How Is Gender of Top Billing Actor Affecting Cinematic Success of a Film?

Yiqing Zhu

The University of Chicago

Abstract

Films have been frequently scrutinized to discern how different genders are perceived and evaluated. While most former researches focus on the portrayal of different genders in the motion picture industry, this research is interested in the effect of gender of leading cast on cinematic success, which may lead to an empirical description of the gender disparity in the motion picture industry. Two models defining commercial success and aesthetic success are formulated and gender effect is observed while other variables acting as controls. A filtered dataset from IMDb is collected and the two models are fitted in both the whole dataset and subsets of it, to see the gender effect varying chronologically in different genres. The research finds that the gender of top-billing cast is statistically significant for both commercial success and aesthetic success of a film and it is easier for female-leading films to achieve commercial success and male-leading films to achieve aesthetic success.

Keywords: gender, film, commercial success, aesthetic success

How Is Gender of Top Billing Actor Affecting Cinematic Success of a Film?

Introduction

Films, acting a prominent role in popular culture, have been frequently scrutinized to discern how different genders are perceived and evaluated. Presumably, mainstream motion pictures largely reflect prevailing cultural attitudes about gender roles, norms, attitudes, and expectations (Rosen, 1973; Haskell, 1987), and also create sociocultural images (Millburn, Mather, & Conrad, 2000). Film industry exhibits what it itself believes (Simonton, 2004) through cinematic products and share it with the society at large. The worldwide dominance of Hollywood "blockbusters" has enlarged and extended this effect. Thus, detecting the biases and misconceptions especially regarding to gender in the motion picture industry has been an essential issue.

While most former researches relevant to this topic focus on the portrayal of different genders in the motion picture industry, this research is interested in the effect that gender of leading cast has on cinematic success, which may lead to an empirical description of the gender disparity in the motion picture industry. More specifically, this research intends to explore if gender acts as an impact factor of cinematic success in aspect of art and business, and if this impact varies among genres and changes over time.

To observe the effect of gender of top billing actor on cinematic success, two models defining commercial success and aesthetic success are formulated. While several indicators of commercial success and aesthetic success act as controlling variables in the models, gender of top billing actor is the main concern, and the estimated coefficient and statistical significance are examined to indicate the gender effect. A filtered dataset from IMDb is collected and the two models are fitted in both the whole dataset and subsets of it, to see if the gender effect varying

chronologically in different genres.

The research finds that the gender of top-billing cast is statistically significant for both commercial success and aesthetic success of a film and it is easier for female-leading films to achieve commercial success and male-leading films to achieve aesthetic success. The gender effect for commercial success is significant for Drama, Mystery, and Western films and is increasing, while the gender effect for aesthetic success is significant for Drama, Family, Mystery, Romance, and Thriller films, and is decreasing.

Literature Review

Cinematic Success

Beginning in nickelodeons as lucrative venture in popular entertainment, film is transformed into a serious art form which partly relied on film scholars and critics who could discuss the medium independent of marketing and box office (Baumann, 2001). Naturally, art and business are two antithetical categories films have been often assessed in, in other words, cinematic success consists of aesthetic success and commercial success (Landes, 2002; Hammad Afzal, 2016), which provides basic guidance for the evaluation of films.

Among empirical studies, several investigations have addressed the predictors of critical acclaim or movie awards (e.g., Zickar & Slaughter, 1999; Simonton, 2002, 2004c) and a very large literature has been focused on the factors indicating box office success of a film (e.g., Dodds & Holbrook, 1988; Litman & Kohl, 1989; Wallace, Seigerman, & Holbrook, 1993; Prag & Casavant, 1994; De Vany & Walls, 1999). A few researchers also examined both questions simultaneously (e.g., Simonton, 2005, 2009), however, no research has been conducted directly relevant to the relationship between gender and cinematic success.

Aesthetic success. "The aesthetic evaluation of artworks is, and always has been, a very controversial exercise." (Ginsburgh & Weyers, 2005) There are three broadly acknowledged methods to evaluate beauty: one decomposes an artwork into attributes and rate each (De Piles, 1708; Beardsley, 1958; Vermazen, 1975; Dickie, 1988, 1997), and some art philosophers "locate the ground of judgments of taste, not in some object which is the target of the judgment, but in the maker of the judgment" (Shiner, 1996), and others take "test of time" and "test of space" to examine beauty (Hume, 1757; Savile, 1982; Dickie, 1988, 1997; Budd, 1995).

Instead of analyzing the aesthetic characteristics or examining the temporal and spatial spread of a film, taking judgment from judges as the predictor of aesthetic success is a more available method. In the context of motion picture industry, the judges would be film critics and consumers with the appearance of the film ratings.

"Philosophers typically put the burden of proving quality on experts, while economists often argue that the actual choices made by consumers are a better measure." (Ginsburg, 2003) Actually, the views of consumers correlate positively with the opinions of film critics (Wanderer, 1970; Boor, 1992; Holbrook, 1999). Moreover, since individual judge may be prone to judgement errors and short-sighted and consumers largely outnumber critics especially in bad movies ratings, and consumer ratings displays a longer accumulating period, both during theatrical run and post-theatrical period, the consumer ratings appears to be valid as well as more reliable than critic ratings. Thus, the consumer rating would be the predictor of aesthetic success of films in this research.

Commercial success. Rather than artistic expression, financial performance has been the goal of "film industry" since its emergence, which is even intensified with the booming of the highly profitable "blockbusters". A quite large number of films are little more than elaborate "get

rich" schemes – "products replete with movie stars and special effects but sadly lacking in plot, dialogue, and characterization" (Simonton, 2005).

The commercial success can be evaluated by the actual profit a film made, however, due to proprietary information including cost of production and budget, the actual profit of a film is unavailable to public in most cases (Litman & Ahn, 1998) and researchers have to estimate the financial performance of a film by various criteria, for example, gross box office earnings or receipts (e.g., Sochay, 1994; Pat Topf, 2010), first weekend gross (e.g., Basuroy, Chatterjee, & Ravid, 2003; Simonton, 2005), the total length of the theatrical run (Sochay, 1994), distributor rental revenue. Though the film revenue encompasses several parts including box office revenue, DVD revenue, and television showing revenue, the box office revenue appears to be the best way to estimate the commercial success of a film at this time because the information is readily available and movie theaters are still accepted as the major source of revenue for a particular film.

Gender and Film

Gender schema and Social learning theory. Gender schema theory (Bem, 1981) describes how young children learn and internalize information about gender roles. Gender schemas are the "cognitive structures stored in memory that organize gender-related knowledge, beliefs, attitudes, and preferences" (Liben & Signorella 1993) that help children develop extensive networks of knowledge about gender (Bem, 1981). Gender schemas are important for the development of children's gender identities; however, they also are the source of gender stereotypes that lead to gender-stereotyped behaviors (Levy & Carter 1989; Nihlen & Bailey 1988).

Social learning theory (later called social cognitive theory; Bandura, 1986) explains how children learn specific attitudes and behaviors from the images and characters they encounter in the media. According to this theory, children learn cultural patterns of behavior through repeated observations of both actual models in their social environments, such as parents and teachers, and symbolic models in their social environments, such as those depicted in the media (Bandura 1969).

One important source for children developing gender schemas are the images and characters they encounter in the media, where films play a prominent role in. The amount of time children exposed to media each day and their increasing dependency on the media for information during the adolescent years clearly underscores the importance of the media as pervasive and influential socializing agents in the lives of many children and adolescents (Faber, Brown, & McLeod 1979; Signorelli 1997). Therefore, the portrayals of different genders in films has been a strong source of gender schemas for people especially for children and the potential imbalance of the portrayal may be of much problem.

Gender stereotypes in films. A certain number of researchers have examined the portrayals of different genders, especially women, in films using the content analysis method. Most of them have found gender hegemony present both behind and in front of the camera, which has not changed much over time (Downs, 1981; Barcus, 1983; Steinke, 1999, 2004; Smith & Marc, 2010; Smith, 2013;) and has strengthened societal stereotypes about women (Downing, Mohammadi, & Sreberny-Mohammadi, 1990).

In a 30-year study of television, Signorielli and Bacue (1999) found that women are underrepresented as compared to men. Witt (2000) found that approximately two-thirds of characters on television are male, a percentage that has stayed consistent since the 1950s. He also

reported in several studies during the 1980s and 1990s that "Most females on nighttime television are young, attractive, thin and ornamental. Most female characters are either under 35 or over 50. Middle-aged women are rare. Females are consistently placed in situations where looks count more than brains and helpless and incompetent behaviors are expected. Men are twice as likely as women to be shown as competent and able to solve problems. Gender stereotypes abound on television, with women depicted as sex objects more frequently than men, and men portrayed as inept when handling children's needs." Smith and Donnerstein (1998) found that in movies and on TV, males are often characterized as being prone to violent behavior, 44% of the time, perpetrators are attractive, and in 75% of the cases, they received no immediate punishment for the crime. Men's violence against women in the media is commonplace, and is often portrayed as a form of "heterosexual-based eroticism," in which women are seduced by masculine (and abusive) behavior (Hooks, 1994). Researcher expressed the concern that a steady diet of viewing these types of depictions "may send the message that girls are less valuable and capable than boys" (Smith & Marc, 2010).

The observed stereotyped depiction of different genders calls for future research especially on empirical evidence for the disparity. At the same time, it also indicates that genres may correlate with the gender-cinematic success relationship and should be taken into consideration.

Method

Model

In this research, two models will be fitted in, investigating the impact of gender on commercial success and aesthetic success of films respectively.

As concluded from former researches, significant predictors for commercial success of a film contains film ratings, whether the film has cinematic predecessors, budget, whether the film is released in holiday season or summer time, and genre (e.g., Dodds & Holbrook, 1988; Litman & Kohl, 1989; Wallace, Seigerman, & Holbrook, 1993; Prag & Casavant, 1994; De Vany & Walls, 1999). Thus, the commercial success of a film is obtained by

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \varepsilon_i \quad (1)$$

where Y is the domestic box office gross of a film, X_1 is a binary variable indicating whether the film is female leading or male leading, X_2 is the consumer rating, X_3 is a binary variable indicating whether the film has cinematic predecessors, X_4 is the budget of a film, X_5 is a binary variable indicating whether the film is released in December, June, July, or August, X_6 is a categorical variable indicating the genre of a film, X_7 is a categorical variable indicating the production year range of the film, ε_i is the error term.

Significant attributes for aesthetic success of a film include whether the film has cinematic predecessors, budget, runtime, genre, and characteristics of screenplay (e.g., Zickar & Slaughter, 1999; Simonton, 2002, 2004, 2005, 2009). In this research, the first four variables would be included in the model and since the dependent variable consumer rating is highly correlated with the number of voters, it would also be included in the model. Hence the model evaluating aesthetic success of a film is estimated as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7} + \varepsilon_i \quad (2)$$

where Y is the consumer rating of a film, X_1 is a binary variable indicating whether the film is female leading or male leading, X_2 is the number of voters, X_3 is a binary variable indicating whether the film has cinematic predecessors, X_4 is the budget of a film, X_5 is the runtime of a

film, X_6 is a categorical variable indicating the genre of a film, X_7 is a categorical variable indicating the production year range of the film, ε_i is the error term.

IMDb Data

The dataset applied to this research is obtained from Internet Movie Database (IMDb)[1], which can be accessed as compressed plain text files from the ftp sites or extracted using the Unix command-line interface tools[2].

IMDb is currently the world's most popular and authoritative source for movie, TV and celebrity content, offering a searchable database of more than 185 million data items including more than 3.5 million movies, TV and entertainment programs and 7 million cast and crew members. IMDb is a user-contributed encyclopedia, with information mainly gathered by people in the industry and entertainment fans around the world and constantly verified with studios and filmmakers through on-screen credits, press kits, official bios, autobiographies, and interviews.

The IMDb data is regarded as information source about both films and netizens and has been analyzed in a large number of literatures including researches on popular geopolitics of films (Dodds, 2006; Jung, 2012), recommendation systems (Lamprecht, 2015), and online social networking (Fatemi, Maryam, & Tokarchuk, 2012).

Since the United States film industry dominates the world market (Acheson & Maule, 1994), and cinematic products do not transport well across linguistic and culture boundaries (Lee, 2006), this research is confined among the American films. The currency difference, inflation, and different development of the commercialization of motion picture also restrict the validity of the evaluation commercial success of multinational films.

[1] http://www.imdb.com/
[2] http://www.imdb.com/interfaces

The IMDb dataset used in this research is updated on May 16th, 2017. Among 4,282,600 media products available in the dataset, 6,095 USA films are filtered which have complete and valid information on domestic box office gross, the gender of top-billing cast, consumer rating, remake / sequel information, budget, release date, genre, number of voters, runtime, and are produced before 2017. Overall 17,676 observations are included in the dataset considering that multiple genres can be assigned to a single film.

Variable

Domestic box office gross. The domestic box office gross information is provided by IMDb and the most recent cumulative gross is collected. In order to get rid of the impact of inflation, the dollar amount of gross is adjusted and lined up with the dollar value in 2016 with help of US Inflation Calculator[3] by using the latest Bureau of Labor Statistics (BLS)[4] inflation information provided in the Consumer Price Index (CPI)[5].

Gender of top-billing cast. The billing order of credits generally signifies their importance. The actors whose names appear first on credit are said to have "top billing", who usually play the principal characters in the film and have the most screen time (Wikipedia, 2017). Thus, a film with top-billing actor can be regarded as male-leading film and one with top-billing actress can be regarded as female-leading film. This is identified with cast gender and billing order of each film in the IMDb dataset.

IMDb rating & Number of voters. The IMDb rating is the weighted average of consumer rating on a scale of one to ten. IMDb has applied various filters to the raw ratings which are not disclosed to public in order to avoid ballot stuffing, and IMDb claims that the

[3] http://www.usinflationcalculator.com/
[4] https://www.bls.gov/cpi/home.htm
[5] http://www.usinflationcalculator.com/inflation/consumer-price-index-and-annual-percent-changes-from-1913-to-2008/

weighted vote average is a more accurate vote average. Each IMDb rating comes with both a numeric rating and the number of voters. A close glance at the number of voters versus rating data can be observed as below.

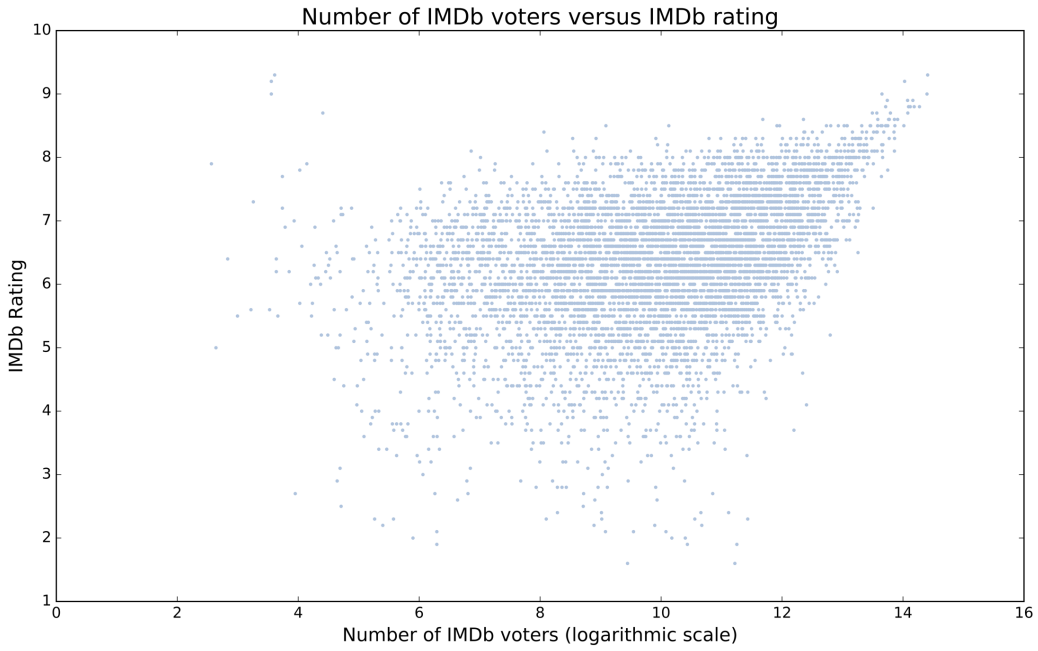

Figure 1

A strong pattern shows that films with approximately more than 100 votes (equals 4.6 at logarithmic scale) trend towards a higher rating with more votes. However, with fewer than 100 votes, there is little structure of the data. To add the reliability of IMDb ratings, the number of voters is included in the Model (2) and films with fewer than 100 votes (n = 61) are dropped from the dataset.

Cinematic predecessors. A film may bear a special relationship with one or more previous films. First, it may be a remake of an earlier film. Second, it may be a sequel to a previous film. Some research has found that sequels tend to be lower quality films (Prag & Casavant, 1994) but somewhat more successful with respect to film rentals (Litman & Kohl,

1989). Movie links information in IMDb dataset is examined carefully and films with the links "follows", "remake of", and "spin off from" are defined as cinematic predecessors available.

Budget. The budget figures on IMDb are estimates instead of exact amounts. IMDb claims that the budget numbers are based on media reports and often supplied by sources close to the production. Same as domestic box office gross, the budget amount included in the model is adjusted to 2016 dollar values.

Release season. Researchers have found that movies released in summer time and Christmas season are more likely to get a higher box office revenue. (e.g., Simonton, 2005) The release date in USA of each film is collected and films released in June, July, August, and December is marked as released in holiday season.

Runtime. On the average, the longer the film, the greater is its potential richness in terms of plot and characterization. The runtime collected in this research is the duration in minutes of the USA theatrical version of each film.

Genre. Some investigations have shown that a film's box office, awards, and critical assessment may partly depend on its genre (Litman, 1983; Prag & Casavant, 1994; Simonton, 2004). Dramas and comedies do not have the same likelihood of success, nor do romances or musicals. Yet the various studies do not agree on which genre have the highest odds of success, a disagreement that likely reflects the fact that they differ regrading whether the criteria are aesthetic or financial. Hence, a comparison among genres of the gender impact is a necessity. Since one film could be defined as several genres, in this research, one film might be repeated in different genres.

Production year. The information about production year of each film is collected from IMDb dataset. All films are subdivided chronologically into the interval of year 1913 – 1919,

year 1920 – 1929, year 1930 – 1939, year 1940 – 1949, year 1950 – 1959, year 1960 – 1969, year 1970 – 1979, year 1980 – 1989, year 1990 – 1999, year 2000 – 2009, and year 2010 – 2016.

Here are some statistical descriptions of the variables.

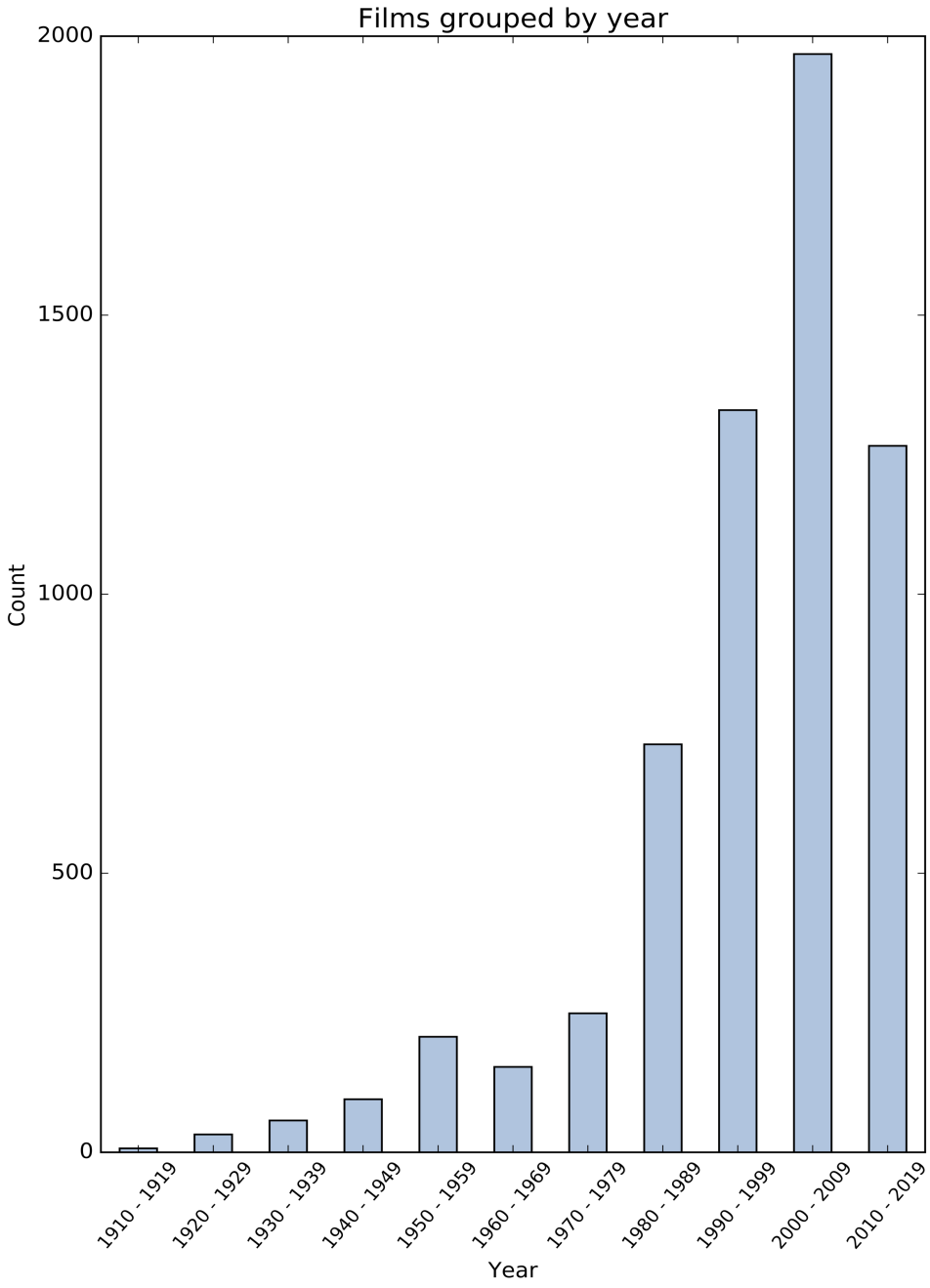

Figure 2

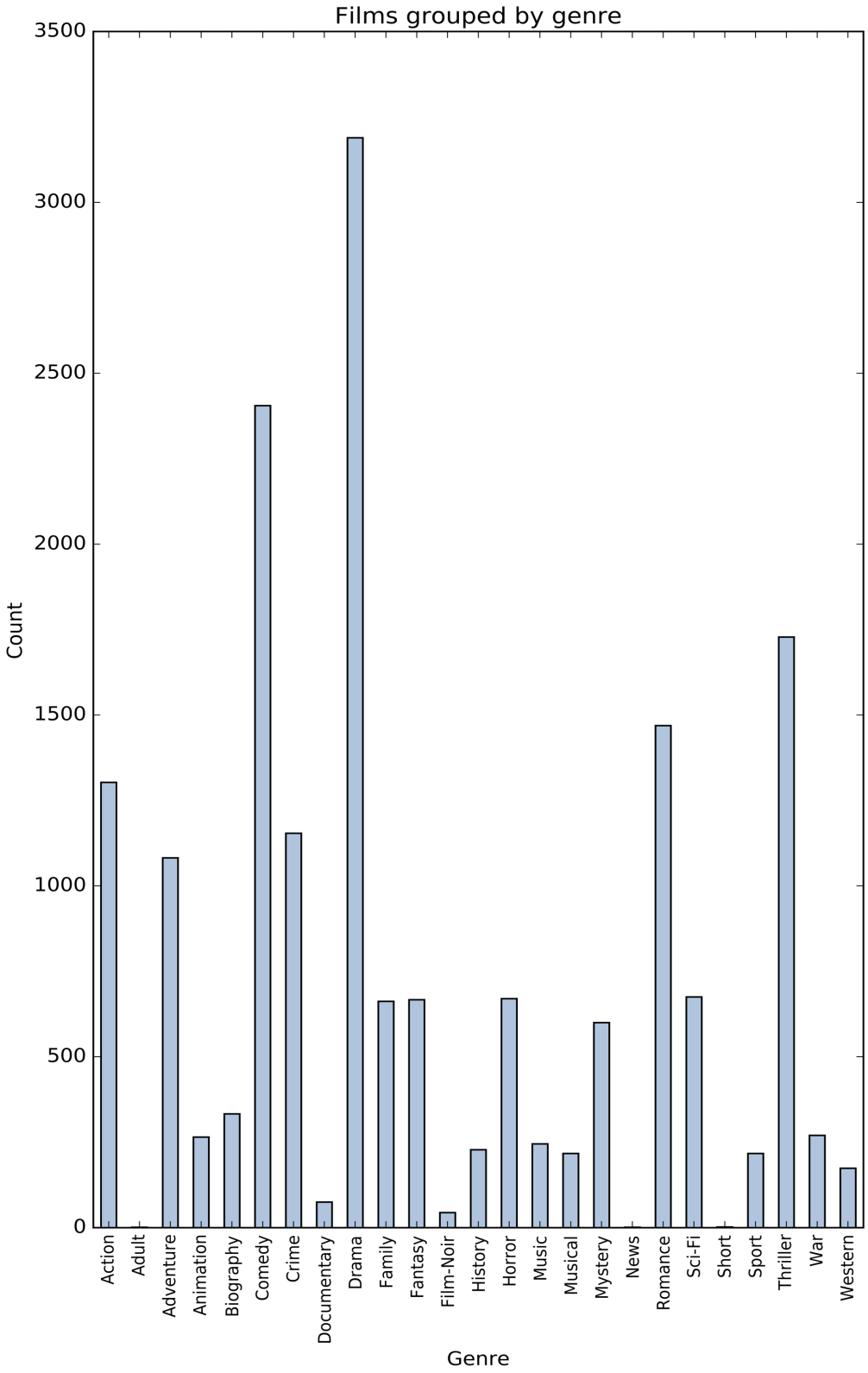
Films grouped by genre
Count
Genre
0
500
1000
1500
2000
2500
3000
3500
Action
Adult
Adventure
Animation
Biography
Comedy
Crime
Documentary
Drama
Family
Fantasy
Film-Noir
History
Horror
Music
Musical
Mystery
News
Romance
Sci-Fi
Short
Sport
Thriller
War
Western

Figure 3

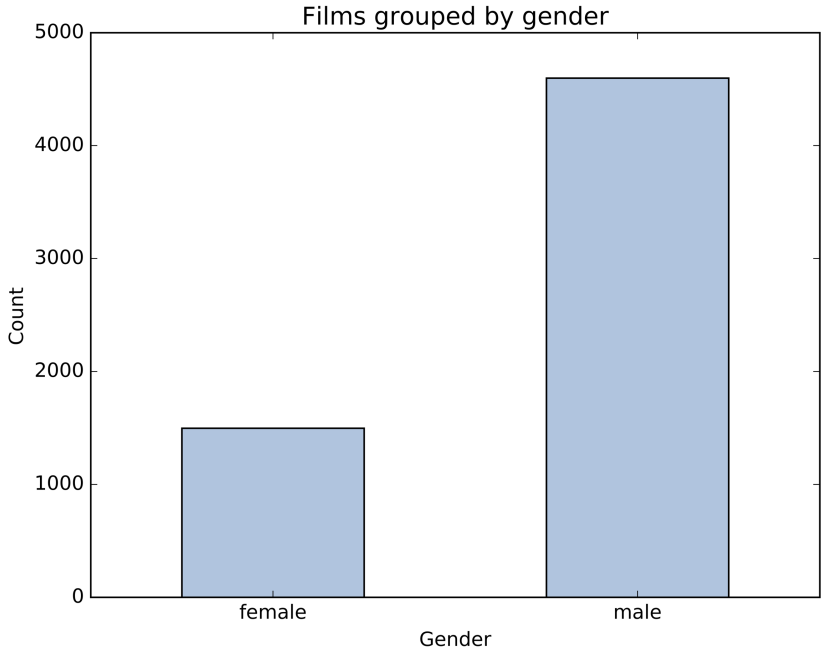

Figure 4

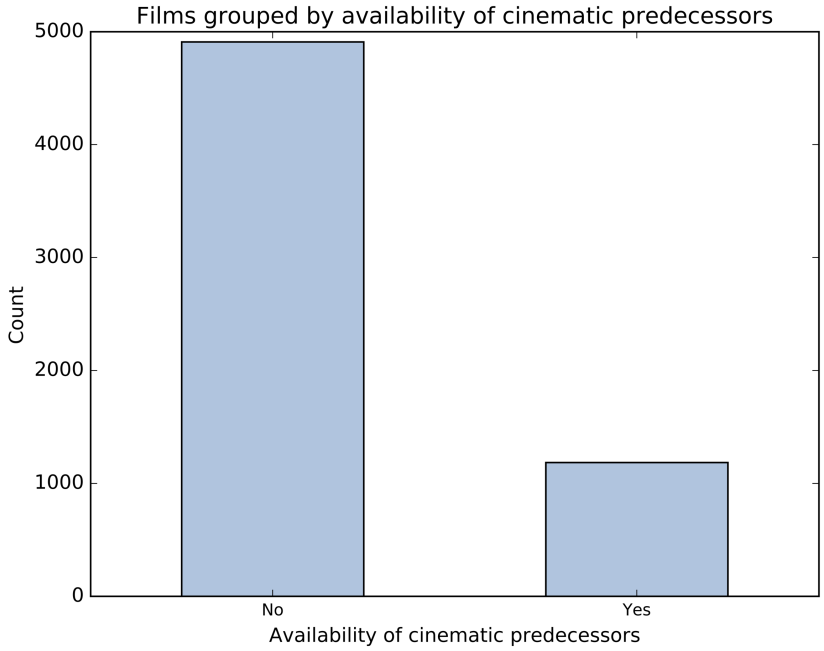

Figure 5

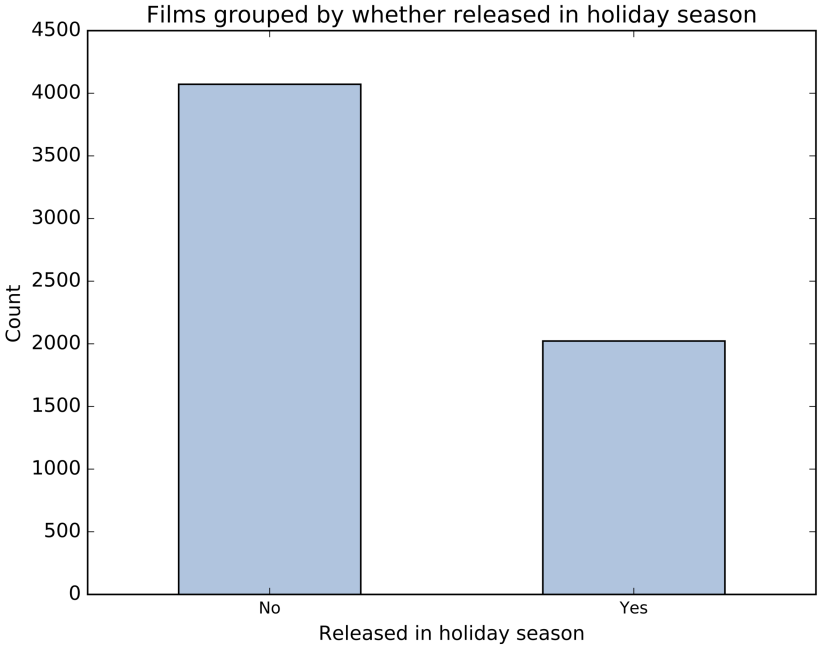

Figure 6

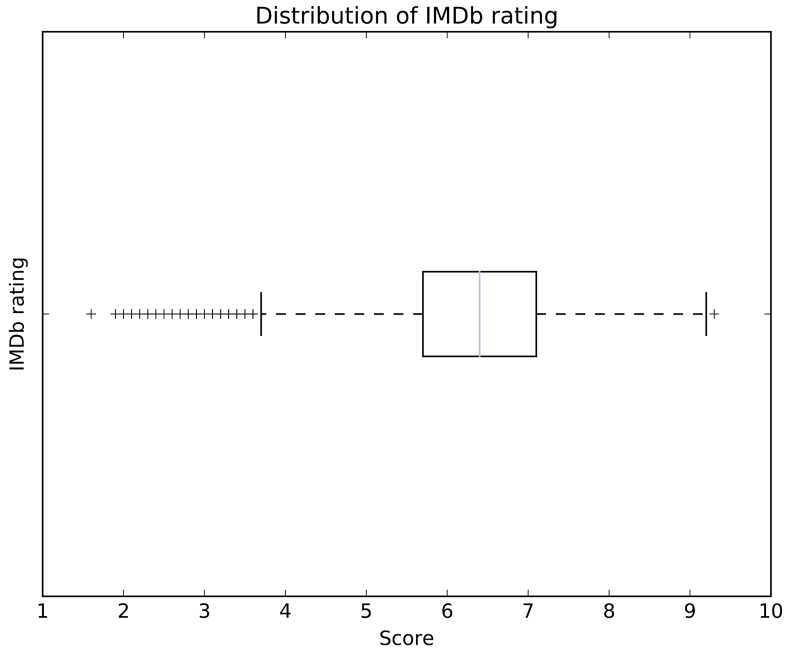

Figure 7

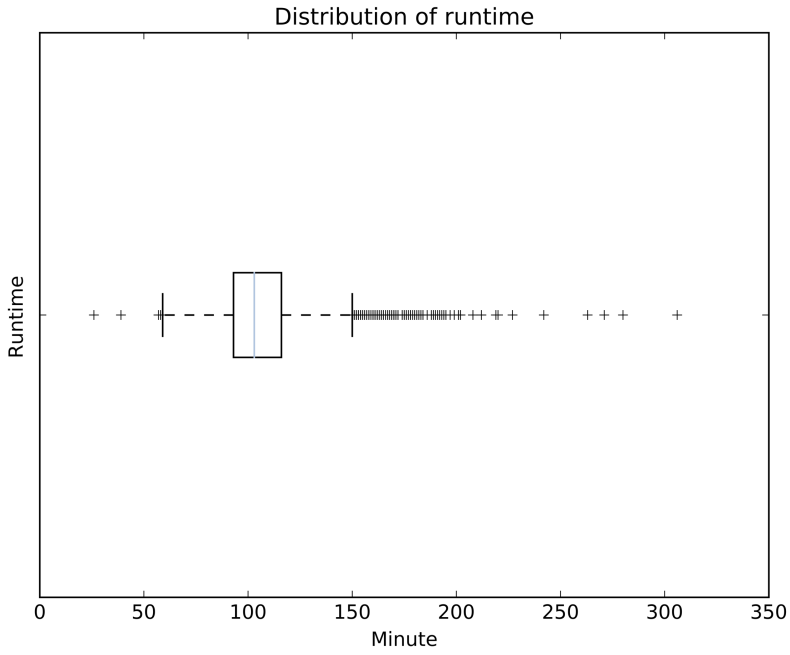

Figure 8

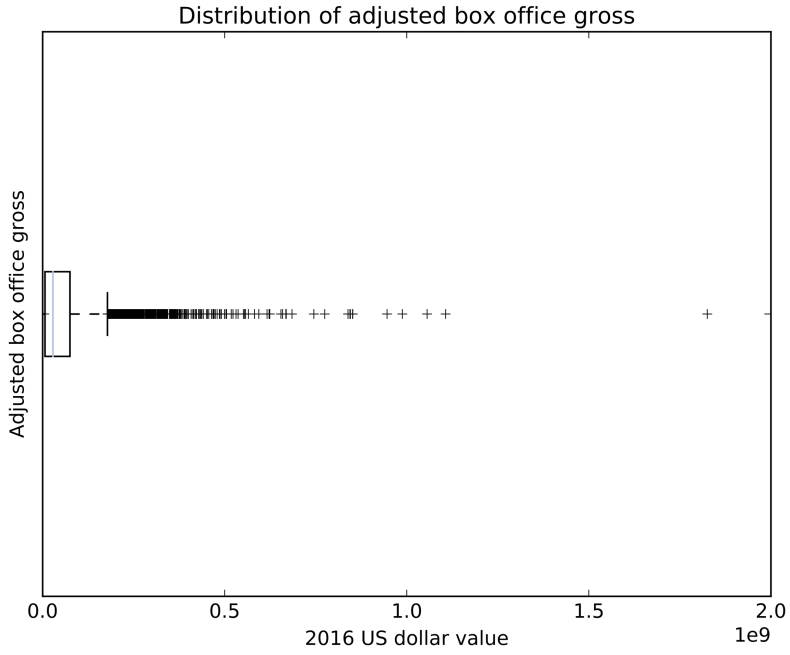

Figure 9

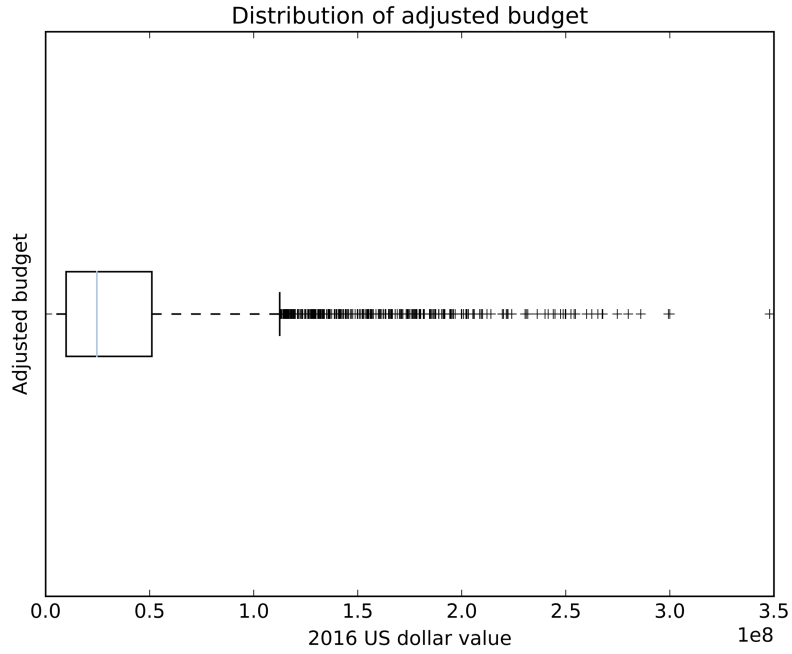

Figure 10

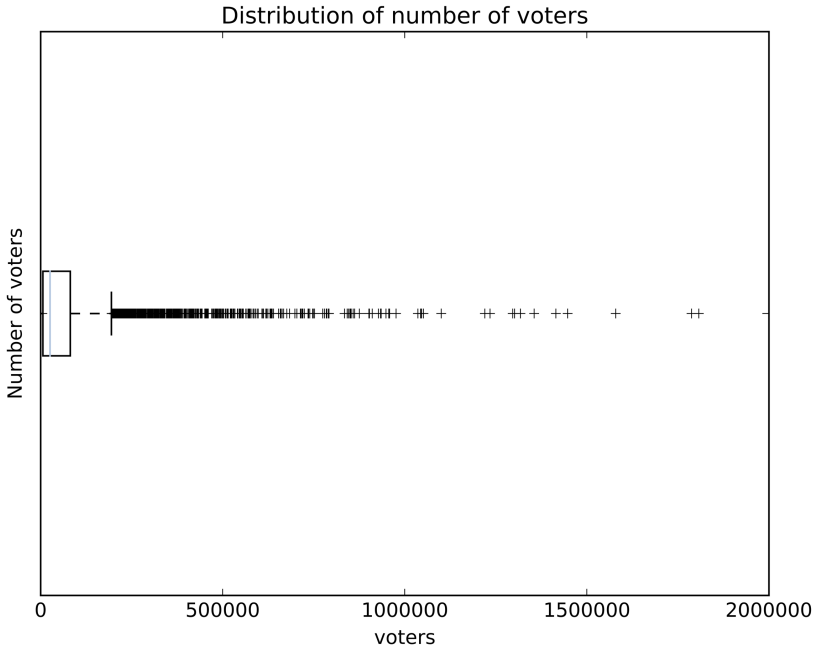

Figure 11

Results

After regression diagnostics[6] for Model (1) and Model (2), one observation in the dataset is defined as a mistake and therefore dropped, and several variables are transformed in order to meet the linear regression model assumption. The Models are corrected as

$$Y_i^{.3} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}{}^2 + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 \log(X_{i4}) + \beta_6 X_{i5} + \beta_7 X_{i6} + \beta_8 X_{i7} + \varepsilon_i \quad (1^*)$$

$$Y_i{}^2 = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 \log(X_{i2}) + \beta_4 X_{i3} + \beta_5 X_{i4} + \beta_6 X_{i5} + \beta_7 X_{i5}{}^2 + \beta_8 X_{i6} + \beta_9 X_{i7} + \varepsilon_i \quad (2^*)$$

The models are fitted in both the whole dataset and subsets of the dataset to examine the significance of the gender term in different situations.

Here are the gender term significance test results for the whole dataset in condition of different genres and years.

(Year 1913 – 2016)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	Male-leading Effect (Std. Error)
	Total[7]	6095	1497 (24.56%)	4598 (75.44%)	-2.911** (1.093)	0.757*** (0.1594)
Genre[8]	**Action**	1303	140 (10.74%)	1163 (89.26%)	0.9102 (5.066)	1.157 (0.7390)
	Adult	1	0 (0.00%)	1 (100.00%)		
	Adventure	1082	167 (15.43%)	915 (84.87%)	-3.051 (5.276)	0.8440 (0.7803)
	Animation	265	55 (20.75%)	210 (79.25%)	5.927 (9.742)	-1.720 (1.267)

[6] See Appendix.
[7] The Models (1*) and (2*) are fitted.
[8] The Reduced Models excluding the genre term X_6 $Y_i^{.3} = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2}{}^2 + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 \log(X_{i4}) + \beta_6 X_{i5} + \beta_7 X_{i7} + \varepsilon_i$ (1**) and $Y_i{}^2 = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 \log(X_{i2}) + \beta_4 X_{i3} + \beta_5 X_{i4} + \beta_6 X_{i5} + \beta_7 X_{i5}{}^2 + \beta_8 X_{i7} + \varepsilon_i$ (2**) are fitted.

Biography	333	86 (25.83%)	247 (74.17%)	4.164 (8.028)	0.08797 (0.856)
Comedy	2405	626 (26.03%)	1779 (73.97%)	-8.092 (2.788)	0.3656 (0.4389)
Crime	1154	163 (14.12%)	991 (85.88%)	-3.617 (4.635)	1.223 . (0.6334)
Documentary	75	14 (18.67%)	61 (81.33%)	-3.295 (16.16)	3.859 (4.857)
Drama	3189	875 (27.44%)	2314 (72.56%)	-8.043** (2.468)	0.9742** (0.3231)
Family	662	152 (22.96%)	510 (77.03%)	6.994 (5.735)	-2.445** (0.9294)
Fantasy	667	150 (22.49%)	517 (77.51%)	11.09 . (5.658)	-1.313 (0.9212)
Film-Noir	44	8 (18.19%)	36 (81.82%)	-47.20 . (25.09)	0.5494 (1.690)
History	228	44 (19.30%)	187 (80.70%)	8.966 (11.22)	-0.3445 (1.365)
Horror	670	237 (35.37%)	433 (64.63%)	-9.598 (4.671)	0.9611 (0.6253)
Music	245	83 (33.88%)	162 (66.12%)	2.211 (9.146)	1.630 (1.506)
Musical	217	76 (35.02%)	141 (64.98%)	2.176 (11.02)	0.8861 (1.420)
Mystery	600	166 (27.67%)	434 (72.33%)	-13.34* (5.334)	2.087** (0.7111)
News	1	0 (0.00%)	1 (100.00%)		
Romance	1469	554 (37.71%)	915 (62.29%)	-3.646 (3.359)	1.368** (0.4662)
Sci-Fi	675	113 (16.74%)	562 (83.26%)	-0.7287 (5.678)	0.1671 (0.9628)
Short	2	0 (0.00%)	2 (100.00%)		
Sport	217	22 (10.14%)	195 (89.86%)	1.512 (13.90)	3.486 (2.175)
Thriller	1728	343 (19.85%)	1385 (80.15%)	-10.52 (3.382)	1.891*** (0.4589)
War	270	41 (15.19%)	229 (84.81%)	20.35 (1.234)	0.2365 (1.405)
Western	174	14 (8.05%)	160 (91.95%)	42.57* (17.16)	-1.054 (1.957)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 1

Table 1 shows that the gender of top-billing cast is statistically significant[9] for both commercial success and aesthetic success of a film.

When considering different genres, the gender term in commercial success model is only statistically significant in Drama, Mystery, and Western films, and gender term in aesthetic success model is statistically significant in Drama, Family, Mystery, Romance, and Thriller films.

Moreover, the male-leading films overwhelmingly outnumber the female-leading films, even in seemingly female dominated film genres Romance, Music, and Musical, the percentage of female-leading films is about 35%, and female-leading ones take 35.37% in Horror films.

Another notable phenomenon is that most of the estimated coefficient of gender term in commercial success model is negative, while most of them is positive in aesthetic model, which means that male-leading is negatively affect the financial performance and positively affect the consumer rating, in other words, female-leading films are easier to achieve commercial success than male-leading films, and male-leading films are easier to achieve aesthetic success than female-leading films. However, due to the imbalanced number of female-leading and male-leading films, this conclusion need some verifications.

[9] The significance level this research takes is 0.5.

(All genres)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[10]	6095	1497 (24.56%)	4598 (75.44%)	-2.911** (1.093)	0.757*** (0.1594)
Year[11]	**1913 – 1919**	7	4 (57.14%)	3 (42.86%)		
	1920 – 1929	32	14 (43.75%)	18 (56.25%)	36.09** (13.17)	-2.815 (2.075)
	1930 – 1939	57	35 (61.40%)	22 (38.60%)	-2.894 (10.42)	0.1127 (1.292)
	1940 – 1949	95	35 (36.84%%)	60 (63.16%)	-7.899 (8.321)	1.232* (0.5732)
	1950 – 1959	207	45 (21.74%)	162 (78.26%)	-11.97 . (6.176)	0.2155 (0.7308)
	1960 – 1969	153	35 (22.88%)	118 (77.12%)	18.51 . (10.08)	-0.7234 (0.7667)
	1970 – 1979	249	48 (19.28%)	201 (80.72%)	2.588 (8.696)	1.541 . (0.855)
	1980 – 1989	731	133 (18.19%)	598 (81.81%)	4.934 (3.832)	0.9477 . (0.5122)
	1990 – 1999	1330	293 (22.03%)	1037 (77.97%)	-1.913 (2.305)	0.5267 (0.356)
	2000 – 2009	1968	514 (26.12%)	1454 (73.88%)	-6.200*** (1.698)	1.209*** (0.2765)
	2010 – 2016	1266	341 (26.94%)	925 (73.06%)	-7.989*** (2.136)	0.2432 (0.3267)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 2

Table 2 shows the effect of gender term varying by time. The ratio of female-leading films is not significantly increasing overtime, rather, it experiences a sharp decrease from 1930 to 1949, remains stable from 1950 – 1999, and increases slightly in 2000 – 2016.

The male-leading effect for commercial success is statistically significant in year 1920 – 1929, 2000 – 2009, and 2010 – 2016, and is mostly negative. A remarkable trend is that from 1970 to 2016, the male-leading effect switches from positive to negative and is continue decreasing, meaning that it is getting hard for male-leading films to achieve commercial success, however, one explanation for this is the large percentage of male-leading films in the dataset.

[10] The Models (1*) and (2*) are fitted.

[11] The Reduced Model excluding the year term X_7 $Y_i^{.3} = \beta_0 + \beta_1 X_{i1} + \beta_2 {X_{i2}}^2 + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 \log(X_{i4}) + \beta_6 X_{i5} + \beta_7 X_{i6} + \varepsilon_i$ (1***) and $Y_i^2 = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 \log(X_{i2}) + \beta_4 X_{i3} + \beta_5 X_{i4} + \beta_6 X_{i5} + \beta_7 {X_{i5}}^2 + \beta_8 X_{i6} + \varepsilon_i$ (2***) are fitted.

The male-leading effect for aesthetic success is statistically significant in year 1940 – 1949, and 2000 – 2009. The estimated coefficient is mostly positive, and overall decreasing through 1970 till now, meaning that though the male-leading films are easier to achieve aesthetic success than female-leading films, the difference is diminishing.

The genres in which gender of top-billing actor makes a significant difference as well as some main genres are picked out for a closer observation. The significance of gender of top-billing cast in genre Action, Adventure, Comedy, Crime, Drama, Family, Mystery, Romance, and Thriller over time is listed in Table 3 to 11.

(Genre: Action)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[12]	1303	140 (10.74%)	1163 (89.26%)	0.9102 (5.066)	1.157 (0.7390)
Year[13]	**1913 – 1919**	2	0 (0.00%)	2 (100.00%)		
	1920 – 1929					
	1930 – 1939	3	0 (0.00%)	3 (100.00%)		
	1940 – 1949	5	2 (40.00%)	3 (60.00%)		
	1950 – 1959	16	1 (6.25%)	15 (93.75%)	-49.72 (34.47)	1.859 (7.810)
	1960 – 1969	25	0 (0.00%)	25 (100.00%)		
	1970 – 1979	57	3 (5.26%)	54 (94.74%)	97.08 . (56.85)	-5.502 (5.644)
	1980 – 1989	181	9 (4.97%)	172 (95.03%)	-11.48 (21.13)	5.087 . (2.827)
	1990 – 1999	301	29 (9.63%)	272 (90.37%)	-3.414 (10.24)	1.955 (1.703)
	2000 – 2009	381	55 (14.44%)	326 (85.56%)	12.18 (7.489)	0.1263 (1.255)
	2010 – 2016	332	41 (12.35%)	291 (87.65%)	-12.99 (8.801)	1.771 (1.263)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 3

[12] The Reduced Models excluding the genre term X_6 (1**) and (2**) are fitted.

[13] The Reduced Model excluding the genre term X_6 and year term X_7 $Y_i^{.3} = \beta_0 + \beta_1 X_{i1} + \beta_2 {X_{i2}}^2 + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 \log(X_{i4}) + \beta_6 X_{i5} + \varepsilon_i$ (1****) and $Y_i^2 = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 \log(X_{i2}) + \beta_4 X_{i3} + \beta_5 X_{i4} + \beta_6 X_{i5} + \beta_7 {X_{i5}}^2 + \varepsilon_i$ (2****) are fitted.

Table 3 shows that the percentage of female-leading films of Action genre is increasing since 1980, but is still quite small. The male-leading Action films are much easier than female-leading films to achieve aesthetic success.

(Genre: Adventure)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	Male-leading Effect (Std. Error)
	Total[14]	1082	167 (15.43%)	915 (84.87%)	-3.051 (5.276)	0.8440 (0.7803)
Year[15]	**1913 – 1919**	2	1 (50.00%)	1 (50.00%)		
	1920 – 1929	6	0 (0.00%)	6 (100.00%)		
	1930 – 1939	7	3 (42.86%)	4 (57.14%)		
	1940 – 1949	14	5 (35.71%)	9 (64.29%)	-2.164 (64.86)	4.761* (1.350)
	1950 – 1959	51	7 (13.73%)	44 (86.27%)	-17.43 (25.14)	1.077 (2.977)
	1960 – 1969	39	1 (2.56%)	38 (97.44%)	-89.47 (118.6)	-6.936 (6.227)
	1970 – 1979	40	6 (15.00%)	34 (85.00%)	-5.884 (46.14)	7.434 . (3.929)
	1980 – 1989	138	24 (17.39%)	114 (82.61%)	2.668 (15.71)	5.672** (2.095)
	1990 – 1999	178	22 (12.36%)	156 (87.64%)	0.4988 (14.19)	0.3831 (2.263)
	2000 – 2009	339	47 (13.86%)	292 (86.14%)	2.785 (8.899)	-1.356 (1.513)
	2010 – 2016	268	51 (19.03%)	217 (80.97%)	-14.42 . (7.962)	-0.9287 (1.366)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 4

Table 4 shows that the percentage of female-leading films of Adventure genre is not obviously increasing recently. The male-leading effect for achieving aesthetic success is decreasing since 1970 from highly positive to negative, which means that it used to be easier for male-leading Adventure films to achieve aesthetic success but since 2000, it has been easier for female-leading Adventure films to achieve aesthetic success.

[14] The Reduced Models excluding the genre term X_6 (1**) and (2**) are fitted.
[15] The Reduced Model excluding the genre term X_6 and year term X_7 (1****) and (2****) are fitted.

(Genre: Comedy)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[16]	2405	626 (26.03%)	1779 (73.97%)	-8.092 (2.788)	0.3656 (0.4389)
Year[17]	**1913 – 1919**					
	1920 – 1929	5	2 (40.00%)	3 (60.00%)		
	1930 – 1939	22	11 (50.00%)	11 (50.00%)	-7.959 (27.23)	3.884 (2.723)
	1940 – 1949	22	11 (50.00%)	11 (50.00%)	18.54 (25.83)	1.622 (2.509)
	1950 – 1959	28	10 (35.71%)	18 (64.29%)	-60.03* (24.97)	3.633 (2.715)
	1960 – 1969	54	18 (33.33%)	36 (66.67%)	49.01 . (24.96)	-4.04 . (2.181)
	1970 – 1979	78	17 (21.79%)	61 (78.21%)	17.30 (25.16)	1.237 (2.477)
	1980 – 1989	296	51 (17.23%)	245 (82.77%)	9.679 (10.30)	-0.7699 (1.424)
	1990 – 1999	592	151 (25.51%)	441 (74.49%)	-0.5885 (5.567)	-1.446 (0.9027)
	2000 – 2009	848	225 (26.53%)	623 (73.47%)	-12.22 (4.169)	1.739* (0.7313)
	2010 – 2016	460	130 (28.26%)	330 (71.74%)	3.589 (5.591)	0.435 (0.9088)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 5

Table 5 shows that the percentage of female-leading films of Comedy genre is steadily increase since 1980. Recently, it is the male-leading Comedy film that are easier to achieve both commercial and aesthetic success.

[16] The Reduced Models excluding the genre term X_6 (1**) and (x**) are fitted.
[17] The Reduced Model excluding the genre term X_6 and year term X_7 (1****) and (2****) are fitted.

(Genre: Crime)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[18]	1154	163 (14.12%)	991 (85.88%)	-3.617 (4.635)	1.223 . (0.6334)
Year[19]	**1913 – 1919**	1	1 (100.00%)	0 (0.00%)		
	1920 – 1929	2	1 (50.00%)	1 (50.00%)		
	1930 – 1939	3	1 (33.33%)	2 (66.67%)		
	1940 – 1949	16	3 (18.75%)	13 (81.25%)	-74.43 (65.68)	0.4099 (4.359)
	1950 – 1959	28	2 (7.14%)	26 (92.86%)	-43.1 (35.3)	1.947 (3.14)
	1960 – 1969	18	4 (22.22%)	14 (77.78%)	13.03 (29.2)	-3.849 (5.819)
	1970 – 1979	51	7 (13.72%)	44 (86.27%)	41.17 (32.19)	2.054 (2.711)
	1980 – 1989	138	8 (5.80%)	130 (95.42%)	25.15 (20.70)	-2.340 (2.501)
	1990 – 1999	325	46 (14.15%)	279 (85.85%)	-5.773 (8.078)	3.683** (1.310)
	2000 – 2009	375	63 (16.80%)	312 (83.20%)	-2.238 (7.083)	0.7799 (1.050)
	2010 – 2016	197	27 (13.71%)	170 (86.29%)	-23.73* (11.29)	-0.4417 (1.385)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 6

Table 6 shows that the percentage of female-leading films of Crime genre is not increasing. Since 1990, it is female-leading Crime films that are easier to achieve commercial success.

[18] The Reduced Models excluding the genre term X_6 (1**) and (2**) are fitted.
[19] The Reduced Model excluding the genre term X_6 and year term X_7 (1****) and (2****) are fitted.

(Genre: Drama)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[20]	3189	875 (27.44%)	2314 (72.56%)	-8.043** (2.468)	0.9742** (0.3231)
Year[21]	**1913 – 1919**	6	4 (66.67%)	2 (33.33%)		
	1920 – 1929	26	12 (46.15%)	14 (53.85%)	30.72 (19.89)	-1.118 (3.890)
	1930 – 1939	33	24 (72.73%)	9 (27.27%)	25.78 (24.6)	-2.098 (3.357)
	1940 – 1949	60	18 (0.30%)	42 (0.70%)	-19.04 (20.36)	1.856 (1.406)
	1950 – 1959	129	29 (22.48%)	100 (77.52%)	-11 (12.59)	-0.8692 (1.574)
	1960 – 1969	85	23 (27.06%)	62 (72.94%)	1.708 (22.49)	1.878 (1.876)
	1970 – 1979	134	29 (21.64%)	105 (78.36%)	6.583 (19.54)	-0.4085 (1.784)
	1980 – 1989	317	64 (20.19%)	253 (79.81%)	-8.884 (9.556)	0.2054 (1.105)
	1990 – 1999	737	191 (25.92%)	546 (74.08%)	-7.786 (4.933)	0.6594 (0.641)
	2000 – 2009	1028	286 (27.82%)	742 (72.18%)	-11.61** (3.849)	1.875** (0.5771)
	2010 – 2016	634	195 (30.76%)	439 (69.24%)	-11.78* (4.775)	0.6059 (0.7043)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 7

Table 7 shows that the percentage of female-leading films of Drama genre is steadily increasing. From 1980 till now, it is much easier for female-leading Drama films to achieve commercial success and male-leading Drama films to achieve aesthetic success.

[20] The Reduced Models excluding the genre term X_6 (1**) and (2**) are fitted.
[21] The Reduced Model excluding the genre term X_6 and year term X_7 (1****) and (2****) are fitted.

(Genre: Family)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[22]	662	152 (22.96%)	510 (77.03%)	6.994 (5.735)	-2.445** (0.9294)
Year[23]	**1913 – 1919**					
	1920 – 1929	3	0 (0.00%)	3 (100.00%)		
	1930 – 1939	3	1 (33.33%)	2 (66.67%)		
	1940 – 1949	4	2 (50.00%)	2 (50.00%)		
	1950 – 1959	17	7 (41.18%)	10 (58.82%)		
	1960 – 1969	14	5 (35.71%)	9 (64.29%)	227.9** (60.46)	-4.252 . (2.054)
	1970 – 1979	16	5 (31.25%)	11 (68.75%)	-13.92 (58.06)	12.3 . (6.557)
	1980 – 1989	46	8 (17.39%)	38 (82.61%)	28.38 (25.32)	2.070 (4.203)
	1990 – 1999	151	28 (18.54%)	123 (81.46%)	9.498 (14.44)	-5.722* (2.264)
	2000 – 2009	255	60 (23.53%)	195 (76.47%)	10.68 (8.302)	-3.213* (1.504)
	2010 – 2016	153	36 (23.53%)	117 (76.47%)	-5.041 (9.007)	-2.783 (1.837)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 8

Table 8 shows that the percentage of female-leading films of Family genre is slightly increasing. It is much easier for female-leading Family films to achieve aesthetic success.

[22] The Reduced Models excluding the genre term X_6 (1**) and (2**) are fitted.
[23] The Reduced Model excluding the genre term X_6 and year term X_7 (1****) and (2****) are fitted.

(Genre: Mystery)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[24]	600	166 (27.67%)	434 (72.33%)	-13.34* (5.334)	2.087** (0.7111)
Year[25]	**1913 – 1919**	1	1 (100.00%)	0 (0.00%)		
	1920 – 1929	1	0 (0.00%)	1 (100.00%)		
	1930 – 1939					
	1940 – 1949	8	2 (25.00%)	6 (75.00%)		
	1950 – 1959	14	2 (14.29%)	12 (85.71%)	-53.25 . (25.61)	4.199 (6.721)
	1960 – 1969	13	3 (23.08%)	10 (76.92%)		
	1970 – 1979	15	5 (33.33%)	10 (66.67%)	61.72 (36.88)	5.275 (3.929)
	1980 – 1989	64	12 (18.75%)	52 (81.25%)	10.89 (16.11)	-0.9838 (2.027)
	1990 – 1999	129	28 (21.71%)	101 (78.29%)	-5.856 (12.49)	3.002 . (1.628)
	2000 – 2009	227	74 (32.60%)	153 (67.40%)	-20.91* (8.342)	2.763* (1.170)
	2010 – 2016	128	39 (30.47%)	89 (69.53%)	-19.48 . (11.06)	0.8264 (1.413)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 9

Table 9 shows that the percentage of female-leading films of Mystery genre is steadily increasing. It is much easier for female-leading Mystery films to achieve commercial success and it is easier for male-leading Mystery films to achieve aesthetic films but the difference is diminishing.

[24] The Reduced Models excluding the genre term X_6 (1**) and (2**) are fitted.
[25] The Reduced Model excluding the genre term X_6 and year term X_7 (1****) and (2****) are fitted.

(Genre: Romance)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[26]	1469	554 (37.71%)	915 (62.29%)	-3.646 (3.359)	1.368** (0.4662)
Year[27]	**1913 – 1919**	1	0 (50.00%)	1 (100.00%)		
	1920 – 1929	21	9 (42.86%)	12 (57.14%)		
	1930 – 1939	33	23 (69.70%)	10 (30.30%)	30.28 (18.1)	-0.5372 (3.072)
	1940 – 1949	43	23 (53.49%)	20 (46.51%)	-1.211 (18.94)	1.045 (1.306)
	1950 – 1959	86	28 (32.56%)	58 (67.44%)	-12.21 (14.63)	-0.9519 (1.486)
	1960 – 1969	47	18 (38.30%)	29 (61.70%)	-7.572 (26.84)	0.1491 (2.449)
	1970 – 1979	33	12 (36.36%)	21 (63.64%)	-57.32 (29.44)	2.831 (3.324)
	1980 – 1989	144	40 (27.78%)	104 (72.22%)	3.653 (13.4)	2.831 (3.324)
	1990 – 1999	308	94 (30.52%)	214 (69.48%)	4 (7.751)	1.025 (1.071)
	2000 – 2009	518	211 (40.73%)	307 (59.27%)	-6.240 (5.156)	2.898*** (0.8135)
	2010 – 2016	235	96 (40.85%)	139 (59.15%)	-1.287 (7.878)	0.4066 (1.126)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 10

Table 10 shows that the percentage of female-leading films of Romance genre is higher in the early time of films and experience a slump during 1950s but get high again recently. It is much easier for female-leading Romance films to achieve commercial success and male-leading Romance films to achieve aesthetic success.

[26] The Reduced Models excluding the genre term X_6 (1**) and (2**) are fitted.
[27] The Reduced Model excluding the genre term X_6 and year term X_7 (1****) and (2****) are fitted.

(Genre: Thriller)

		# Obs	# Female-leading	# Male-leading	Commercial Success Model	Aesthetic Success Model
					Male-leading Effect (Std. Error)	**Male-leading Effect (Std. Error)**
	Total[28]	1728	343 (19.85%)	1385 (80.15%)	-10.52 (3.382)	1.891*** (0.4589)
Year[29]	**1913 – 1919**					
	1920 – 1929					
	1930 – 1939	1	0 (0.00%)	1 (100.00%)		
	1940 – 1949	18	5 (27.78%)	13 (72.22%)	-6.324 (55.39)	-4.390 (2.547)
	1950 – 1959	32	3 (9.38%)	29 (90.63%)	-37.55 (28.97)	-3.239 (3.481)
	1960 – 1969	25	5 (20.00%)	20 (80.00%)		
	1970 – 1979	66	11 (16.67%)	55 (83.33%)	26.12 (29.32)	1.078 (2.598)
	1980 – 1989	213	33 (15.49%)	180 (84.51%)	15.46 (9.749)	-0.8945 (1.406)
	1990 – 1999	426	71 (16.67%)	355 (83.33%)	-1.709 (6.689)	2.832** (1.019)
	2000 – 2009	565	124 (21.95%)	441 (78.05%)	-15.25** (5.528)	2.482** (0.7735)
	2010 – 2016	382	91 (23.82%)	291 (76.18%)	-2.182** (7.029)	1.959* (0.9057)

. p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 11

Table 11 shows that the percentage of female-leading films of Thriller genre is slightly increasing. It is easier for male-leading Thriller films to achieve commercial success and male-leading Thriller films to achieve aesthetic success.

A graphic conclusion helps to draw a more straight-forward impression of the male-leading effect across genres overtime.

[28] The Reduced Models excluding the genre term X_6 (1**) and (2**) are fitted.
[29] The Reduced Model excluding the genre term X_6 and year term X_7 (1****) and (2****) are fitted.

Commercial success model

Total	1920 – 1929	1930 – 1939	1940 – 1949	1950 – 1959	1960 – 1969	1970 – 1979	1980 – 1989	1990 – 1999	2000 – 2009	2010 – 2016
Action										
Adventure										
Comedy										
Crime										
Drama										
Family										
Mystery										
Romance										
Thriller										

Aesthetic success model

Total	1920 – 1929	1930– 1939	1940 – 1949	1950 – 1959	1960 – 1969	1970 – 1979	1980 – 1989	1990 – 1999	2000 – 2009	2010 – 2016
Action										
Adventure										
Comedy										
Crime										
Drama										
Family										
Mystery										
Romance										
Thriller										

Negative Coefficient	p > 0.1	. p < 0.1	*p < 0.05	**p < 0.01	***p < 0.001
Positive Coefficient	p > 0.1	. p < 0.1	*p < 0.05	**p < 0.01	***p < 0.001

Cell remains blank if model not applicable.

Figure 12

Conclusion

Discussion

In conclusion, the gender of top-billing actor is statistically significant for both commercial success and aesthetic success of a film, and the gender effect for commercial success is decreasing and that for aesthetic success is increasing.

The gender effect is significant for both commercial and aesthetic success of Drama and Mystery films, and is significant solely for commercial success of Western films, significant

solely for aesthetic success of Family, Romance, and Thriller films. According to Fiebert (1990), the traditional female role includes four dimensions: woman's interest in appearance and beauty, women's interest in domestic skills, women's concern for the care and nature of others, and women' preoccupation with romance. This coincides with the findings of this research that gender effect is significant in Family and Romance films.

The percentage of female-leading films is not significantly increasing overtime and the male-leading overwhelmingly outnumber female-leading films. This is surprising given that females comprise over half of United States population.

It is easier for female-leading films to achieve commercial success. This may partially result from the imbalanced number of male-leading and female-leading films. Another possible interpretation is about rarity. Since female-leading films are much less than male-leading films, they can be more novel and attracting, thus achieve better box office gross.

It is easier for male-leading films to achieve aesthetic success. This may due to the difference of career development of actors and actresses. Levy (1989) studied the demographic origins and early career development of movie stars active between 1932 and 1984, and found that though women more often enter the acting profession at younger ages but male actors tend to have somewhat longer careers as stars (Lehman, 1941) whereas female actors are more prone to have to assume increasingly unattractive roles if they want to continue in the business (Bazzini et al, 1997). Moreover, female performers earn much less money than male performers, both over the career course and at their career peak (Lehman, 1941; markson & Taylor, 1993). In line with these differences, gender is intimately related to the attainment of the much-coveted position of the "movie star". For example, according to one investigation, 78% of movie stars were identified as male (Wallace, Seigerman, & Holbrook, 1993). The quality leading cast

highly correlated with the budget and therefore the quality of the films, hence, this curious gender contrast may help explain that the females may less than males to appear as a leading role in aesthetically successful films. What's more, Simonton (2004) in his "Best Actress" Paradox research stated that male roles are more strongly associated with great stories than are female roles. One evidence supporting it is the finding that the nest-selling novels, in contrast to unsuccessful novels, tend to have central male characters who are far more conspicuous than the female characters (Harvey, 1953).

Some disagreements with the overall trend when examining each genre carefully are that the percentage of female-leading films of Action, Comedy, Drama, Family, Mystery, Thriller genre is steadily increasing, and in among Adventure films and Family films, it is easier for female-leading films to achieve aesthetic success, while among Comedy films, it is easier for male-leading films to achieve commercial success.

Limitation

As earlier discussed, the results obtained from this research may suffer from the fact that male-leading films is much more than female-leading films, so some adjust may be needed for further interpretation.

Moreover, a detailed count-level comparison between male-leading films and female-leading films can be obtained from a bigger subset of IMDb dataset with only the gender of top-billing cast information available. However, this research only includes films with all variables information available and thus sacrifice the data amount.

To even more improve this research, more control variables should be found and included in the models since the R-squared in both models are about 0.5, which is not very stratifying, though all controlling variables included are statistically significant.

References

Acheson, K., & Maule, C. J. (1994). Understanding Hollywood's organization and continuing success. Journal of Cultural Economics, 18(4), 271-300.

Archer, S. L. (1985). Identity and the choice of social roles. New Directions for Child and Adolescent Development, 1985(30), 79-99.

Arnheim, R. (1957). Film as art. Univ of California Press.

Bandura, A., Ross, D., & Ross, S. A. (1963). A comparative test of the status envy, social power, and secondary reinforcement theories of identificatory learning. The Journal of Abnormal and Social Psychology, 67(6), 527.

Bandura, A. (1969). Social-learning theory of identificatory processes. Handbook of socialization theory and research, 213, 262.

Bandura, A. (1986). Social foundations of thought and action: A social cognitive perspective. Englewood Cliffs, NJ: Princeton-Hall.

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How critical are critical reviews? The box office effects of film critics, star power, and budgets. Journal of marketing, 67(4), 103-117.

Barcus, F. E. (1983). Images of life on children's television: Sex roles, minorities, and families. Praeger Publishers.

Baumann, S. (2001). Intellectualization and art world development: Film in the United States. American Sociological Review, 404-426.

Bazzini, D. G., McIntosh, W. D., Smith, S. M., Cook, S., & Harris, C. (1997). The aging woman in popular film: Underrepresented, unattractive, unfriendly, and unintelligent. Sex roles, 36(7), 531-543.

Bem, S. L. (1993). The lenses of gender: Transforming the debate on sexual inequality. Yale University Press.

Bem, S. L. (1994). In a male-centered world, female differences are transformed into female disadvantages. Chronicle of Higher Education, 40(50), B1-B3.

Beuf, A. (1974). Doctor, lawyer, household drudge. Journal of Communication, 24(2), 142-145.

Bleakley, A., Jamieson, P. E., & Romer, D. (2012). Trends of sexual and violent content by gender in top-grossing US films, 1950–2006. Journal of Adolescent Health, 51(1), 73-79.

Boor, M. (1990). Reliability of ratings of movies by professional movie critics. Psychological Reports, 67(1), 243-257.

Brown, L. M., & Gilligan, C. (1993). Meeting at the crossroads: Women's psychology and girls' development. Feminism & Psychology, 3(1), 11-35.

Budd, M. (1995). Values of art.

De Piles, R. 1708. Cours de Peinture par Principes. Jacques Thuillier, ed. Paris: Gallimard.

De Vany, A., & Walls, W. D. (1999). Uncertainty in the movie industry: Does star power reduce the terror of the box office?. Journal of cultural economics, 23(4), 285-318.

Delmestri, G., Montanari, F., & Usai, A. (2005). Reputation and strength of ties in predicting commercial success and artistic merit of independents in the Italian feature film industry. Journal of Management Studies, 42(5), 975-1002.

Desai, K. K., & Basuroy, S. (2005). Interactive influence of genre familiarity, star power, and critics' reviews in the cultural goods industry: The case of motion pictures. Psychology & Marketing, 22(3), 203-223.

Chang, B. H., & Ki, E. J. (2005). Devising a practical model for predicting theatrical movie success: Focusing on the experience good property. Journal of Media Economics, 18(4),

247-269.

Chyng, S. (2001). Mickey Mouse Monopoly [VHS]. Northampton, MA: Media Education Foundation.

Dickie, G. (1997). Introduction to aesthetics: An analytic approach (p. 22). New York: Oxford University Press.

Dickie, G. (2010). Evaluating art. Temple University Press.

Dodds, J. C., & Holbrook, M. B. (1988). What's an Oscar worth? An empirical estimation of the effects of nominations and awards on movie distribution and revenues. Current research in film: Audiences, economics and the law, 4, 72-88.

Dodds, K. (2006). Popular geopolitics and audience dispositions: James Bond and the internet movie database (IMDb). Transactions of the Institute of British Geographers, 31(2), 116-130.

Comstock, G., & Scharrer, E. (2001). The use of television and other film-related media. Handbook of children and the media, 47-72.

Downing, J., Mohammadi, A., & Sreberny, A. (1995). Questioning the media: A critical introduction. Sage.

Downs, A. C. (1981). Sex-role stereotyping on prime-time television. The Journal of Genetic Psychology, 138(2), 253-258.

Eliashberg, J., & Shugan, S. M. (1997). Film critics: Influencers or predictors?. The Journal of Marketing, 68-78.

Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2007). From story line to box office: A new approach for green-lighting movie scripts. Management Science, 53(6), 881-893.

Faber, R. J., Brown, J. D., & McLeod, J. M. (1979). Coming of age in the global village:

Television and adolescence. Children communicating: Media and development of thought, speech, understanding, 215-249.

Fatemi, M., & Tokarchuk, L. (2012, April). An empirical study on IMDb and its communities based on the network of co-reviewers. In Proceedings of the First Workshop on Measurement, Privacy, and Mobility (p. 7). ACM.

Fouts, G., & Burggraf, K. (1999). Television situation comedies: Female body images and verbal reinforcements. Sex roles, 40(5), 473-481.

Gilberg, M., & Hines, T. (2000). Male entertainment award winners are older than female winners. Psychological Reports, 86(1), 175-178.

Ginsburgh, V. (2003). Awards, success and aesthetic quality in the arts. The Journal of Economic Perspectives, 17(2), 99-111.

Ginsburgh, V., & Weyers, S. (2005). Quantitative approaches to valuation in the arts, with an application to movies. Mimeo: Free University Bruxelles.

Hand, C. (2001). Increasing returns to information: further evidence from the UK film market. Applied Economics Letters, 8(6), 419-421.

Haskell, M. (2016). From reverence to rape: The treatment of women in the movies. University of Chicago Press.

Harvey, J. (1953). The content characteristics of best-selling novels. Public Opinion Quarterly, 17(1), 91-114.

Hedley, M. (2002). The geometry of gendered conflict in popular film: 1986–2000. Sex Roles, 47(5-6), 201-217.

Hill, J. P., & Lynch, M. E. (1983). The intensification of gender-related role expectations during early adolescence. In Girls at puberty (pp. 201-228). Springer US.

Hirschman, E. C., & Pieros, A. (1985). Relationships among indicators of success in Broadway plays and motion pictures. Journal of Cultural Economics, 9(1), 35-63.

Hoerrner, K. L. (1996). Gender roles in Disney films: Analyzing behaviors from Snow White to Simba. Women's Studies in Communication, 19(2), 213-228.

Holbrook, M. B. (1999). Popular appeal versus expert judgments of motion pictures. Journal of consumer research, 26(2), 144-155.

Hooks, B. (1994). Outlaw culture: Resisting representations. Routledge.

Hume, D. (1965). Of the standard of taste, and other essays.

Huston, A. C., & Alvarez, M. M. (1990). The socialization context of gender role development in early adolescence.

Lamprecht, D., Geigl, F., Karas, T., Walk, S., Helic, D., & Strohmaier, M. (2015, October). Improving recommender system navigability through diversification: A case study of IMDb. In Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business (p. 21). ACM.

Lauzen, M. M., & Dozier, D. M. (2005). Maintaining the double standard: Portrayals of age and gender in popular films. Sex Roles, 52(7), 437-446.

Lee, F. L. (2006). Cultural discount and cross-culture predictability: Examining the box office performance of American movies in Hong Kong. Journal of Media Economics, 19(4), 259-278.

Lehman, H. C. (1941). The chronological ages of some recipients of large annual incomes. Social Forces, 196-206.

Levy, E. (1989). The democratic elite: America's movie stars. Qualitative Sociology, 12(1), 29-54.

Levy, G. D., & Carter, D. B. (1989). Gender schema, gender constancy, and gender-role knowledge: The roles of cognitive factors in preschoolers' gender-role stereotype attributions. Developmental Psychology, 25(3), 444.

Liben, L. S., & Signorella, M. L. (1993). Gender-schematic processing in children: The role of initial interpretations of stimuli. Developmental Psychology, 29(1), 141.

Lincoln, A. E. (2007). Cultural honours and career promotions: Re-conceptualizing prizes in the field of cultural production. Cultural Trends, 16(1), 3-15.

Litman, B. R., & Ahn, H. (1998). Predicting Financial Success of Motion Pictures. BR Litman The Motion Picture Mega-Industry.

Jung, J. J. (2012). Attribute selection-based recommendation framework for short-head user group: An empirical study by MovieLens and IMDB. Expert Systems with Applications, 39(4), 4049-4054.

Maltin, L. (2005). Leonard Maltin's classic movie guide. Plume.

Markson, E. W., & Taylor, C. A. (1993). Real versus reel world: Older women and the Academy Awards. Women & Therapy, 14(1-2), 157-172.

Mazzarella, S. R., & Pecora, N. O. (Eds.). (1999). Growing up girls: Popular culture and the construction of identity (Vol. 9). Peter Lang Pub Incorporated.

Milburn, M. A., Mather, R., & Conrad, S. D. (2000). The effects of viewing R-rated movie scenes that objectify women on perceptions of date rape. Sex Roles, 43(9), 645-664.

Nihlen, A. S., & Bailey, B. A. (1988). Children's display of gender schemas through interaction with nontraditional workers. Anthropology & Education Quarterly, 19(2), 155-162.

Orenstein, P. (1994). School girls: Younv women. self-esteem, and the confidence gap. New York: Bantam Doubleday DeIl Publishing Group.

Prag, J., & Casavant, J. (1994). An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. Journal of Cultural Economics, 18(3), 217-235.

Reinstein, D. A., & Snyder, C. M. (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. The journal of industrial economics, 53(1), 27-51.

Rosen, M. (1973). Popcorn Venus: Women, Movies & the American Dream.

Ruvolo, A. P., & Markus, H. R. (1992). Possible selves and performance: The power of self-relevant imagery. Social cognition, 10(1), 95-124.

Savile, A. (1982). The test of time: An essay in philosophical aesthetics (p. 157). London: Clarendon Press.

Sawhney, M. S., & Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. Marketing Science, 15(2), 113-131.

Schank, R. C., & Abelson, R. P. (2013). Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Psychology Press.

Shiner, R. A. (1996). Hume and the causal theory of taste. The Journal of Aesthetics and Art Criticism, 54(3), 237-249.

Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children′ s gender schemata about others: A meta-analytic review. Developmental Review, 13(2), 147-183.

Signorielli, N. (1991). A sourcebook on children and television. Greenwood Publishing Group.

Signorielli, N. (1997). A content analysis: Reflections of girls in the media. Children Now and the Kaiser Family Foundation.

Signorielli, N., & Bacue, A. (1999). Recognition and respect: A content analysis of prime-time television characters across three decades. Sex roles, 40(7), 527-544.

Simonton, D. K. (2002). Collaborative aesthetics in the feature film: Cinematic components predicting the differential impact of 2,323 Oscar-nominated movies. Empirical Studies of the Arts, 20(2), 115-125.

Simonton, D. K. (2004). The “best actress” paradox: Outstanding feature films versus exceptional women's performances. Sex Roles, 50(11), 781-794.

Simonton, D. K. (2004). Group Artistic Creativity: Creative Clusters and Cinematic Success in Feature Films1. Journal of Applied Social Psychology, 34(7), 1494-1520.

Simonton, D. K. (2005). Film as art versus film as business: Differential correlates of screenplay characteristics. Empirical Studies of the Arts, 23(2), 93-117.

Simonton, D. K. (2005). Cinematic creativity and production budgets: Does money make the movie?. Journal of Creative Behavior, 39(1).

Simonton, D. K. (2007). Is bad art the opposite of good art? Positive versus negative cinematic assessments of 877 feature films. Empirical Studies of the Arts, 25(2), 143-161.

Simonton, D. K. (2007). Film music: Are award-winning scores and songs heard in successful motion pictures?. Psychology of Aesthetics, Creativity, and the Arts, 1(2), 53.

Simonton, D. K. (2009). Cinematic success criteria and their predictors: The art and business of the film industry. Psychology & Marketing, 26(5), 400-420.

Smith, S. L., & Donnerstein, E. (1998). Harmful effects of exposure to media violence: Learning of aggression, emotional desensitization, and fear.

Smith, S. L., Choueiti, M., & Gall, S. (2010). Gender inequality in popular films: Examining on

screen portrayals and behind-the-scenes employment patterns in motion pictures released between 2007–2009. Los Angeles, CA: Annenberg School for Communication & Journalism.

Smith, S. L., Pieper, K. M., Granados, A., & Choueiti, M. (2010). Assessing gender-related portrayals in top-grossing G-rated films. Sex Roles, 62(11-12), 774-786.

Steinke, J. (1999). Women Scientist Role Models on Screen A Case Study of Contact. Science Communication, 21(2), 111-136.

Steinke, J. (2004). Science in cyberspace: Science and engineering World Wide Web sites for girls. Public Understanding of Science, 13(1), 7-30.

Steinke, J. (2005). Cultural representations of gender and science: Portrayals of female scientists and engineers in popular films. Science Communication, 27(1), 27-63.

Sochay, S. (1994). Predicting the performance of motion pictures. Journal of Media Economics, 7(4), 1-20.

Tenenbaum, H. R., & Leaper, C. (2003). Parent-child conversations about science: The socialization of gender inequities?. Developmental Psychology, 39(1), 34.

Topf, P. (2010). Examining success in the motion picture industry. The Park Place Economist, 18(15), 1.

Vogel, H. L. (2014). Entertainment industry economics: A guide for financial analysis. Cambridge University Press.

Wallace, W. T., Seigerman, A., & Holbrook, M. B. (1993). The role of actors and actresses in the success of films: How much is a movie star worth?. Journal of cultural economics, 17(1), 1-27.

Walker, J. (2000). Halliwell's Film and Video Guide 2001. HarperCollins Canada, Limited.

Wanderer, J. J. (1970). In defense of popular taste: Film ratings among professionals and lay audiences. American Journal of Sociology, 76(2), 262-272.

Zickar, M. J., & Slaughter, J. E. (1999). Examining creative performance over time using hierarchical linear modeling: An illustration using film directors. Human Performance, 12(3-4), 211-230.

Appendix - Model Calibration

The original models this research intends to fit in are

$$Box\ office\ gross_i = \beta_0 + \beta_1 Gender_i + \beta_2 Rating_i + \beta_3 Cinematic\ predecessors_i + \beta_4 Budget_i + \beta_5 Holiday\ season_i + \beta_6 Genre_i + \beta_7 Year_i + \varepsilon_i \quad (1)$$

$$Rating_i = \beta_0 + \beta_1 Gender_i + \beta_2 Voters_i + \beta_3 Cinematic\ predecessors_i + \beta_4 Budget_i + \beta_5 Runtime_i + \beta_6 Genre_i + \beta_7 Year_i + \varepsilon_i \quad (2)$$

(For easier recognition, the variables are named differently in the Appendix. Refer to Variable Section for more detailed explanation of variables.)

Identify outliers

Outlier observation – Model (1)

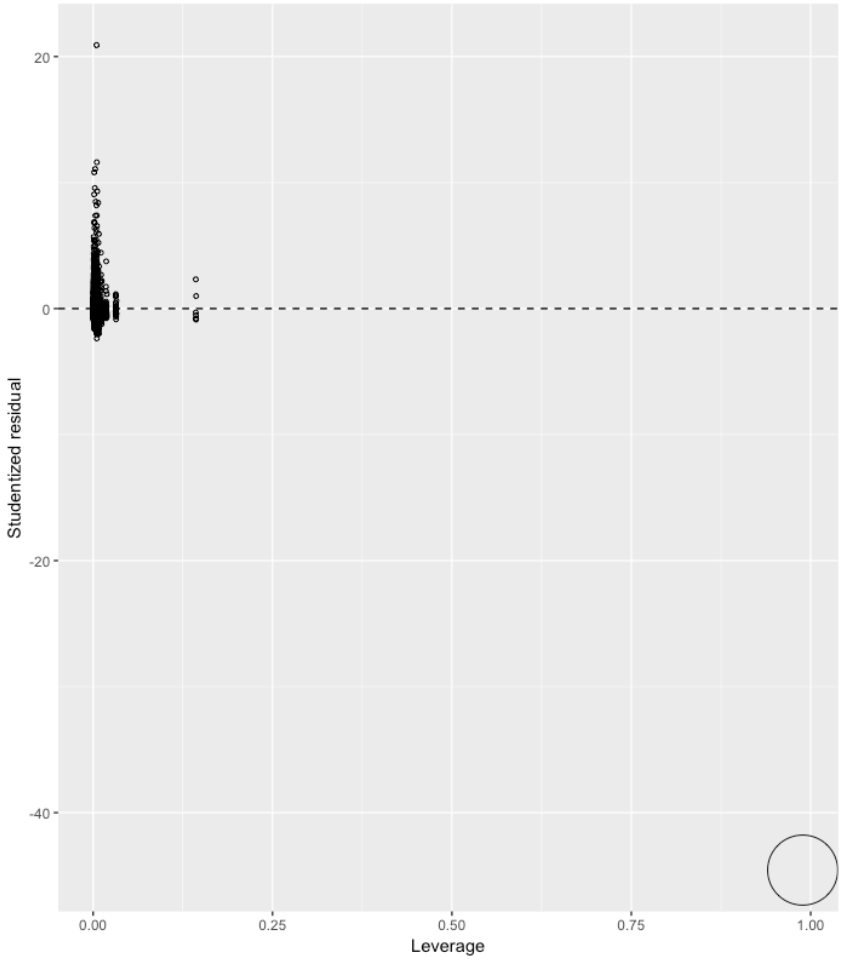

Figure 1

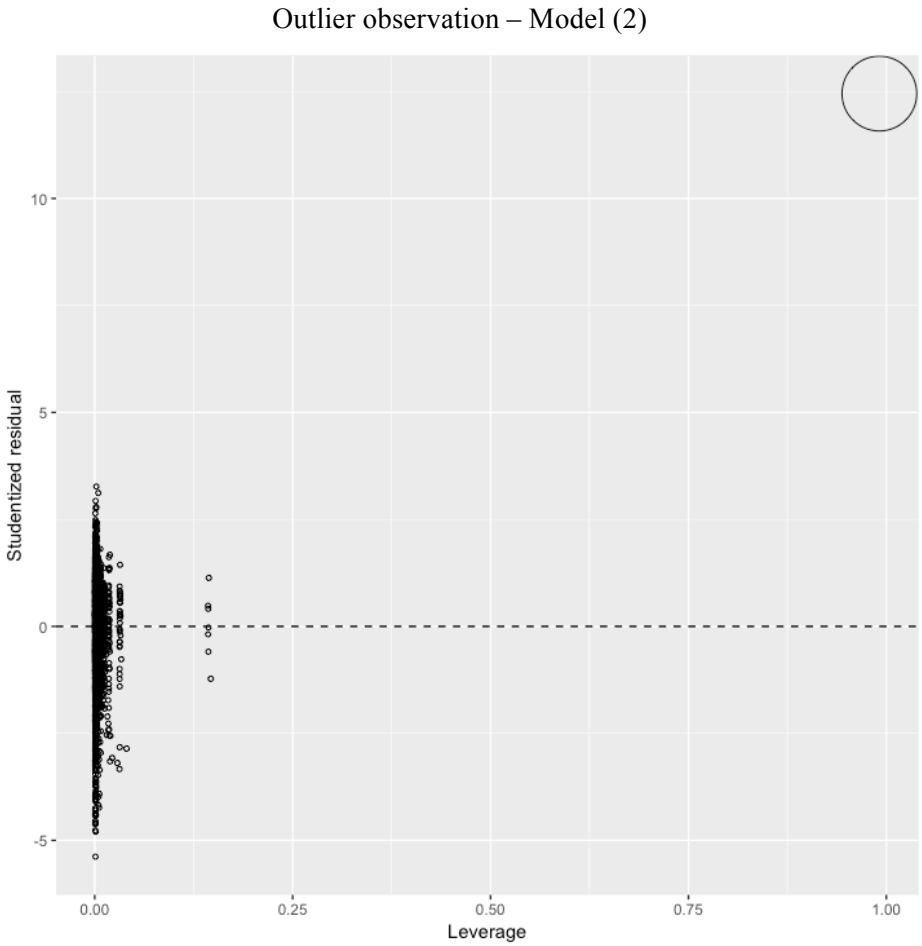

Figure 2

Figure 1 and 2 are bubble plots indicating the leverage (x axis), discrepancy (y axis), and influence (proportion) of each observation. As the two plots indicated, there are many outliers existed in both models.

After careful examination of the identified outliers in both models, one observation which has biggest leverage and influence in both models is identified as a miscoded data. The observation is summarized as follows:

Title	Gender	Production year	Genre	Voters	Rating
30 Minutes or Less	Male	2011	Action, Comedy, Crime	80960	6.1
Cinematic Predecessors	**Holiday Season**	**Adjusted budget**	**Adjusted box office gross**		
No	Yes	29,960,000,000	39,647,699		

Table 1

The problematic number is adjusted budget. The budget of this film is far more than the averaged budget and are around 756 times of its box office gross, which is rarely happened. Thus, this observation is identified as a misentered outlier and is omitted from the dataset. The model after omitting the outlier is marked as (1-1) and (2-1) and the adjusted R-squared of the two models increase by 0.1666 and 0.0073 respectively.

Detect non-normally distributed errors and non-constant variance errors

Model (1-1)

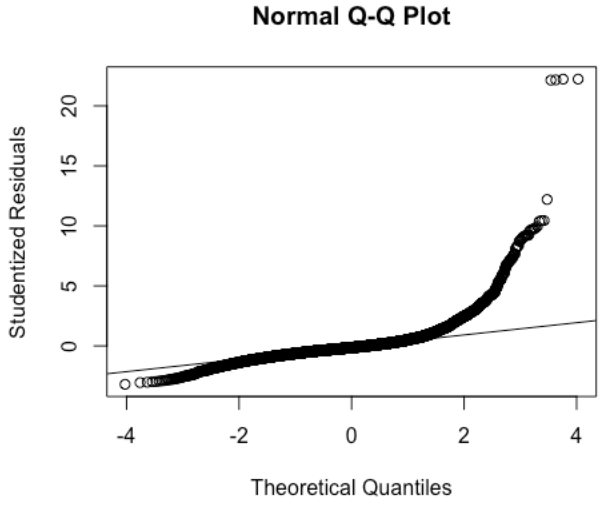

Figure 3

Model (2-1)

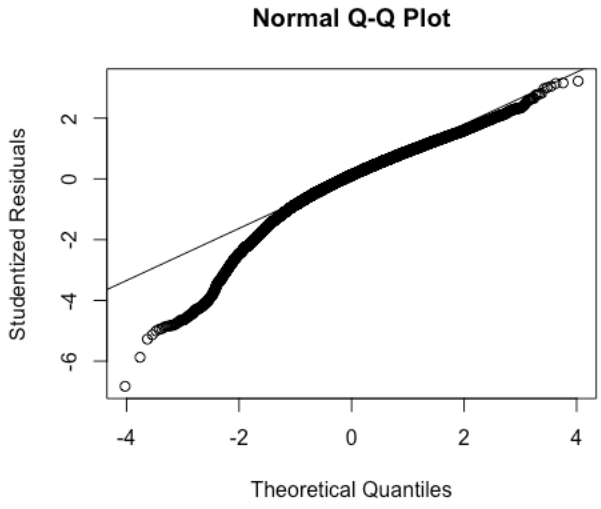

Figure 4

Figure 3 and 4 are normal quantile plots indicating the violation of the assumption of normal distributed errors in both models.

The studentized Breusch-Pagan test on both models are reported as follows:

	BP	df	p-value
Model (1-1)	611.41	39	2.2e-16
Model (2-1)	548.79	39	2.2e-16

Table 2

Hence, the both models violated the assumption of constant variance of errors.

To deal with the non-normally distributed errors and non-constant variance errors, Box-Cox method is applied.

Model (1-1)

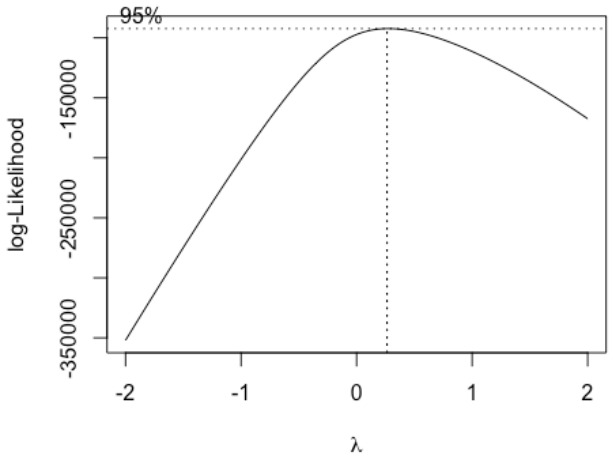

Figure 5

Model (2-1)

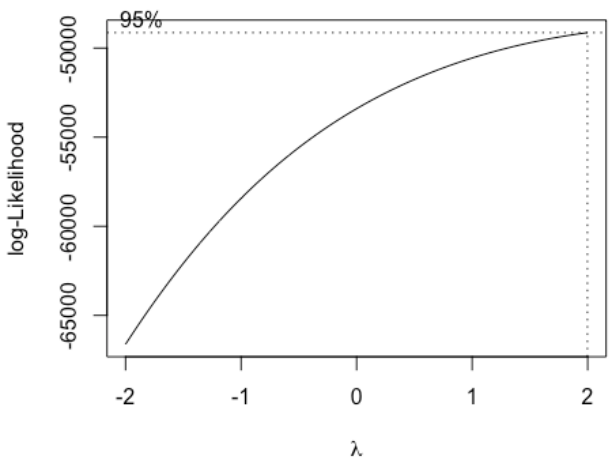

Figure 6

Figure 5 suggests the Box_office_gross^0.3 transformation and figure 6 suggests Rating^2 transformation. After the transformation, the two errors are well dealt with, as the following normal quantile plots Figure 7 and 8 indicated. The new models are marked Model (1-2) and Model (2-2).

Model (1-2)

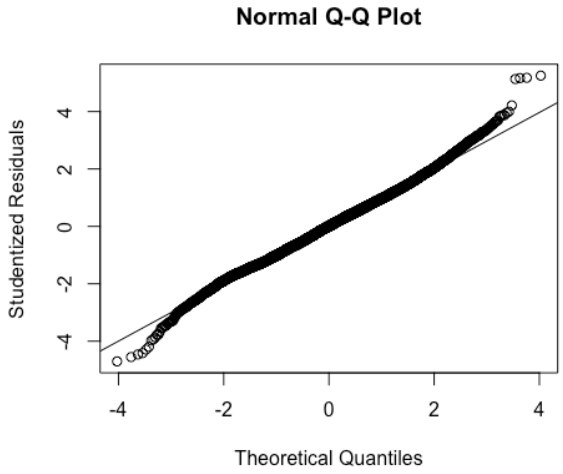

Figure 7

Model (2-2)

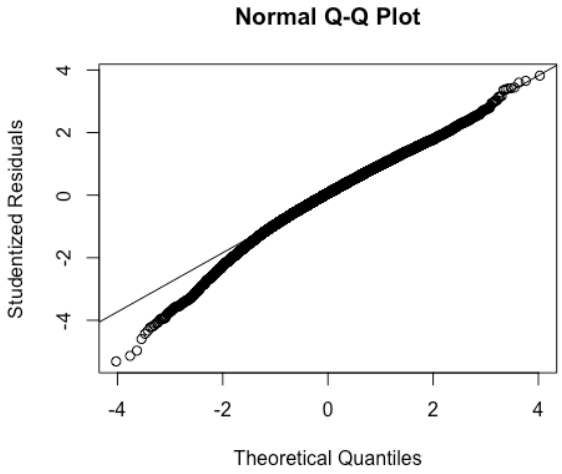

Figure 8

The adjusted R-squared of model (1-2) and (2-2) increase by 0.0953 and 0.0568 respectively after transformation.

Detect non-linearity in the data

The model (1-2) contains numeric variables rating and budget, thus the nonlinearity of the relationship between rating and box office gross and the relationship between budget and box office gross are examined.

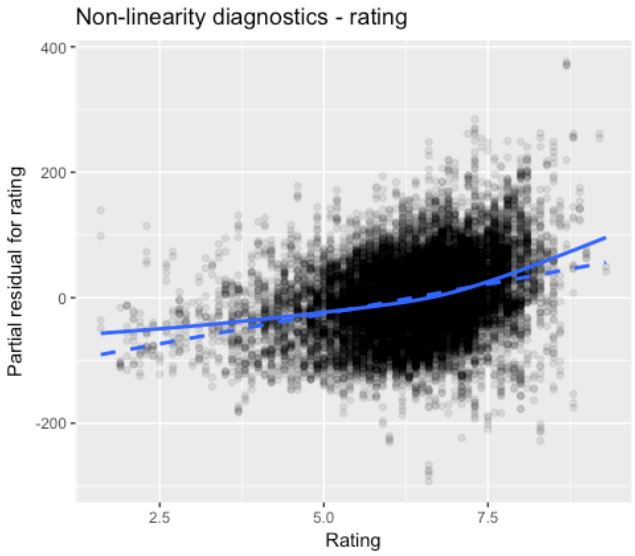

Figure 9

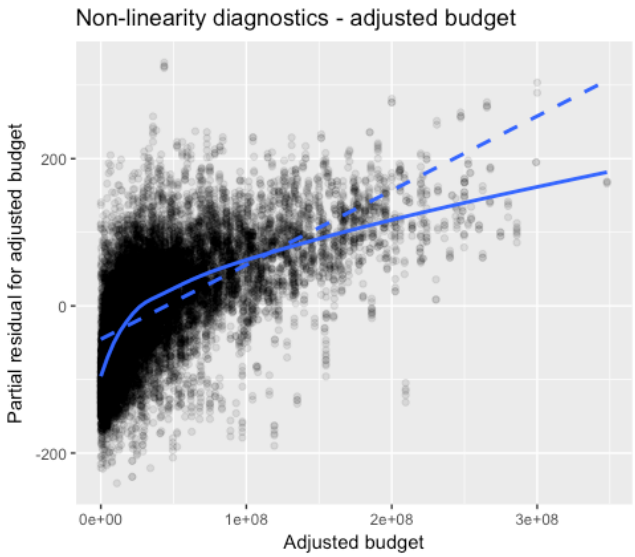

Figure 10

The partially residual plots Figure 9 indicates that for rating, the partial relationship with box_office_gross^0.3 is close to but not strictly linear. Figure 10 indicates that for budget, the partial relationship with box_office_gross^0.3 is not linear. A remedy to correct this is to square the rating term, and add a log-transformed term of budget.

The partial fits for the two explanatory variables after transformation are plotted below.

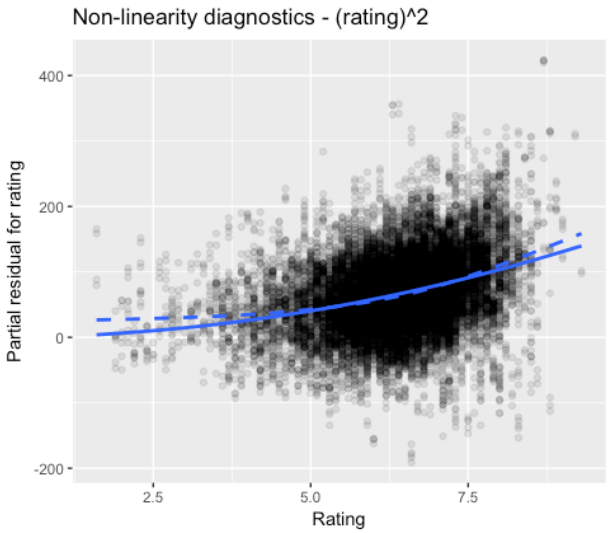

Figure 11

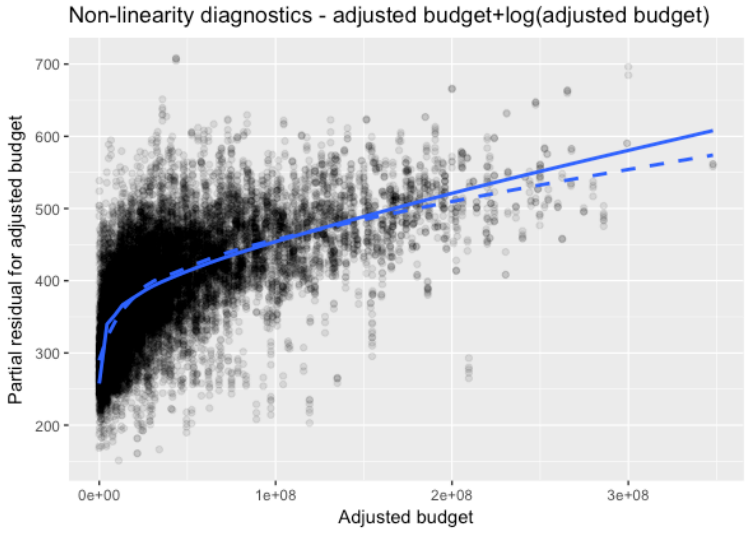

Figure 12

The partial fits and the partial residuals in Figure 11 and 12 overlap significantly, which means the revised model does a good job accounting for the nonlinearity.

The model after transformation is marked as (1-3) and the adjusted R-squared of model (1-3) increases by 0.0586.

The model (2-2) contains numeric variables voters, budget, and runtime, thus the nonlinearity of the relationship between voters and rating, relationship between budget and rating, and relationship between runtime and rating are examined.

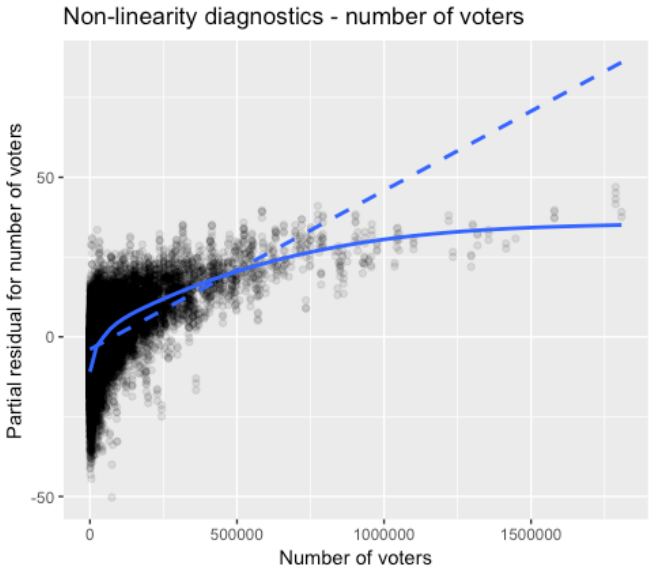

Figure 13

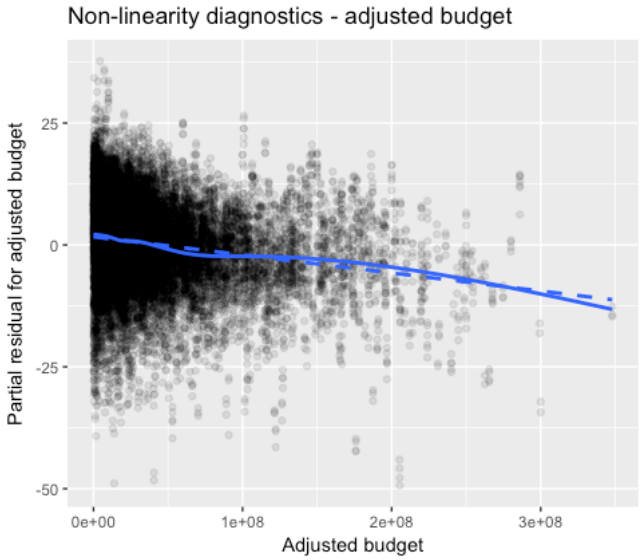

Figure 14

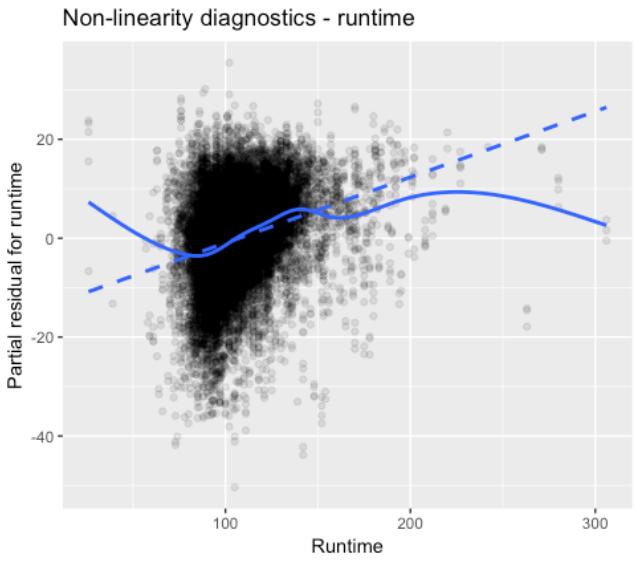

Figure 15

The partially residual plots Figure 13 indicates that for the number of voters, the partial relationship with rating^2 is not linear. Figure 14 indicates that for budget, the partial relationship with rating^2 is linear so no transformation needed. Figure 15 indicates that for runtime, the partial relationship with rating^2 is not linear. A remedy to correct this is to add a log-transformed term of voters and add a squared runtime term.

The partial fits for the two explanatory variables after transformation are plotted below.

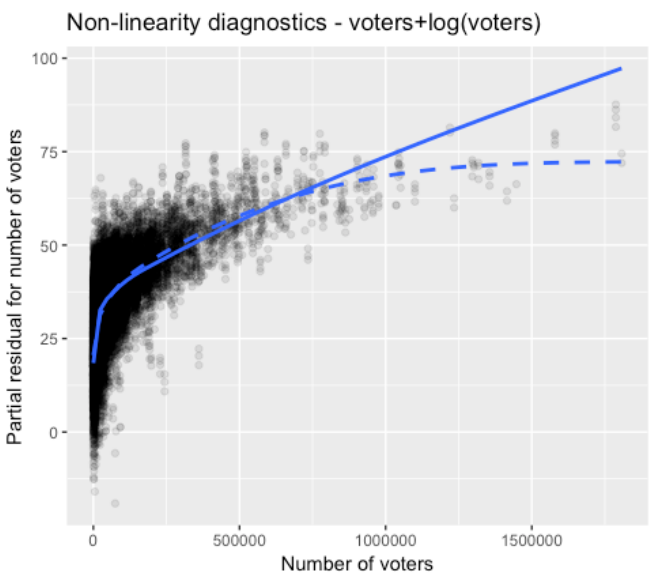

Figure 16

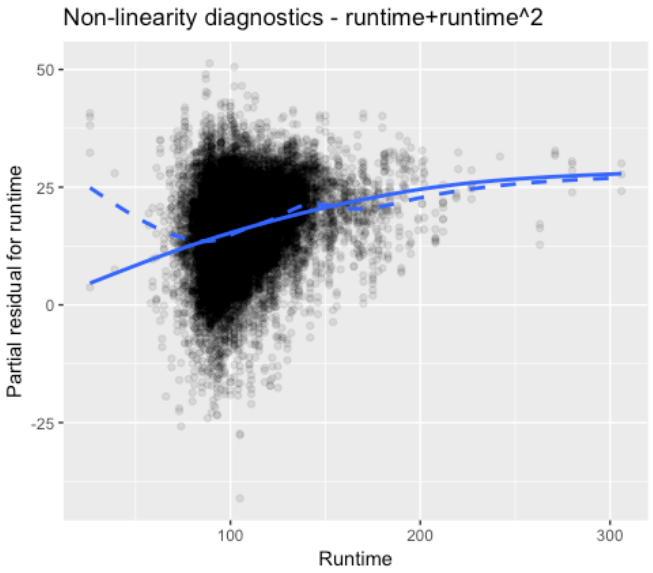

Figure 17

The overlap of partial fits and the partial residuals in Figure 16 and 17 are significantly improved, which means the revised model does a good job accounting for the nonlinearity.

The model after transformation is marked as (2-3) and the adjusted R-squared of model (2-3) increases by 0.0895.

Detect collinearity

The numerical variables rating and budget in model (1-3) are detected as no collinearity.

	rating	*stand_budget*
rating		0.097***
stand_budget	0.097***	

Computed correlation used pearson-method with listwise-deletion.

Table 3

The numerical variables voters, runtime, and budget in model (2-3) are detected as no collinearity.

	votes	*USA_runtime*	*stand_budget*
votes		0.317***	0.461***
USA_runtime	0.317***		0.333***
stand_budget	0.461***	0.333***	

Computed correlation used pearson-method with listwise-deletion.

Table 4

Regression Report

The final model (1-3) after transformation are

$$Box\ office\ gross_i{}^{0.3} = \beta_0 + \beta_1 Gender_i + \beta_2 Rating_i{}^2 + \beta_3 Cinematic\ predecessors_i +$$

$$\beta_4 Budget_i + \beta_5 \log(Budget_i) + \beta_6 Holiday\ season_i + \beta_7 Genre_i + \beta_8 Year_i + \varepsilon_i \qquad (1-3)$$

The evaluations of Model (1), Model (1-1), Model (1-2), Model (1-3) are listed below.

Commercial Success Model

	Dependent Variable			
	Box office gross		***Box office gross$^{0.3}$***	
	(1)	**(1-1)**	**(1-2)**	**(1-3)**
Gender	3.051e+06* (1.548e+06)	-3.279e+06* (1.379e+06)	-1.835e+00 (1.159e+00)	-2.911e+00** (1.093e+00)
Rating	2.943e+07*** (6.511e+05)	2.246e+07*** (5.874e+05)	1.907e+01*** (4.935e-01)	
Rating^2				1.618e+00*** (3.857e-02)
Cinematic predecessors	3.881e+07*** (1.630e+06)	2.230e+07*** (1.468e+06)	2.377e+01*** (1.234e+00)	2.264e+0*** (1.163e+00)
Budget	9.504e-03*** (1.631e-03)	9.545e-01*** (1.385e-02)	1.010e-06*** (1.163e-08)	5.148e-07*** (1.540e-08)
log(Budget)				2.186e+01*** (4.793e-01)
Holiday season	2.282e+07*** (1.369e+06)	1.034e+07*** (1.230e+06)	1.629e+01*** (1.034e+00)	1.354e+01*** (9.762e-01)
Genre (Adult)	1.524e+08 . (8.497e+07)	1.938e+08* (7.549e+07)	1.545e+02* (6.343e+01)	2.462e+02*** (5.984e+01)

Genre (Adventure)	1.603e+07*** (3.499e+06)	4.863e+06 (3.113e+06)	-6.077e-01 (2.616e+00)	-1.241e-01 (2.467e+00)
Genre (Animation)	1.949e+07*** (5.734e+06)	1.049e+07* (5.096e+06)	7.652e+00 . (4.282e+00)	7.356e+00 . (4.038e+00)
Genre (Biography)	-4.990e+07*** (5.264e+06)	-2.005e+07*** (4.697e+06)	-1.494e+01*** (3.947e+00)	-1.873e+01*** (3.724e+00)
Genre (Comedy)	-2.164e+07*** (2.937e+06)	4.040e+06 (2.637e+06)	4.611e+00* (2.215e+00)	6.229e+00** (2.089e+00)
Genre (Crime)	-3.823e+07*** (3.446e+06)	-1.058e+07*** (3.089e+06)	-9.718e+00*** (2.596e+00)	-1.047e+01*** (2.448e+00)
Genre (Documentary)	-7.571e+07*** (1.010e+07)	-1.394e+07 (9.018e+06)	-3.681e+01*** (7.577e+00)	-3.115e+00 (7.193e+00)
Genre (Drama)	-4.259e+07*** (2.842e+06)	-1.248e+07*** (2.564e+06)	-1.706e+01*** (2.154e+00)	-1.457e+01*** (2.032e+00)
Genre (Family)	1.204e+07** (4.059e+06)	1.406e+07*** (3.607e+06)	1.748e+01*** (3.031e+00)	1.501e+01*** (2.858e+00)
Genre (Fantasy)	8.703e+06* (4.045e+06)	6.985e+06 . (3.594e+06)	2.207e+00 (3.020e+00)	2.953e+00 (2.848e+00)
Genre (Film-Noir)	-7.445e+07*** (1.340e+07)	-4.207e+07*** (1.192e+07)	-4.312e+01*** (1.001e+01)	-3.644e+01*** (9.443e+00)
Genre (History)	-4.095e+07*** (6.145e+06)	-3.073e+07*** (5.462e+06)	-2.753e+01*** (4.590e+00)	-3.157e+01*** (4.329e+00)
Genre (Horror)	-3.547e+07*** (4.067e+06)	-2.394e+05 (3.650e+06)	-3.384e-01 (3.067e+00)	8.999e+00** (2.899e+00)
Genre (Music)	-2.959e+07*** (5.928e+06)	3.833e+06 (5.290e+06)	4.478e+00 (4.445e+00)	3.611e+00 (4.191e+00)
Genre (Musical)	-2.145e+07*** (6.281e+06)	-9.107e+06 (5.583e+06)	-8.914e+00 . (4.691e+00)	-1.300e+01** (4.424e+00)
Genre (Mystery)	-3.378e+07*** (4.205e+06)	-9.376e+06* (3.753e+06)	-3.978e+00 (3.154e+00)	-5.462e+00 . (2.974e+00)
Genre (News)	-1.232e+08 (8.492e+07)	-5.065e+07 (7.545e+07)	-1.174e+02 . (6.340e+01)	-8.860e+01 (5.979e+01)
Genre (Romance)	-2.927e+07*** (3.287e+06)	-3.742e+06 (2.944e+06)	-3.234e+00 (2.474e+00)	-2.802e+00 (2.332e+00)
Genre (Sci-Fi)	9.933e+05 (4.026e+06)	9.344e+05 (3.578e+06)	-5.240e+00 . (3.006e+00)	-1.384e+00 (2.836e+00)
Genre (Short)	-5.513e+07 (6.007e+07)	2.805e+06 (5.338e+07)	4.591e+01 (4.485e+01)	4.919e+01 (4.229e+01)
Genre (Sport)	-3.352e+07*** (6.236e+06)	-4.749e+06 (5.557e+06)	2.749e+00 (4.669e+00)	-4.984e-02 (4.403e+00)

Genre (Thriller)	-2.957e+07*** (3.124e+06)	-8.714e+06** (2.792e+06)	-5.837e+00 * (2.346e+00)	-5.502e+00* (2.212e+00)
Genre (War)	-3.547e+07*** (5.743e+06)	-2.221e+07*** (5.107e+06)	-2.193e+01*** (4.291e+00)	-2.634e+01*** (4.047e+00)
Genre (Western)	-4.362e+07*** (6.925e+06)	-2.609e+07*** (6.158e+06)	-2.175e+01*** (5.174e+00)	-2.264e+01*** (4.879e+00)
Year (1920-1929)	-5.085e+07* (2.192e+07)	-5.798e+07** (1.948e+07)	-4.070e+01* (1.637e+01)	-7.178e+01*** (1.544e+01)
Year (1930-1939)	-4.732e+07* (2.070e+07)	-6.128e+07*** (1.839e+07)	-5.087e+01*** (1.545e+01)	-8.965e+01*** (1.459e+01)
Year (1940-1949)	-2.091e+07 (2.013e+07)	-3.779e+07* (1.789e+07)	-2.414e+01 (1.503e+01)	-6.762e+01*** (1.420e+01)
Year (1950-1959)	-2.387e+07 (1.979e+07)	-3.794e+07* (1.759e+07)	-3.484e+01* (1.478e+01)	-7.588e+01*** (1.396e+01)
Year (1960-1969)	3.178e+07 (1.989e+07)	-1.717e+06 (1.768e+07)	-1.115e+01 (1.485e+01)	-5.467e+01*** (1.403e+01)
Year (1970-1979)	2.881e+07 (1.977e+07)	7.471e+06 (1.757e+07)	-3.428e+00 (1.477e+01)	-4.119e+01** (1.394e+01)
Year (1980-1989)	-2.230e+07 (1.960e+07)	-4.648e+07** (1.741e+07)	-4.504e+01** (1.463e+01)	-8.800e+01*** (1.383e+01)
Year (1990-1999)	-2.102e+07 (1.955e+07)	-6.182e+07*** (1.738e+07)	-6.258e+01*** (1.461e+01)	-1.039e+02*** (1.380e+01)
Year (2000-2009)	-1.754e+07 (1.954e+07)	-6.350e+07*** (1.737e+07)	-7.143e+01*** (1.460e+01)	-1.090e+02*** (1.379e+01)
Year (2010-2016)	-1.632e+07 (1.956e+07)	-6.243e+07*** (1.739e+07)	-7.176e+01*** (1.461e+01)	-1.078e+02*** (1.380e+01)
Constant	-1.018e+08*** (2.004e+07)	-6.693e+07*** (1.782e+07)	6.907e+01*** (1.497e+01)	-1.851e+02*** (1.547e+01)
R^2	0.2111	0.3774	0.4724	0.531
Adjusted R^2	0.2094	0.376	0.4713	0.5299
Residual Std. Error	84860000 (df = 17639)	75400000 (df = 17636)	63.35 (df = 17636)	59.74 (df = 17635)
F Statistics	121.1*** (df = 39; 17639)	274.1*** (df = 39; 17636)	404.9*** (df = 39; 17636)	499.1*** (df = 40; 17635)

Note: #Observations = 17676 for all models presented above. . p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 5

The final model (2-3) after transformation are

$$Rating_i^2 = \beta_0 + \beta_1 Gender_i + \beta_2 Voters_i + \beta_3 \log(Voters_i) + \beta_4 Cinematic\ predecessors_i + \beta_5 Budget_i + \beta_6 Runtime_i + \beta_7 Runtime_i^2 + \beta_8 Genre_i + \beta_9 Year_i + \varepsilon_i \qquad (2-3)$$

The evaluations of Model (2), Model (2-1), Model (2-2), Model (2-3) are listed below.

Aesthetic Success Model

	Dependent Variable			
		Rating	*Rating²*	
	(1)	**(1-1)**	**(1-2)**	**(1-3)**
Gender	4.906e-02** (1.511e-02)	6.056e-02*** (1.505e-02)	7.752e-01*** (1.737e-01)	7.570e-01*** (1.594e-01)
Voters	3.421e-06*** (5.137e-08)	3.661e-06*** (5.375e-08)	4.976e-05*** (6.204e-07)	2.930e-05*** (6.758e-07)
log(Voters)				3.185e+00*** (5.663e-0)
Cinematic predecessors	-2.948e-01*** (1.582e-02)	-2.595e-01*** (1.592e-02)	-3.029e+00*** (1.837e-01)	-3.423e+00*** (1.690e-01)
Budget	-1.864e-11 (1.591e-11)	-2.435e-09*** (1.689e-10)	-3.705e-08*** (1.950e-09)	-6.150e-08*** (1.841e-09)
Runtime	9.188e-03*** (3.590e-04)	1.080e-02*** (3.741e-04)		1.835e-01*** (1.767e-02)
Runtime²			1.331e-01*** (4.317e-03)	-3.044e-04*** (6.819e-05)
Genre (Adult)	-7.457e-01 (8.286e-01)	-7.571e-01 (8.239e-01)	-9.707e+00 (9.509e+00)	-8.684e+00 (8.733e+00)
Genre (Adventure)	1.207e-01*** (3.410e-02)	1.502e-01*** (3.397e-02)	2.023e+00*** (3.921e-01)	2.440e+00*** (3.601e-01)
Genre (Animation)	7.147e-01*** (5.627e-02)	7.831e-01*** (5.615e-02)	9.897e+00*** (6.481e-01)	9.977e+00*** (5.981e-01)
Genre (Biography)	7.510e-01*** (5.139e-02)	6.722e-01*** (5.139e-02)	8.470e+00*** (5.932e-01)	8.570e+00*** (5.446e-01)
Genre (Comedy)	2.100e-01*** (2.883e-02)	1.698e-01*** (2.881e-02)	2.129e+00*** (3.325e-01)	2.363e+00*** (3.054e-01)

Genre (Crime)	3.757e-01*** (3.357e-02)	3.174e-01*** (3.364e-02)	3.732e+00*** (3.883e-01)	3.474e+00*** (3.565e-01)
Genre (Documentary)	1.164e+00*** (9.853e-02)	1.061e+00*** (9.823e-02)	1.430e+01*** (1.134e+00)	1.817e+01*** (1.043e+00)
Genre (Drama)	5.049e-01*** (2.768e-02)	4.374e-01*** (2.793e-02)	5.313e+00*** (3.223e-01)	5.987e+00*** (2.963e-01)
Genre (Family)	2.614e-01*** (3.976e-02)	2.853e-01*** (3.957e-02)	3.941e+00*** (4.567e-01)	4.196e+00*** (4.202e-01)
Genre (Fantasy)	9.361e-02* (3.949e-02)	1.092e-01** (3.928e-02)	1.620e+00*** (4.534e-01)	1.614e+00*** (4.164e-01)
Genre (Film-Noir)	7.350e-01*** (1.307e-01)	6.837e-01*** (1.300e-01)	8.631e+00*** (1.500e+00)	7.801e+00*** (1.377e+00)
Genre (History)	4.405e-01*** (6.041e-02)	3.951e-01*** (6.015e-02)	5.077e+00*** (6.942e-01)	6.428e+00*** (6.379e-01)
Genre (Horror)	-8.148e-02* (3.977e-02)	-1.504e-01*** (3.983e-02)	-1.677e+00*** (4.598e-01)	-2.372e+00*** (4.226e-01)
Genre (Music)	3.703e-01*** (5.789e-02)	3.055e-01*** (5.774e-02)	4.037e+00*** (6.664e-01)	4.726e+00*** (6.118e-01)
Genre (Musical)	3.477e-01*** (16.126e-02)	3.337e-01*** (6.092e-02)	4.279e+00*** (7.031e-01)	5.101e+00*** (6.456e-01)
Genre (Mystery)	3.193e-01*** (4.097e-02)	2.661e-01*** (4.091e-02)	3.058e+00*** (4.722e-01)	2.490e+00*** (4.336e-01)
Genre (News)	2.417e+00** (8.280e-01)	2.313e+00** (8.233e-01)	3.150e+01*** (9.502e+00)	3.409e+01*** (8.723e+00)
Genre (Romance)	3.162e-01*** (3.211e-02)	2.643e-01*** (3.214e-02)	3.127e+00*** (3.709e-01)	3.485e+00*** (3.407e-01)
Genre (Sci-Fi)	-6.049e-02 (3.932e-02)	-6.056e-02 (3.910e-02)	-5.759e-01 (4.512e-01)	-5.924e-01 (4.143e-01)
Genre (Short)	2.469e+00*** (5.862e-01)	2.494e+00*** (5.829e-01)	3.208e+01*** (6.728e+00)	3.858e+01*** (6.199e+00)
Genre (Sport)	4.350e-01*** (6.089e-02)	3.765e-01*** (6.068e-02)	4.816e+00*** (7.003e-01)	5.067e+00*** (6.431e-01)
Genre (Thriller)	2.007e-01*** (3.046e-02)	1.555e-01*** (3.045e-02)	1.787e+00*** (3.515e-01)	1.390e+00*** (3.227e-01)
Genre (War)	4.116e-01*** (5.625e-02)	3.668e-01*** (5.602e-02)	4.683e+00*** (6.465e-01)	5.385e+00*** (5.936e-01)
Genre (Western)	3.902e-01*** (6.755e-02)	3.551e-01*** (6.721e-02)	4.233e+00*** (7.758e-01)	5.851e+00*** (7.127e-01)
Year (1920-1929)	8.240e-01*** (2.138e-01)	8.689e-0*** (2.126e-01)	1.223e+01*** (2.453e+00)	1.002e+01*** (2.253e+00)

Year (1930-1939)	6.144e-01** (2.019e-01)	6.726e-01*** (2.007e-01)	9.519e+00*** (2.317e+00)	6.052e+00** (2.128e+00)
Year (1940-1949)	5.680e-01** (1.963e-01)	6.261e-01** (1.952e-01)	8.828e+00*** (2.253e+00)	5.483e+00** (2.070e+00)
Year (1950-1959)	2.917e-01 (1.930e-01)	3.373e-01 . (1.919e-01)	5.180e+00* (2.215e+00)	-1.400e+00 (2.035e+00)
Year (1960-1969)	2.337e-01 (1.940e-01)	2.912e-01 (1.929e-01)	4.886e+00* (2.227e+00)	-6.411e-01 (2.047e+00)
Year (1970-1979)	1.939e-01 (1.928e-01)	2.403e-01 (1.917e-01)	3.996e+00 . (2.213e+00)	-2.848e+00 (2.036e+00)
Year (1980-1989)	-1.409e-01 (1.911e-01)	-8.214e-02 (1.900e-01)	-1.734e-01 (2.193e+00)	-7.325e+00*** (2.018e+00)
Year (1990-1999)	-3.227e-01 . (1.907e-01)	-2.346e-01 (1.897e-01)	-2.089e+00 (2.189e+00)	-9.703e+00*** (2.015e+00)
Year (2000-2009)	-4.656e-01* (1.906e-01)	-3.708e-01 . (1.896e-01)	-3.807e+00 . (2.188e+00)	-1.222e+01*** (2.014e+00)
Year (2010-2016)	-3.866e-01* (1.908e-01)	-2.984e-01 (1.898e-01)	-3.186e+00 (2.190e+00)	-1.237e+01*** (2.018e+00)
Constant	5.123e+00*** (1.954e-01)	4.977e+00*** (1.946e-01)	2.358e+01*** (2.246e+00)	5.689e-02 (2.281e+00)
R^2	0.368	0.3754	0.432	0.5214
Adjusted R^2	0.3667	0.374	0.4308	0.5203
Residual Std. Error	0.8275 (df = 17639)	0.8228 (df = 17636)	9.496 (df = 17636)	8.717 (df = 17634)
F Statistics	263.4*** (df = 39; 17639)	271.7*** (df = 39; 17636)	344*** (df = 39; 17636)	468.6*** (df = 41; 17634)

Note: #Observations = 17676 for all models presented above. . p<0.1; * p<0.05; ** p<0.01, *** p<0.001

Table 6