for a film. Given the suspicion that positive critic reviews are influenced by the film industry pressures, it is expected that negative critic reviews are more significant than positive critic reviews. For both user and critic review, due to the negativity effect, it is expected that:

H1 Negative reviews from both users and critic reviews is expected to have a larger effect on box office revenue.

Finally, it has been found that consensus is the most influential external cue for potential movie-goers, thus the following hypothesis is expected to be true. (Eliashaberg et al., 1997)

H2 The effect of consumer review on box office revenue is moderated by the critic ratings and vice versa to be true as well.

For critics and their influence film performance via box office revenues, there are two different fields of thoughts presented by Eliashaberg et al. (1997). First, critics serve as an opinion leader or influencer because they are seen as the expert within the industry and are typically able to review films even before release and have asymmetrical information compared to consumers. As earlier research shows, consumers value the opinions of experts and the invested, critics possess the ability to influence consumers action and affect box office revenues especially in the early weeks of film release where asymmetric information is at its largest. The second field of thought is that, because critics have dedicated their lives to watching and evaluating films, they have developed the ability to accurately predict how a film will perform.

This theory suggests that critics merely represent the views of consumers and that critics do not affect film performance, and thus critic review is expected to be correlated with box office revenue in later weeks. Considering Critics roles as both Predictors and Critics, the following hypothesizes is expected to be true.

H3 If critics are influencers, critical reviews are correlated with box office revenue in the first few weeks only, and not with box office revenue in the later weeks or with the entire run.

H4 If critics are predictors, critical reviews are correlated with box office revenue in the later weeks and the entire run, not necessarily with box office revenue in the earlier weeks where the consumer consensus has yet to be formed.

Lastly, as to whether critic or consumer reviews are expected to have a dominant effect on film performance is varied. The information overload theory states that the overwhelming volume of consumer reviews leads to selective processing and biased conclusions and consumer reviews will not have a significant effect on box office revenues compared to critic reviews. And secondly, due to the influx of information due to the digital age and the information overload that has resulted, consumers are more interested in the expert's opinions on films than ever before, as critic opinions are in agreement in general more than consumers.

On the other hand, it has also been proposed that due to the fact that critics review films as a living and are the most invested within the consumer base, they have different criteria when it comes to evaluating films that do not align with the interests of consumers. Critics tend to value the deeper meaning, sophistication, consistency, camera work and various other aspects, that not only may not be important but may be considered boring to casual consumers thus proposing that critic reviews are expected to have a less significant effect or even a negative effect on film performance.

Thus, due to the contradicting theories the following hypotheses are capable of being true:

H5 User Reviews have a larger impact on box office revenues than critic reviews.
H6 Critic Reviews have a larger impact on box office revenues than user reviews.
H7 Positive Critic Rating have a negative impact on film performance.

V. Empirical Methodology

The empirical models for this paper are mainly borrowed from Basuroy et al. (2020) with the addition of various control variables to account for omitted variable bias as this paper looks to conduct a linear regression model to measure the effects of user versus critic reviews on film performance. This paper will run four different regression models: Model 1 or Critic and User model will include variables to both critic and user aspects, Model 2 or User Model including $UserVolume_{it}$, $UserReview_{it}$, $UserVariance_{it}$, Model 3 or Critic Model including $CriticReview_{i}$, $CriticVolume_{i}$, Model 4 or Critic Vs User is a model that includes the *CriticVsUser* variable, which is a variable that measures the difference between final critic rating and final user rating respective to their average. This goes against the most recent literature where Basuroy et al. (2020) relies on a panel data analysis. The rationale for this decision is as follows. A main variable of interest here is critic reviews, but this variable is relatively constant over time for any given movie: over 90 percent of critic reviews are released prior to film release, and the rest following within the first week or two of movie release. Due to the relative invariability of critic reviews, the model would suffer from multicollinearity if movie fixed effects are used. Thus, this study looks to build upon the research methodology of Eliashberg et al. (1997) and Basuroy et al. (2003) utilizing a linear regression model while building upon the model presented by Basuroy et al. (2020).

Model 1: Critic and User Model

$$Log(Box_{it}) = Const + \beta_1 Log(WeeklyUserVolume_{it}) + \beta_2 WeeklyUserReviews_{it} + \beta_3 Log(UserVariance_{it}) + \beta_4 CriticReviewVolume_i + \beta_5 EarlyWeek * FinalCriticRating_i + \beta_6 LaterWeek * FinalCriticRating_i + \beta_7 X_{it} + \beta_8 Y_i + \beta_9 Z_t + \varepsilon$$

The main dependent variable will be measured in the form of *Log(Box$_{it}$)* which is the log of *Weekly Box Office Revenues* for ith movie, on tth week. The log of the box office revenue is analyzed since the levels of this variable is highly skewed, and the use of log values for the dependent variable allows for a better economic interpretation of coefficients.

For user reviews, I collect individual user reviews from the IMDB website with the review score and date, using Anaconda in conjunction with Python to scrape the data for online user reviews. *UserVolume$_{it}$* represents the number of movies reviews to date for each week similar to Liu (2006), Chintagunta et al. (2010), and Basuroy et al. (2020). UserReview$_{it}$ is the average rating of users who have reviewed the movie for each week. And finally, *UserVariance$_{it}$* is a measure of variance using average ratings within each week (weighted by the respective number of reviews within the week) relative to the average of all user ratings up to and including week *t* which is borrowed from Basuroy et al. (2020) as demonstrated below.

CriticVolume$_i$ measures the total number of critic reviews that was sourced from Metacritic and available side by side to user reviews on the IMDB website for each film. *Final Critic Rating* is separated using interaction terms for early and later weeks following the work of Eliashaberg in 1997. This is done to account for the relationship between critic reviews and box office revenue for Critic's possible different roles, influencer and/or predictor, that critics may play on film performance. This was proposed by Eliashaberg that Critics will act as influencers in earlier weeks due to the larger asymmetric information and predictors in later weeks.

X_{it} indicates other control variables included in the model that are variable over time and across movies: theaters, or the total number of theaters a film is being played at in any given week.

Y_i includes the control variables that are different for each film, but constant over time. This includes, number of awards a film has received, number of nominations a film has received, production budget for a film, and a variable that accounts for leading actors utilizing the Star Power available on IMDB. Star Power is a measure of popularity of an actor, founded from the millions of users and interactions IMDB receives on their website. This variable will also have a squared term, as it has been seen in previous literature to have a non-linear effect on box office revenues Basuroy et al. (2003).

Z_t includes the control variable that are constant over films, but different in respect to time. This variable includes a set of dummy variables to account for seasonality accounting for the different seasons which have been said to have a drastically different levels of box office revenues by both Liran (2007) and Vogel (2001).

VI. Results

According to the correlation matrix in Table 3, most variables have low correlation values amongst each other. This is apart from number of theaters a film is played in and the *Weekly Box Office Revenue* which is to be expected because the number of theaters a film is played in is directly linked to the performance of a film in a given week. This result is similar to the findings of Basuroy et al. (2021) and Eliashaberg et al. (1997). The other instance of a high correlation level is the final critic rating and the number of nominations and awards a film has received which once again is to be expected as a critically acclaimed film is expected to receive more nominations and awards all thing held constant.

Figure 1 is the depiction of the residuals versus the estimated values of the dependent variable in Model 1. This model which includes variables for both critics and user reviews. Upon analysis we can see that there are some outliers in the lower levels of the estimated *Weekly Box Office Revenue*. These are more niche films that do not perform as well as on average and have a higher variability in film performance compared to Blockbuster films. However, even with the inclusion of these outliers in the lower performing *Weekly Box Office Revenues* the residuals look to behave in a random way with not particular pattern, and issues of normality. Although heteroskedasticity seems to be a slight problem for lower levels of *Weekly Box Office Revenues*, the model was calculated using heteroskedastic robust standard errors and thus this is a problem that is accounted for in the estimation procedure.

The same models without heteroskedastic robust standard errors were run for a smaller dataset with 54 films from 2018 which exclude the more niche movies as well which can be seen in Table 5. And upon observing Figure 3, the amount of variation for lower levels of box office revenue is decreased. On top of this there does not appear to be a pattern in the way the Model 1 for 54 films fails to estimate the *Weekly Box Office Revenue* which is also a plus. And upon analyzing the parameter estimates from Table 5, the results are similar to the Models with all 72 films and thus it is assumed that the estimations are appropriate.

Figure 2 depicts the estimated and actual values of *Weekly Box Office Revenue* levels based upon the various parameters within the model: with the estimated values being depicted as red dots, and the actual *Weekly Box Office Revenue* levels in black dots. We can see that the distribution of the red dots very closely follows the distribution of the actual *Weekly Box Office*

Revenue levels and there appears to be no systematic pattern in how the model fails to accurately estimate Weekly Box Office Revenue.

Model 1 includes parameter estimates for both critic and user review. Within the model there are a couple of important things to note. Firstly, the parameter estimates for *Weekly User Review Volume* is positive and statistically significant ranging from .134 to .292, stating that if the number of Weekly User Reviews were to double, *Weekly Box Office Revenue* is expected to increase 13.4 percent to 29.2 percent. *Weekly User Review Volume* is statistically significant in every single model analyzed within this paper which supports the literature that publicity plays an important role in film performance (Duan et al., 2008). The parameter for number of *Theaters, Number of Nominations* received, and *Star Power Top 1000* were all positive and statistically significant, which is in accordance with the literature, as an increase in any of these variables, all else held constant, is expected to yield higher film performance. The most significant being *Theaters* where a doubling in the number of theaters a film is shown in for any given week is expected to increase *Weekly Box Office Revenue* anywhere from 109.7 percent to 120.3 percent, in other words, more than double film performance. This is in accordance with past literature and makes sense as doubling the number of theaters a film is shown in should also around double film revenue (Eliashaberg, 1997). The inclusion of the dummy variables for seasons was account for any season-specific variables that would affect the *Weekly Box Office Revenue* of all movies similarly.

As for the two crucial parameters estimates, firstly as discussed in the methodology section, an interaction term between week from movie release and final critic rating was created in order to account and allow the possibility of critics functioning as influencers and or predictors in film performance. Critics are seen to have a positive and statistically significant effect on film performance in the earlier weeks of release, supporting hypothesis #3 that critics can influence consumer decisions in the earlier weeks of film release where asymmetric information is at its greatest. The parameter estimate states that for every increase in ten points out of the hundred scale for the final critic review average, film performance is expected to increase 13 percent as well. As for the parameter estimate for final critic review in the later weeks, it is not only statistically insignificant, but also holds little to no economic value, suggesting that critics do not act as predictors of a film's success. Although this contradicts the findings of Basuroy et al. (2003), this is supported within the literature as well, stating that because critics are so invested, they value different aspects of a film, which may contribute to a boring or non-enjoyable film to the casual audience, and thus those reviews will not serve as accurate predictions of how a film will perform in the later weeks (Chen et al., 2004; Goes and Mayzlin, 2004). Model 2 and 3 analyze the significance of User Review and Critic review separately, and although there is slight change in the parameter estimates of, the results are similar to Model 1. Finally, Model 4 looks to address the research question of whether it is more profitable to market a product towards the invested or the casual audience directly by creating an interaction term between final critic review and final user review of a film, and their difference in percentages from their respective average ratings for all films within 2018. This parameter measures the difference between critic review and consumer review where a critically acclaimed film and a film with poor user reviews will receive a higher value for this interaction term and vice versa.

The parameter estimate is positive and statistically significant which indicates that it is more profitable to create a product that is more targeted towards the casual audience. Specifically, for every 1 percentage points increase in difference between final critic rating and final user rating, film performance will decrease by 0.6 percent on average.

VII. Conclusion

Due to the rapid and ever-growing power of the internet and readily available user reviews, it seems that, although in the short run, critics may have an influence over film performance, overall, critics do not act as predictors of film performance, nor are the most appealing group to market towards in modern times contrary to the findings of earlier papers. (Eliashaberg et at., 1997; Basuroy et al., 2020). This could be attributed to the fact that the most recent study done by Basuroy et al. in 2020 shows a depiction of the interaction of critic and user review on film performance in 2008 which undoubtably has evolved over the course of a decade to 2018. Even more significant than user reviews however were the user volume which could imply that publicity is the main driver of film performance rather than user or critic review which is in accordance with previous literature as well (Duan et al., 2008). Furthermore, analyzing Model 4 and the *CriticVsUser* variable, it was found that it is better to market a film towards the casual consumer rather than the invested minority or critics if film performance maximization in the long term is desired. This does not support the initial hypothesis proposed in the paper and could indicate that. Although further research and implementation needs to be done in other industries, the current standard of appealing to the masses seems to be profit maximizing strategy for films in the modern age.

VIII. References

Alzate, Miriam, Marta Arce-Urriza, and Javier Cebollada. "Online Reviews and Product Sales: The Role of Review Visibility." *Journal of Theoretical and Applied Electronic Commerce Research* 16.4 (2021): 638–669. *Crossref*. Web.

A. Banerjee, The economics of rumours, Review of Economic Studies 60 (1993) 309–327.

Austin, B.A. (1984). Portrait of an art film audience. *Journal of Communication*, **34**(Winter), 74–87.

Basuroy, Suman, et al. "How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets." Journal of Marketing, vol. 67, no. 4, (2003), 103–117., www.jstor.org/stable/30040552..

Boor, Myron. "Relationships among Ratings of Motion Pictures by Viewers and Six Professional Movie Critics." Psychological Reports, vol. 70, no. 3_suppl, June 1992, pp. 1011–1021, doi:10.2466/pr0.1992.70.3c.1011.

Box Office Mojo. "Number of Movies Released in The United States and Canada from 2000 to 2020." *Statista*, Statista Inc., 17 Jan 2021, https://www-statista-com.ezproxy.uakron.edu:2443/statistics/187122/movie-releases-in-north-america-since-2001/

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie

box office performance: Accounting for sequential rollout and aggregation across local markets.

Marketing Science, *29*(5), 944–957.

De Vany, A. and Walls, W.D. (1999). Uncertainty in the movie industry: does star power reduce the terror of the box office? *Journal of Cultural Economics*, **23**, 285–318.

D. Godes, D. Mayzlin, Using online conversations to study word of mouth communication, Marketing Science 23 (4) (2004) 545–560 Fall.

Dillio, Richard, "Different Scores: Video Gamers' Use of Amateur and Professional Reviews" (2013). Thesis. Rochester Institute of Technology.

Einav, Liran. "Seasonality in the U.S. Motion Picture Industry." *The RAND Journal of Economics*, vol. 38, no. 1, 2007, pp. 127–145. *JSTOR*, www.jstor.org/stable/25046296. Accessed 18 Feb. 2021.

Eliashberg, Jehoshua, and Steven M. Shugan. "Film Critics: Influencers or Predictors?" Journal of Marketing, vol. 61, no. 2, (1997), 68–78., www.jstor.org/stable/1251831. Accessed 14 Feb. 2021.

Ginsburgh, V. andWeyers, S. (1999). On the perceived quality of movies. *Journal of Cultural Economics*, **23**, 269–283.

Hadida, Allègre L. "Motion Picture Performance: A Review and Research Agenda." International Journal of Management Reviews 11.3 (2009): 297–335..

Hadida, A.L. (2003). *Strategic Assets, Institutional Factors and Performance: An Application of the Resource Based View and of New Institutional Eco- nomics to Cinema Projects in France and the United States (1988–1997)*. Jouy en Josas: HEC.

Hadida, A.L. (2004). *Reputation Resources, Commit- ment and Performance of Film Projects in the USA and Canada (1988–1997),* WP 03/2004. Cambridge, UK: Judge Business School.

KRC Research. (2012). *Buy it, try it, rate it*. Available at, https://www.webershandwick.com/uploads/news/ files/ReviewsSurveyReportFINAL.pdf. Accessed August 2, 2015.

Michael P. Keane and Nada Wasi, 2013. "The Structure of Consumer Taste Heterogeneity in Revealed vs. Stated Preference Data," Economics Papers 2013-W10, Economics Group, Nuffield College, University of Oxford. (Keane and Wasi, 2013)

Neelamegham, R., & Chintagunta, P. (1999). A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Science*, *18*(2), 115–136.

P.-Y. Chen, S.-Y. Wu, J. Yoon, The impact of online recommendations and consumer feedback on sales, Proceedings of the International Con-ference on Information Systems, 2004, pp. 711–724, ICIS.

Rozin, Paul, and Edward B. Royzman. "Negativity Bias, Negativity Dominance, and Contagion." *Personality and Social Psychology Review*, vol. 5, no. 4, Nov. 2001, pp. 296–320, doi:10.1207/S15327957PSPR0504_2.

Suman Basuroy & S. Abraham Ravid & Richard T. Gretz & B. J. Allen, 2020. "Is everybody an expert? An investigation into the impact of professional versus user reviews on movie revenues," Journal of Cultural Economics, Springer;The Association for Cultural Economics International, vol. 44(1), pages 57-96, March.

Shi, Hongyan, et al. "Consumer Heterogeneity, Product Quality, and Distribution Channels." Management Science, vol. 59, no. 5, 2013, 1162–1176., www.jstor.org/stable/23443933.

Tsao, WC. Which type of online review is more persuasive? The influence of consumer reviews and critic ratings on moviegoers. Electron Commer Res 14, (2014) 559–583. https://doi.org/10.1007/s10660-014-9160-5

Vogel, Harold L. (2001), *Entertainment Industry Economics*, 5th ed. Cambridge, UK: Cambridge University Press.

Wenjing Duan, Bin Gu, Andrew B. Whinston, Do online reviews matter? — An empirical investigation of panel data, Decision Support Systems, Volume 45, Issue 4, 2008, Pages 1007-1016,

Y. Chen, S. Fay, Q. Wang, Marketing implications of online consumer product reviews, Working paper, department of marketing, University of Florida, 2004.

Y.Liu, Word of mouth for movies: its dynamics and impact on box office revenue, Journal of Marketing 70 (July 2006) 74–89.

Zhang, Julie & Zhang, Zili & Yang, Yang. (2016). The power of expert identity: How website-recognized expert reviews influence travelers' online rating behavior. Tourism Management. 55. 10.1016/j.tourman.2016.01.004.

Zhang, X., & Dellarocas, C. (2006). The lord of the ratings: Is a movie's fate influenced by reviews? In *ICIS*

2006 proceedings, 1959–1978.

IV. Appendix

Table 1: List of variables used in the analysis, their definitions, references, and sources

Variables	Variable Definition	Literature Support
$Log(Box_{it})$	Main dependent variable. Box office revenue of film *i* in week *t*.	Basuroy et al. (2003), Liu (2006)
$UserVolume_{it}$	Exogenous variable. Number of user reviews for film *i* by week *t*.	Liu (2006), Zhang and Dellarocas (2006), Chintagunta et al. (2010)
$UserReview_{it}$	Exogenous variable. Average user rating of film *i* by week *t*	Liu (2006), Zhang and Dellarocas (2006), Chintagunta et al. (2010)
$UserVariance_{it}$	Exogenous variable. Variance of average user ratings of film *i* by week *t*.	Zhang and Dellarocas (2006), Chintagunta et al. (2010)
$CriticVolume_{i}$	Number of critic reviews for film I by week *t*.	Zhang and Dellarocas (2006)
$CriticRating_{i}$	Exogenous variable. Average critic rating of film *i* by week *t*.	Basuroy et al. (2003), Liu (2006), Zhang and Dellarocas (2006), Gopinath et al. (2013) Eliashaberg et al. (1996)
$CriticVsUser_{i}$	Exogenous variable. Measures difference between Critic and User Review from each other respective to their averages.	Basuroy et al. (2020)
$Theaters_{it}$	Exogenous variable. Number of theaters a film *i* airs in week *t*.	Basuroy et al. (2003), Liu (2006), Chintagunta et al. (2010), Gopinath et al. (2013)
$Budget_{i}$	Exogenous variable. Total amount of money spent in production of movie *i*.	Basuroy et al. (2003), De Vany and Walls (2002), Sorenson and Waguespack (2006)
$NumofOscars_{i}$	Exogenous variable. Number of Oscars received by film *i*.	Ginsburgh and Weyers (1999), Hadida (2003), Ginsburgh (2003)
$NumofAwards_{i}$	Exogenous variable. Number of awards received by film *i*.	Ginsburgh and Weyers (1999), Hadida (2003), Ginsburgh (2003)
$NumofNominations_{i}$	Exogenous variable. Number of nominations received by film *i*.	Ginsburgh and Weyers (1999), Hadida (2003), Ginsburgh (2003)
$StarPowerTop1000_{i}$	Exogenous variable. Number of actors within the top 1000 for the IMDB star power rating for film *i*.	Eliashberg (1996), Neelamegham and Chintagunta (1999)
$Genre_{i}$	Exogenous variable. Genre of a film *i*.	Austin (1981), (1984), Litman (1983), Eliashberg *et al.* (2001)
$Rating_{i}$	Exogenous variable. Rating certificate from MPA for film *i*.	Sawhney and Eliashberg (1996), De Vany and Walls (2002), Basuroy *et al.* (2003), Ravid and Basuroy (2004)
$Seasonality_{t}$	Exogenous variable. Dummy variable for each month of the year for week *t*.	Sorenson and Waguespack (2006) (Vogel, 2001)

Source: IMDB and Metacritic with own calculations

Table 2 Descriptive Statistics

Variables	Mean	SD	Min	Max
Box(in $1000s)$_{it}$	11,031	28,254	1,342	338,333
Weekly User Volume$_{it}$	466.47	568.28	5	2744
Weekly User Review$_{it}$	7.01	1.11	4.23	9.25
User Variance$_{it}$	0.055	0.123	0	1.695
Critic Review Volume$_{i}$	45.83	9.23	4	69
Final Critic Rating$_{i}$	65.88	16.84	6.6	93
Critic Vs User$_{it}$	3.84	20.06	-69.00	84.40
Theaters$_{it}$	1,469.41	1,473.33	1.00	4,485.00
Budget(in $1000s)$_{i}$	67,971	70,565	880	321,000
Number of Oscars$_{i}$	0.264	0.727	0	4
Number of Awards$_{i}$	19.917	31.872	0	182
Number of Nominations$_{i}$	52.778	66.865	1	331
Star Power Top 1,000$_{i}$	3.903	3.643	0	28

Source: IMDB and Metacritic with own calculations

Note: Includes 72 films all from 2018

Critic Vs User is a variable that measures the difference between final critic rating and final user rating respective to their average. A higher value for the variable Critic Vs User indicates that the film received better Final Critic Ratings in comparison to Final User Ratings.

Table 3: Correlation Matrix

	LogWeeklyGrossBoxOff	WeeklyUserRev	FinalCriticRating	LogWeeklyUserVol	LogUserVariance2	WeeklyCriticRevVol	LogTheaters	LogBudget	LogNumofOscars	LogNumofAwards	LogNumofNominations	LogStarPowerTop.1000
Log Weekly GrossBox Off	1	0.295	0.123	0.725	-0.214	0.225	0.927	0.272	0.171	0.242	0.278	0.216
Weekly UserRev	0.295	1	0.148	0.068	-0.397	0.025	0.210	-0.031	0.252	0.278	0.251	0.120
FinalCritic Rating	0.123	0.148	1	0.092	0.007	0.446	0.021	-0.244	0.299	0.711	0.670	0.101
Log Weekly UserVol	0.725	0.068	0.092	1	-0.187	0.300	0.625	0.253	0.224	0.320	0.318	0.037
LogUser Variance	-0.214	-0.397	0.007	-0.187	1	0.001	-0.120	-0.053	-0.083	-0.057	-0.047	-0.013
Weekly CriticRev Vol	0.225	0.025	0.446	0.300	0.001	1	0.148	0.183	0.354	0.594	0.728	0.257
Log Theaters	0.927	0.210	0.021	0.625	-0.120	0.148	1	0.325	0.091	0.116	0.156	0.190
LogBudget	0.272	-0.031	-0.244	0.253	-0.053	0.183	0.325	1	0.037	-0.170	-0.029	0.338
LogNumof Oscars	0.171	0.252	0.299	0.224	-0.083	0.354	0.091	0.037	1	0.647	0.570	0.044
LogNumof Awards	0.242	0.278	0.711	0.320	-0.057	0.594	0.116	-0.170	0.647	1	0.882	0.116
LogNumof Nominations	0.278	0.251	0.670	0.318	-0.047	0.728	0.156	-0.029	0.570	0.882	1	0.152
LogStar Power Top.1000	0.216	0.120	0.101	0.037	-0.013	0.257	0.190	0.338	0.044	0.116	0.152	1

Source: IMDB and Metacritic with own calculations

Note: Correlation matrix with log transformations included to show interaction of variables within model

Table 4: Regression Results for 72 movies

	Dependent variable: Log of Weekly Gross Box Office Revenue			
	Critic and User (1)	User Only (2)	Critic Only (3)	Critic Vs User (4)
Log of Weekly User Volume	0.134^{***}	0.292^{***}		0.249^{***}
	(0.038)	(0.047)		(0.042)
Weekly User Review	0.090^{***}	0.130^{***}		
	(0.018)	(0.024)		
Log of User Variance	-0.010	-0.019		-0.029^{**}
	(0.012)	(0.012)		(0.012)
Critic Review Volume	-0.012^{***}		-0.013^{***}	-0.018^{***}
	(0.004)		(0.004)	(0.005)
Final Critic Rating for Early Weeks	0.013^{***}		0.013^{***}	
	(0.003)		(0.003)	
Final Critic Rating Later Weeks	-0.003		-0.005^{**}	
	(0.002)		(0.002)	
Critic Vs User				-0.006^{***}
				(0.002)
Log of Theaters	1.097^{***}	1.162^{***}	1.172^{***}	1.203^{***}
	(0.037)	(0.044)	(0.024)	(0.037)
Log of Budget	0.056^{**}	-0.051	0.064^{**}	-0.033
	(0.026)	(0.036)	(0.025)	(0.037)
Log of Num of Oscars	-0.092	-0.088	-0.060	-0.130
	(0.077)	(0.098)	(0.083)	(0.100)
Log of Num of Awards	0.084^{**}	0.010	0.145^{***}	0.059
	(0.042)	(0.047)	(0.041)	(0.047)
Log of Num of Nominations	0.183^{***}	0.090^{**}	0.189^{***}	0.212^{***}
	(0.045)	(0.041)	(0.044)	(0.046)
Log of Star Power Top 1000	-0.002	0.029	-0.038	-0.002
	(0.049)	(0.071)	(0.049)	(0.065)
Seasons: Jun-Aug	0.211^{***}	0.220^{***}	0.212^{***}	0.317^{***}
	(0.060)	(0.063)	(0.060)	(0.066)
Seasons: Mar-May	-0.059	-0.052	-0.090	0.013
	(0.065)	(0.070)	(0.065)	(0.073)
Seasons: Sep-Nov	-0.092	-0.105	-0.123^{*}	-0.047
	(0.070)	(0.072)	(0.069)	(0.071)
Constant	4.934^{***}	5.484^{***}	5.449^{***}	6.209^{***}
	(0.433)	(0.574)	(0.394)	(0.517)
Observations	720	720	720	720
Adjusted R^2	0.953	0.943	0.951	0.942
Residual Std. Error	0.458	0.514	0.471	0.518
Robust residual standard error:	0.4581	0.5138	0.4714	0.5179

Source: IMDB and Metacritic with own calculations

Notes: 1. *p<.1, **p<.05***, p<0.01

2. Robust standard errors are in parentheses

3. Critic Vs User is a variable that measures the difference of critic review and the average critic review rating in 2018, user review and the average user review rating in 2018 and takes the difference of the two after adjusting user reviews to a 1–100-point scale by multiplying by 10 and taking the difference to see the total difference in evaluation from critics and user reviews for a given film

Table 5: Regression Results for 54 movies

	Dependent variable: Log of Weekly Gross Box Office Revenue			
	(1)	(2)	(3)	(4)
Log of Weekly User Volume	0.210^{***}	0.378^{***}		0.355^{***}
	(0.029)	(0.029)		(0.028)
Weekly User Review	0.081^{***}	0.114^{***}		
	(0.018)	(0.020)		
Log of User Variance	-0.028^{***}	-0.037^{***}		-0.052^{***}
	(0.011)	(0.012)		(0.012)
Critic Review Volume	-0.007		-0.010^{**}	-0.016^{***}
	(0.005)		(0.005)	(0.006)
Final Critic Rating for Early Weeks	0.010^{***}		0.009^{***}	
	(0.002)		(0.002)	
Final Critic Rating for Later Weeks	-0.002		-0.008^{***}	
	(0.002)		(0.002)	
Critic Vs User Final Review Rating				-0.004^{***}
				(0.002)
Log of Theaters	1.116^{***}	1.145^{***}	1.207^{***}	1.165^{***}
	(0.023)	(0.026)	(0.021)	(0.026)
Log of Budget	-0.014	-0.115^{***}	0.038	-0.109^{***}
	(0.027)	(0.030)	(0.029)	(0.031)
Log of Num of Oscars	0.075	0.110	0.055	0.104
	(0.083)	(0.092)	(0.088)	(0.095)
Log of Num of Awards	0.012	-0.061	0.127^{***}	-0.026
	(0.043)	(0.045)	(0.044)	(0.047)
Log of Num of Nominations	0.148^{***}	0.128^{***}	0.171^{***}	0.223^{***}
	(0.047)	(0.045)	(0.050)	(0.054)
Log of Star Power Top 1000	0.191^{***}	0.237^{***}	0.132^{**}	0.274^{***}
	(0.054)	(0.061)	(0.057)	(0.063)
Seasons: Jun-Aug	0.269^{***}	0.314^{***}	0.305^{***}	0.417^{***}
	(0.072)	(0.076)	(0.077)	(0.082)
Seasons: Mar-May	0.041	0.101	0.045	0.165^{**}
	(0.068)	(0.076)	(0.073)	(0.079)
Seasons: Sep-Nov	-0.00000	0.058	0.017	0.118
	(0.071)	(0.079)	(0.076)	(0.082)
Constant	5.382^{***}	6.016^{***}	5.433^{***}	6.788^{***}
	(0.470)	(0.504)	(0.477)	(0.503)
Observations	540	540	540	540
Adjusted R^2	0.938	0.918	0.928	0.915
Residual Std. Error	0.517	0.595	0.557	0.604
F Statistic	543.786^{***}	502.319^{***}	579.665^{***}	449.882^{***}

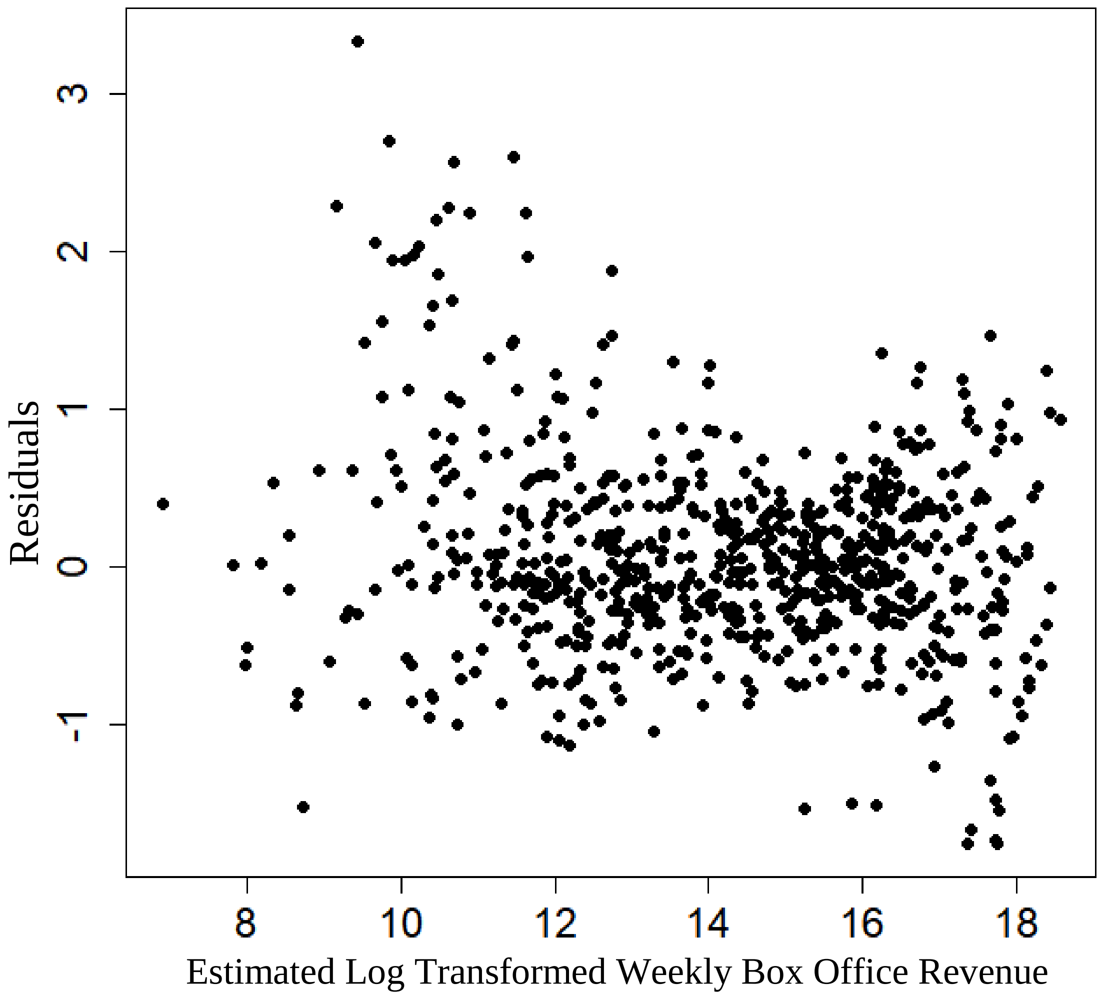
Source: IMDB and Metacritic with own calculations

Note: 1. $^{*}p<.1$, $^{**}p<.05$, $^{***}p<0.01$

2. The variable "Log of User Variance" is a variable that measures the variation of weekly user reviews from all previous user reviews for film i

3. Critic Vs User is a variable that measures the difference of critic review and the average critic review rating in 2018, user review and the average user review rating in 2018 and takes the difference of the two after adjusting user reviews to a 1–100-point scale by multiplying by 10 and taking the difference to see the total difference in evaluation from critics and user reviews for a given film

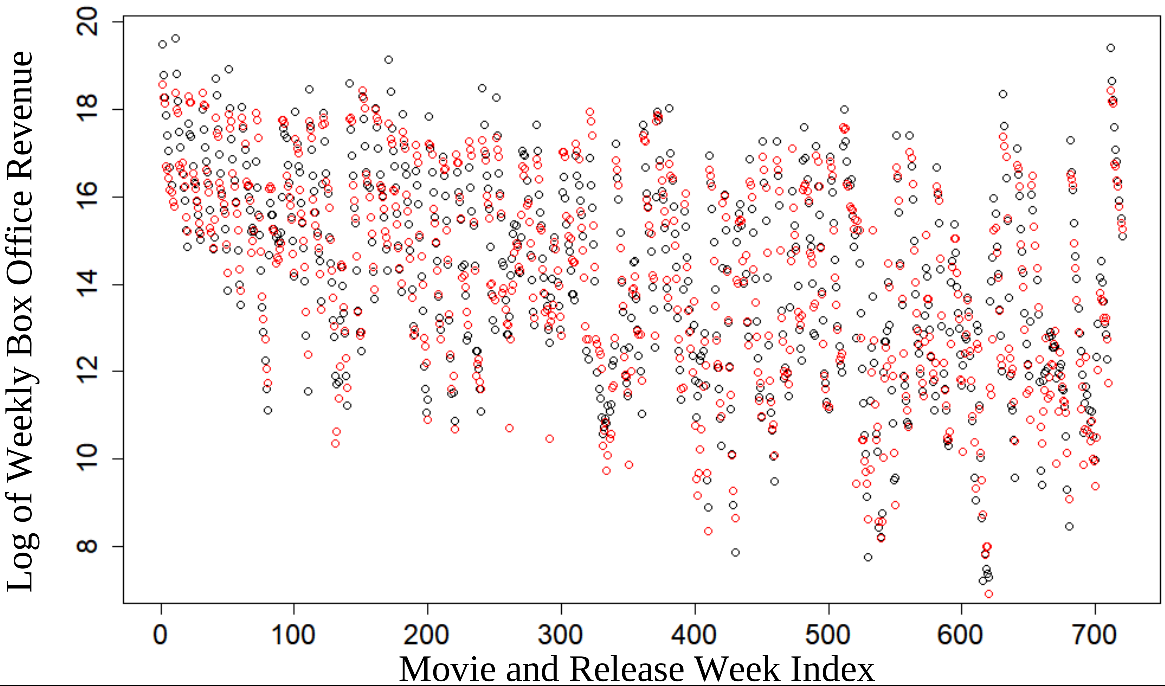
Figure 1: Residuals of Model 1 Estimates for 72 Films



Source: IMDB and Metacritic with own calculations

Note: Includes 72 films all from 2018

Figure 2: Model 1 Estimated VS Actual Box Office Revenue for 72 Films



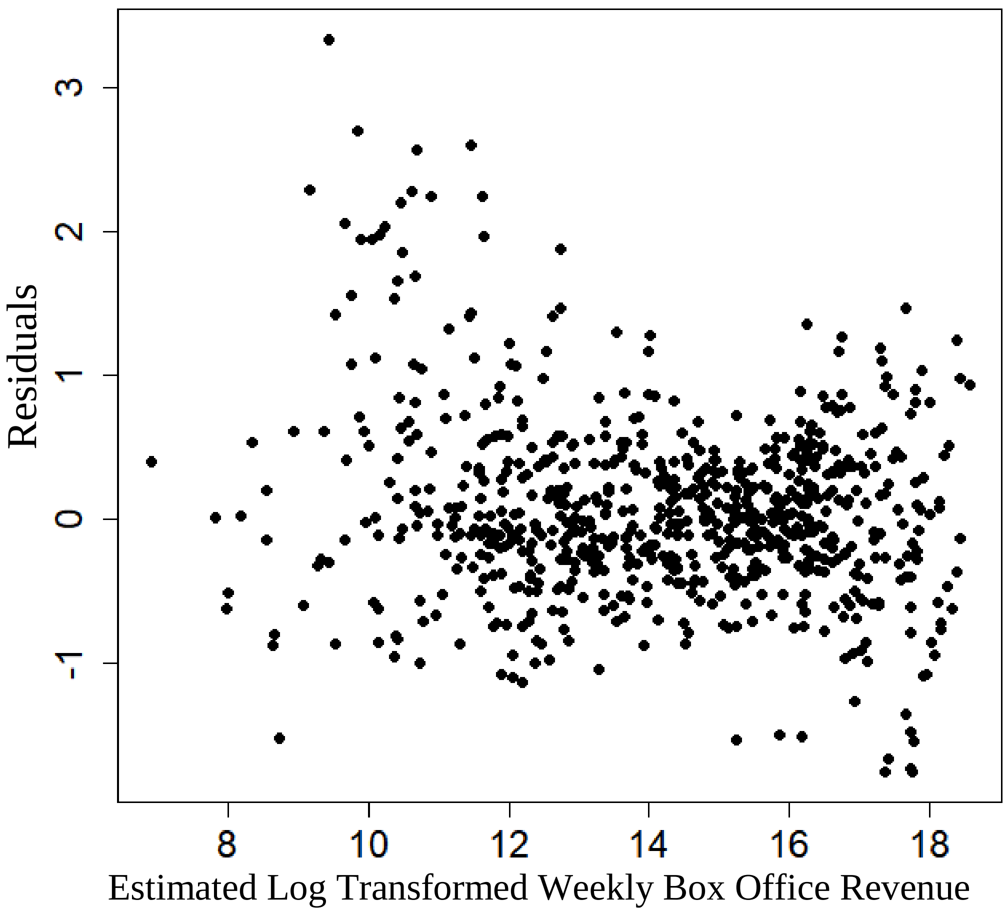
Source: IMDB and Metacritic with own calculations

Note: Includes 72 films all from 2018

Red dots: Predicted Box Office Revenue

Black dots: Actual Box Office Revenue

Figure 3: Residuals of Model 1 Estimates for 54 Films



Source: IMDB and Metacritic with own calculations

Note: Includes 54 films all from 2018

Figure 4: Model 1 Estimated VS Actual Box Office Revenue for 54 Films

Source: IMDB and Metacritic with own calculations

Note: Includes 72 films all from 2018

Red: Predicted Box Office Revenue

Black: Actual Box Office Revenue

X. Coding

Senior Project Coding

```
##Final Coding From Start to Finish
#Reading data set on PC
SPdata <- read.csv("C:/Dropbox/Jojo/Akron-Sp21/DATASP1.csv")
SP <- data.frame(SPdata)

#Creation and transformation of various different variables
SP$FinalCriticRatingSq <- SP$FinalCriticRating^2
SP$StarPowerTopSq.10 <- (SP$StarPowerTop.10)^2
SP$StarPowerTopSq.100 <- (SP$StarPowerTop.100)^2
SP$StarPowerTopSq.1000 <- (SP$StarPowerTop.1000)^2
#SP$LogStarPowerTopSq.1000 <- (SP$LogStarPowerTop.1000)^2
SP$January <- as.numeric(SP$Seasonality== "January")  # as.numeric translates to numerical - here from logical
SP$February<- as.numeric(SP$Seasonality== "February")  # as.numeric translates to numerical - here from logical
SP$March <- as.numeric(SP$Seasonality== "March")  # as.numeric translates to numerical - here from logical
SP$April <- as.numeric(SP$Seasonality== "April")  # as.numeric translates to numerical - here from logical
SP$May <- as.numeric(SP$Seasonality== "May")  # as.numeric translates to numerical - here from logical
SP$June <- as.numeric(SP$Seasonality== "June")  # as.numeric translates to numerical - here from logical
SP$July<- as.numeric(SP$Seasonality== "July")  # as.numeric translates to numerical - here from logical
SP$August <- as.numeric(SP$Seasonality== "August")  # as.numeric translates to numerical - here from logical
SP$September <- as.numeric(SP$Seasonality== "September")  # as.numeric translates to numerical - here from logical
SP$October  <- as.numeric(SP$Seasonality== "October")  # as.numeric translates to numerical - here from logical
SP$November <- as.numeric(SP$Seasonality== "November")  # as.numeric translates to numerical - here from logical
SP$December <- as.numeric(SP$Seasonality== "December")  # as.numeric translates to numerical - here from logical
```

```
SP$UserVariance <- (SP$WeeklyUserVol*(SP$WeeklyUserRev -
SP$FinalReviewRating)^2)/SP$CumulativeUserRevVol

SP$LogWeeklyGrossBoxOff <- log(SP$WeeklyGrossBoxOff)
SP$LogBudget <- log(SP$Budget)
SP$LogNumofAwards <- log(SP$NumofAwards+1)
SP$LogNumofOscars <- log(SP$NumofOscars+1)
SP$LogNumofNominations <- log(SP$NumofNominations+1)
SP$LogStarPowerTop.10 <- log(SP$StarPowerTop.10+1)
SP$LogStarPowerTop.100 <- log(SP$StarPowerTop.100+1)
SP$LogStarPowerTop.1000 <- log(SP$StarPowerTop.1000+1)
SP$LogCumulativeUserRevVol <- log(SP$CumulativeUserRevVol)
SP$LogTheaters <- log(SP$Theaters)
SP$LogUserVariance <- log(SP$UserVariance)
SP$LogUserVariance2 <- log(SP$UserVariance+.0001)
SP$LogWeeklyUserVol <- log(SP$WeeklyUserVol+1)
SP$LogTheaters <- log(SP$Theaters+1)

SP$AvgDiffCritic <- SP$FinalCriticRating - 62.1
SP$AvgDiffUser <- 10*(SP$WeeklyUserRev - 6.75)
SP$CriticVsUser <- SP$AvgDiffCritic - SP$AvgDiffUser

SP$EarlyWeek <- ifelse(SP$Week percentinpercent 1:3, 1, 0)
SP$LaterWeek <- ifelse(SP$Week percentinpercent 4:10, 1, 0)

SP$Time1 <- SP$Week

library(pastecs)
#stat.desc(SP)

##### Histograms for all the variables in my dataset

hist(SP$WeeklyGrossBoxOff)
hist(SP$LogWeeklyGrossBoxOff)
hist(SP$Rank)
hist(SP$LogRank)
```

```
hist(SP$Budget)
hist(SP$LogBudget)
hist(SP$NumofOscars)
hist(SP$NumofAwards)
hist(SP$LogNumofAwards)
hist(SP$NumofNominations)
hist(SP$StarPowerTop.10)
hist(SP$StarPowerTop.100)
hist(SP$StarPowerTop.1000)
hist(SP$LogStarPowerTop.1000)
hist(SP$LogStarPowerTop.10)
hist(SP$LogStarPowerTop.100)
hist(SP$LogStarPowerTopSq.1000)
hist(SP$FinalCriticRating)
hist(SP$FinalReviewRating)
hist(SP$Rating)
hist(SP$WeeklyCriticRevVol)
hist(SP$WeeklyUserRev)
hist(SP$WeeklyUserFinal)
hist(SP$WeeklyUserVol)
hist(SP$FinalCriticAvgComp)
hist(SP$FinalRevAvgComp)
hist(SP$CumulativeUserRevVol)
hist(SP$LogCumulativeUserRevVol)
hist(SP$UserVariance)
hist(SP$LogUserVariance)
hist(SP$LogUserVariance2)
hist(SP$Theaters)
hist(SP$LogTheaters)
hist(SP$LogTheaters)
hist(SP$WeeklyUserVol)
hist(SP$LogWeeklyUserVol)

install.packages("dplyr")
install.packages("tidyr")
install.packages("broom")
library(broom)
```

```
# Installing package and running and creating linear regression with heteroskedastically robust standard errors
install.packages("robustbase")
library(robustbase)

Model1 <- lmrob(LogWeeklyGrossBoxOff ~ LogWeeklyUserVol + WeeklyUserRev + LogUserVariance2 + CriticRevVol + (FinalCriticRating:EarlyWeek) + (FinalCriticRating:LaterWeek) + LogTheaters + LogBudget + LogNumofOscars + LogNumofAwards + LogNumofNominations + LogStarPowerTop.1000 + Seasons, data = SP)
#User Review Version
Model2 <- lmrob(LogWeeklyGrossBoxOff ~ LogWeeklyUserVol + WeeklyUserRev + LogUserVariance2 + LogTheaters + LogBudget + LogNumofOscars + LogNumofAwards + LogNumofNominations + LogStarPowerTop.100 + LogStarPowerTop.1000 + Seasons, data = SP)
#Critic Version
Model3 <- lmrob(LogWeeklyGrossBoxOff ~ (FinalCriticRating:EarlyWeek) + (FinalCriticRating:LaterWeek) + CriticRevVol + LogTheaters + LogBudget + LogNumofOscars + LogNumofAwards + LogNumofNominations + LogStarPowerTop.1000 + Seasons, data = SP)
#Critic vs. User version
Model4 <- lmrob(LogWeeklyGrossBoxOff ~ CriticVsUser + LogWeeklyUserVol + LogUserVariance2 + CriticRevVol + LogTheaters + LogBudget + LogNumofOscars + LogNumofAwards + LogNumofNominations + LogStarPowerTop.1000 + Seasons, data = SP)

# Mixed Effects model for 54 films with normal standard errors

Model1 <- lm (LogWeeklyGrossBoxOff ~ LogWeeklyUserVol + WeeklyUserRev + LogUserVariance2 + CriticRevVol + (FinalCriticRating:EarlyWeek) + (FinalCriticRating:LaterWeek) + LogTheaters + LogBudget + LogNumofOscars + LogNumofAwards + LogNumofNominations + LogStarPowerTop.1000 + Seasons, data = SP)
#User Review Version
Model2 <- lm (LogWeeklyGrossBoxOff ~ LogWeeklyUserVol + WeeklyUserRev + LogUserVariance2 + LogTheaters + LogBudget + LogNumofOscars + LogNumofAwards + LogNumofNominations + LogStarPowerTop.100 + LogStarPowerTop.1000 + Seasons, data = SP)
#Critic Version
Model3 <- lm (LogWeeklyGrossBoxOff ~ (FinalCriticRating:EarlyWeek) + (FinalCriticRating:LaterWeek) + CriticRevVol + LogTheaters + LogBudget + LogNumofOscars + LogNumofAwards + LogNumofNominations + LogStarPowerTop.1000 + Seasons, data = SP)
#Critic vs. User version
Model4 <- lm (LogWeeklyGrossBoxOff ~ CriticVsUser + LogWeeklyUserVol + LogUserVariance2 + CriticRevVol + LogTheaters + LogBudget + LogNumofOscars + LogNumofAwards + LogNumofNominations + LogStarPowerTop.1000 + Seasons, data = SP)

# To export Tables and regression models into downloadable word tables
```

```
install.packages("stargazer")
library(stargazer)
stargazer(Model1, Model2, Model3, Model4, title="Regression Results for 54 movies", align =
TRUE,type="html",out="Regression_Results5.doc")
getwd()

# Creation of correlation matrix and exporting it into a table format for downloading

correlation.matrix <- cor(attitude[,c("rating","complaints","privileges")])
stargazer(correlation.matrix, title="Correlation Matrix")

correlation.matrix <-
cor(SP[,c("LogWeeklyGrossBoxOff","WeeklyUserRev","FinalCriticRating","LogWeeklyUserV
ol","LogUserVariance2","WeeklyCriticRevVol","LogTheaters","LogBudget","LogNumofOscars
","LogNumofAwards","LogNumofNominations","LogStarPowerTop.1000")])
stargazer(correlation.matrix, title="Correlation
Matrix",type="html",out="Correlation_Matrix.doc")
install.packages("robustbase")
library(robustbase)

# Graphs for Presentation slide
#Plotting lbox office revenues for log explanation
par(mfrow=c(2,1))
plot(SP$Week,SP$WeeklyGrossBoxOff, main="Plot of Weekly Box Office Revenue by
Week",xlab="Week", ylab="Weekly Box Office Revenue")
plot(SP$Week,SP$LogWeeklyGrossBoxOff, main="Plot of Log of Weekly Box Office Revenue
by Week",xlab="Week", ylab="Log of Weekly Box Office Revenue")

plot(x,f_x,pch=16,lty=1,ylim=c(0,10),col="blue",main="User Reviews for Aqua
Man",xlab="Week (t)", ylab="Movie (4)")
lines(x[order(x)], f_x[order(x)], xlim=range(x), ylim=range(f_x), pch=16,col="blue")
plot(x,f_y,pch=16,lty=1,ylim=c(0,100),col="red",main="Final Critic Rating for Black
Panther",xlab="Week (t)", ylab="Movie (i)")
lines(x[order(x)], f_y[order(x)], xlim=range(x), ylim=range(f_y), pch=16,col="red")

#Plotting residual of model as well as against predicted vs. actual box office revenue


plot(Logmodel2$resid)
plot(Logmodel2$fitted, Logmodel2$resid)


#Python Coding in order to scrape movie data for User Reviews from IMDB.com


###
```

```
###
###  Web Scraping Template 002 - BS4 on IDMb
###
###    https://www.udemy.com/course/web-scraping-python-tutorial/learn/lecture/12934390#overview
###
##########################################

import requests
from bs4 import BeautifulSoup



movienames = ['12Strong', 'APrivateWar', 'AQuietPlace', 'ASimpleFavor', 'AStarIsBorn', 'AmericanAnimals', 'Ant-ManandtheWasp', 'Aquaman', 'Avengers_InfinityWar', 'BeautifulBoy',
'BlackPanther', 'Blockers', 'BohemianRhapsody', 'Bumblebee', 'Burning', 'CanYouEverForgiveMe', 'Capernaum', 'CrazyRichAsians', 'CreedII', 'Deadpool2', 'DeathWish', 'Destroyer',
'EighthGrade', 'FantasticBeasts', 'FirstMan', 'GameNight', 'GreenBook', 'Halloween', 'Hereditary', 'IfBealeStreetCouldTalk', 'Incredibles2', 'InstantFamily', 'IsleofDogs',
'JurassicWorld_FallenKingdom', 'LeaveNoTrace', 'LittleItaly', 'LoveSimon', 'Mamma_Mia!_Here_We_Go_Again', 'Mary_Poppins_Returns', 'Maze_Runner_TheDeathCure', 'Mid90s',
'Mission_Impossible-Fallout', 'OceansEight', 'OntheBasisofSex', 'OperationFinale', 'PacificRim_Uprising', 'RalphBreakstheInternet', 'Rampage', 'ReadyPlayerOne', 'RedSparrow',
'Searching', 'Shoplifters', 'Sicario_ DayoftheSoldado', 'Skyscraper', 'Solo_AStarWarsStory', 'Sorry to Bother You', 'Spider-Man_IntotheSpider-Verse', 'Stan&Ollie', 'Tag',
'The SistersBrothers', 'TheFavourite', 'TheHateUGive', 'TheHousewithaClockinItsWalls', 'TheMeg', 'TheMule', 'TheNun', 'ThePredator', 'TombRaider', 'Venom', 'Vice', 'Widows', 'Wildlife']



#movienames = ['BlackPanther', '12Strong', 'APrivateWar', 'AQuietPlace']




for moviename in movienames:

    print(moviename)
```

```
input_filename = moviename + '.html'
output_filename = '0_scraped/' + moviename + '.csv'

soup = BeautifulSoup(open(input_filename), 'html.parser')

# Use CSS selector
reviews = soup.select('div[class*="lister-item"]')
print('review length=' + len(reviews))
i=0
for review in reviews:
    rating_tag = review.select('svg[class*="ipl-icon"]+span')  # + means next
    rating     = rating_tag[0].text if rating_tag else "N/A"
    date       = review.select('span[class*="review-date"]')
    title      = review.select('a[class="title"]')
    reviewer   = review.select('span[class="display-name-link"]')
    i=i+1
    if i percent 2 == 0:
        the_line = (str(i) + ", " + rating + ", " + date[0].text +
                    ", " + reviewer[0].text[1:] + ", " + title[0].text[0:-1] + "\n")
        #print(the_line)
        with open(output_filename, 'a') as f: f.write(the_line)
```