

Research Question/Title

How hard are movies with different gender of top billing cast get success respectively by time?

1. Introduction

Acting a prominent role in popular culture, films 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 (Haskell, 1987; Rosen, 1973), and besides mirroring the sociocultural images, films also create them (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 researchers relevant to this topic focus on the portrayal of different genders in the motion picture industry, this research will empirically examine the relationship between gender of leading cast and cinematic success over time, and will try to fit in a logit model to define the gender of top billing cast of a film concerning its cinematic success. More specifically, this research wants 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.

**How do you answer the question? Data, theory, methods*

**Summarize the results*

2. Literature Review

2.1 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). Thus, 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., De Vany & Walls, 1999; Dodds & Holbrook, 1988; Litman & Kohl, 1989; Prag & Casavant, 1994; Wallace, Seigerman, & Holbrook, 1993). 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.

2.1.1 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; Dickie, 1988, 1997; Savile, 1982; 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 (Boor, 1992; Holbrook, 1999; Wanderer, 1970). Moreover, since individual judge may be prone to judgement errors and short-sighted and consumers largely outnumber critics 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 and would be the predictor of aesthetic success of films in this research.

2.1.2 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, 2005b).

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 versions, and television showings, 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.

2.2 Gender and Film

2.2.1 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 behavior (Levy and Carter 1989; Nihlen and 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).

Children develop gender schemas from the images and characters they encounter in the media, where films play a prominent role in. The amount of time children spend using the 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, and 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.

2.2.2 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 and most of them have found gender hegemony present behind and in front of the camera and not changed much over time (Steinke, 1999, 2004; Smith & Marc, 2010; Smith, 2013; Downs, 1981; Barcus, 1983; Downing, Mohammadi & Sreberny-Mohammadi, 1990), which 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. Another researcher (Witt, 2000) found approximately two-thirds of characters on television are male, a percentage that has stayed consistent since the 1950s. Witt (2000) reported on several studies from 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 & Donnerstein (1998) found that in movies and

on TV, males are often characterized as being prone to violent behavior, and 61% television programs contain some violence, 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.

3. Method

3.1 IMDb Data

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), we will confine the research 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.

The dataset obtained is 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 tools².

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).

3.2 Model & Measurement

¹ <http://www.imdb.com/>

² <http://www.imdb.com/interfaces>

The IMDb dataset contains various film information, among which this research will focus on the information about title, cast, aesthetic success indicator consumer rating, commercial success indicator box office gross, production year, and genre.

The logit model this research tries to fit in is in the formula as

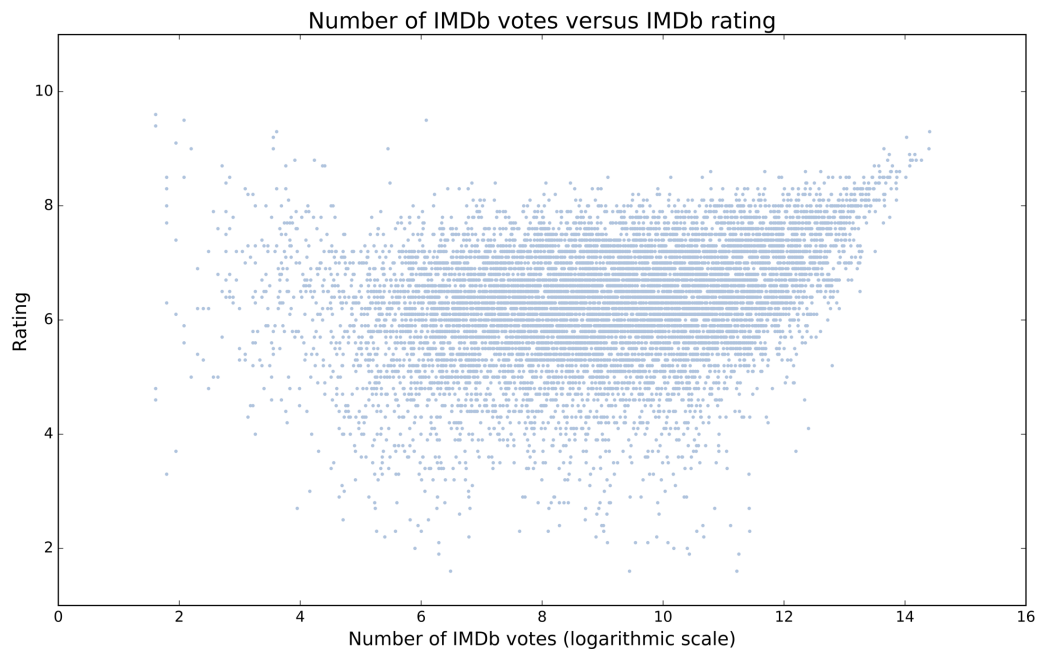
$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \varepsilon_i$$

where independent variable Y_i is binary, indicating whether the film is female leading or male leading; X_1 is a categorical variable indicating the production year range of the film; X_2 is a categorical variable indicating the genre of the film; X_3 is a numerical variable indicating the consumer rating; X_4 is a numerical variable indicating the standardized box office gross.

The IMDb dataset used in this research is updated on May 16th, 2017. Among 4,282,600 media products available in the dataset, 10,080 USA films are filtered which have complete information on the gender of top-billing cast, consumer rating and number of voters, domestic box office gross, genre, and are produced before 2017.

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). 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 can be identified with cast gender and billing order of each film in the IMDb dataset.

The IMDb rating is the weighted vote average on a scale of one to ten. IMDb has applied various filters which are not disclosed to the raw data in order to avoid ballot stuffing, and IMDb claims that the weighted vote average is a more accurate vote average. Each IMDb rating comes with both a numeric rating and the number of votes cast. Concerned that less well-known films with fewer votes would be unreliably rated, a close glance at the rating data can be observed as below.



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 to the data. To add the reliability of ratings, films with fewer than 100 votes ($n = 287$) are dropped from the dataset.

The box office information is heavily affected by inflation and commercialization of motion picture industry thus are not comparable across years. Since this research only cares about the commercial success of a film in specific time slot instead of measuring its success overtime, we simply normalize the box office gross. More specifically, the highest box office gross in each time slot is normalized to 100, and the other box office grosses in that time slot are computed accordingly.

Since one film could be defined as several genres, in this research, the same films might be repeated in different genres.

Variable Information ($n = 9793$)

Gender	Male	Female
Count	2530	7263

Variable Information (n = 9793)

	Min.	25%	50%	75%	Max.	Mean	Std.
Rating	1.6	5.7	6.4	7.0	9.5	6.28	1.04
Box Office Gross	1,003	557,683	6,607,652	28,927,720	1,055,920,000	25,864,440	51,491,160
Standized Box Office Gross	0.00	0.09	1.53	5.37	100	4.63	8.70
Production Year	1913				2016		

Variable Information (n = 26532)

Genre	Action	Adult	Adventure	Animation	Biography	Comedy	Crime
Count	1656	2	1367	325	560	3951	1646
	Documentary	Drama	Family	Fantasy	Film-Noir	Horror	Music
	514	5230	905	849	57	863	468
	Musical	Mystery	News	Romance	Sci-Fi	Short	Sport
	323	832	11	2419	817	23	362
	Thriller	War	Western				
	2336	381	285				