The Impact of Female Characters on a Movie's Return on Investment

Raghav Lakhotia, Sameer Ganu, Anshul Goel, Abhishek Kumar

Abstract— In the age and times where women's empowerment is a significant topic of discussion, we aim to analyze the potential gender diversity influence on box office revenues. The following research is carried out by collecting data from 400 Hollywood movies between the years 2014-2017 and performing regression analysis to find a correlation between the presence of female characters in movies and their return on investment (ROI). The paper finds that there is a positive relationship between the performance of the movies (its ROI) and the gender diversity i.e. the more the number of female characters, the higher the revenue generated. Another factor such as Number of Votes also has a direct impact on the revenue of the movie. The research not only takes into consideration the mere presence of women on screen but also the exchange of at least one dialogue among themselves, which is presented by the Bechdel Score of the movie.

Keywords—Bechdel, diversity, Hollywood, return on investment.

I. Introduction

THE entertainment industry, particularly the film and movie industry, has seen tremendous growth in the last two decades. As per Film and Movie Statistics [1], the global box office revenue has been forecasted to increase from 38 billion USD to a whopping 50 billion USD. The introduction of multiscreen Cineplex complexes added a new dimension to the industry which made it possible to showcase multiple films at the same time, thus generating more revenue and meeting the exponential customer demand. The technological advancement from 2-D to 3-D and now 4-D, the graphics, Ultra HD, Artificial reality and the most recent being drones in the cinema have completely revolutionized the industry. But does this technological advancement instrumental in providing equal opportunity to female characters? The issue of gender representation is inarguably the most urgent and prominent one in the industry. It is often said that the movie industry is still gender biased. We see more male characters than the female counterparts. As per the report entitled "Inequality in 900 popular films" [2], it is observed that the percentage of female speaking characters has moved from 29% in 2009 to 32% in 2016. Only 12% of the films out of 900 depict the correct balance between the gender representations. With a dismal ratio of 2.3 males to every single female, we can conclude that the industry is biased towards men. The reports

also points at interesting facts across 1438 content creators where only 4.2% of the directors on set are female, 13.2% are female writers, 20.7% represent female producers and 1.7% represent female composers. On having a closer look at the earnings of the top 10 actors and actresses in Hollywood [3], we see a huge gender pay gap. The differences in the paycheck amounts to even 3X. Where the top 10 actors banked a cumulative $488.5 million, the top 10 earning women banked only $172.5 million.

II. Motivation for Our Research

Similar enduring patterns of gender inequality have been observed throughout the paid labor force, but it is the film industry that represents so much complexity in employment and pay. Over the past few years, both the government and corporates have worked towards increasing women's representation in the workplace. Even research has concluded that inclusion of women in a Board of Directors resulted in a positive outcome i.e. increase in sales/revenue for several companies across the globe [4]-[6].

If corporate entities can take a noteworthy stance in bridging the gender gap, then why not entertainment industry? Do we require a law to culminate this inequality gap in the industry? The question is why does entertainment industry seems to be so far behind in dealing with this issue. We rarely see films that are powered by ineffaceable female protagonists. It is also not difficult to miss the ever-existing perceptible male dominance. Can all this be attributed to the dearth of female directors, screenwriters and other people behind the camera? In other words, having greater diversity behind the scenes going to change this? Exorbitant paychecks for a male actor clearly points that the industry sees huge returns from a role played by a male character as compared to a role played by a female character. But is it really true?

To have a better understanding of the relationship between the women in the cast and the performance of the movie (e.g. ROI), it is important to study the impact of gender diversity on movie performance. In this paper, movie diversity refers to the number of females in the entire movie cast. For analyzing this effect, we will use regression analysis technique to identify the factors that contribute towards influencing the ROI of a movie and the degree of correlation among them. The data used in the research is collected from Internet Movie Database (IMDb) covering 385 movies between the years 2014-2017 and each having at least one female in the entire movie cast. IMDb is the most trusted website which hosts 180 million data items including movies, TV shows, cast and crew members. Data was scrapped from the IMDb [7] using python code. The

Raghav Lakhotia is with General Electric, India, under the Digital Technology Leadership Program (corresponding author, phone: +91-9591192190; e-mail: raghav.lakhotia.2015@iimu.ac.in).

Sameer Ganu, Anshul Goel, and Abhishek Kumar are with General Electric, India under the Digital Technology Leadership Program. (e-mail: ganusameer90@lgmail.com, anshulgoel23@gmail.com, akiitbhu@gmail.com).

following section of the paper highlights the exact methodology used in the performing the analysis. Fig. 1 describes the journey of our research starting from data extraction to the end results.

Fig. 1 Steps involved in our research

III. RESEARCH METHOD

A. Research Strategy

The study uses the quantitative method. The research is not restricted to one genre of the movie. It covers all the genres: adventure, comedy, drama, romance for the period 2014-2017.

B. Data

We decided to consider only those movies which have at least one female member in the entire cast. The cast here means all the artists who were seen on screen during the movie irrespective of their role duration. There were certain movies for which either budget or gross US income were not available. Hence, we decided not to consider these movies record as we were unable to calculate the ROI for such movies. After deleting the movies record with such irregularities or missing data, we were left with 385 movies spread across the years 2014-2017.

C. Dependent and Independent Variables

The dependent variable is the ROI for the movie which is calculated using Gross Income (GIC) and Budget of the movie (BUG). Although, we considered taking Number of Votes (NOV) as the dependent variable, it is not necessary that all the viewers would have voted for that movie. The first independent variable is the gender ratio (GER). This variable is expressed as the percentage of females in the cast to that of the total cast size. The second independent variable is the number of countries (CON) where the movie was released. The third independent variable is the language (LAN) in which the movie was released.

D. Control Variable

To minimize the bias, control variables are introduced in the regression. The first control variable is the Rating (RAT) of the movie. The second control variable is the Bechdel Test Score (BTS). Bechdel test is a famous test named after American cartoonist Alison Bechdel, which aims to gauge the fact if two women in a work of fiction are talking among themselves and on something other than men. On the blog named after Bechdel [8], the Bechdel score of the sample movies was collected for the research based on the following criteria:

a. If the Bechdel score is 3, all the conditions are satisfied by the movie and the Bechdel rating given by us is 1.
b. Else, if the Bechdel score is either 0, 1 or 2, one of the

three conditions [8] is not satisfied by the movie, and the Bechdel rating assigned is 0.

The third control variable is the Number of Votes (NOV) for a movie.

IV. CONCEPTUAL MODEL

The work aims to highlight the effect of having more female characters in the movie cast. Hence, the hypothesis is as follows:

H1: The presence of more women in a movie cast is positively related to the ROI of the movie.

To test the hypothesis, regression analysis will be performed. The regression function is as follows:

$$ROI = \alpha + \beta_1 GER + \beta_2 CON + \beta_3 LAN + \beta_4 RAT + \beta_5 BTS + \beta_6 NOV + \varepsilon \qquad (1)$$

where, GER = Gender Ratio of a movie cast, CON = Number of countries where a movie was released, LAN = Number of languages in which a movie was released, RAT = IMDB Rating, BTS = Bechdel Score Passed/Failed, NOV = Number of Votes

V. RESULTS

Regression analysis was performed in Microsoft Excel which will help to answer the hypothesis. First, the descriptive statistics of all the variables are presented. Then, the correlation result between the variables is discussed. In the end, the result of multiple linear regression is presented.

A. Descriptive Statistics

The total sample size is 385 observations. Table I shows the statistical description of the variables.

The variables are tested for normality. The input sample contains variables that are not normally distributed namely: ROI, Gender Ratio, Country, Language and Number of Votes. For all the variables, except for the Number of Votes, we found the mean of the sample and compared each individual sample with its sample mean. The variables are assigned the category depending upon the degree of variation of its observed value with the mean. For example, if the gender ratio is greater than or equal to 0 and less than 0.1, then it falls under category 0. Similarly, if the gender ratio is greater than or equal to 0.1 and less than 0.2, then it falls under category 2 and so on. The logarithmic operation was performed on the

variable Number of Votes to transform it to a normal form. Normalization check was performed by going through the descriptive statistics for the variables, as shown in Table I. A window of -2.0 to +2.0 on kurtosis and skewness [9]-[12] for a variable was set as the baseline to decide if the variable is normalized or not. Figs. 2 (a)-(d) show the normally distributed graph for certain normalized variables.

TABLE I
STATISTICAL DESCRIPTION OF THE VARIABLES

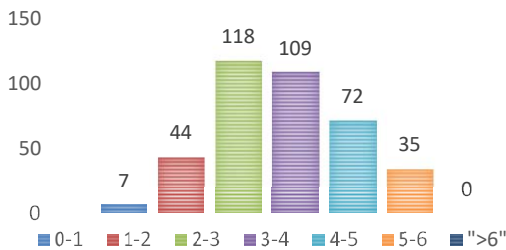|	ROI	GER	CON	LAN	RAT	BTS	NOV
Mean	1.40	2.78	1.52	0.18	6.46	2.15	1.94
Standard Error	0.16	0.06	0.03	0.02	0.05	0.05	0.02
Median	0.00	3.00	1.00	0.00	6.40	3.00	1.95
Standard Deviation	3.16	1.20	0.63	0.39	0.91	1.05	0.46
Sample Variance	9.99	1.43	0.40	0.15	0.82	1.10	0.21
Kurtosis	-1.20	-0.56	1.21	0.75	0.19	-0.94	0.84
Skewness	0.56	0.09	1.05	1.66	-0.36	-0.74	-0.51

B. Correlation

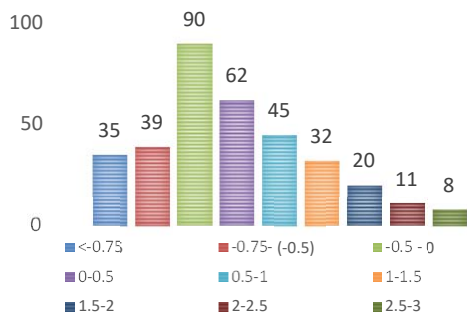Fig. 2 (a) Frequency Distribution of Gender Ratio
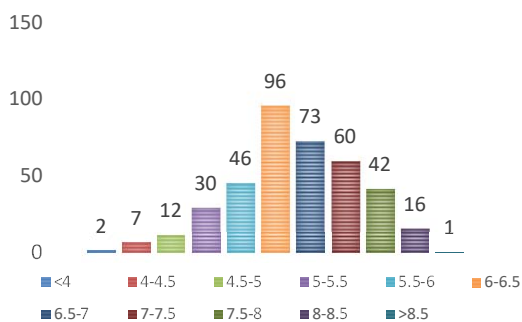

Fig. 2 (b) Frequency Distribution of ROI

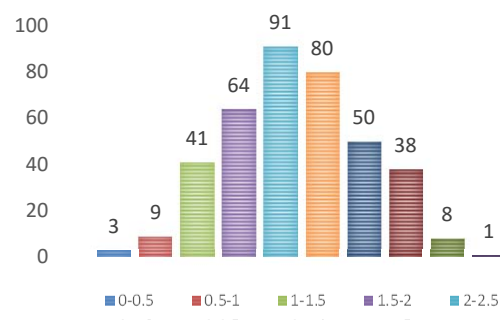Fig. 2 (c) Frequency Distribution of Movie Rating

Fig. 2 (d) Frequency Distribution of Log (NOV)

Table II shows the correlation data among the variables. Two variables are said to be linearly correlated if the value is either -1 or +1 [13]. As the value approaches closer to zero, the variables are said to be less correlated with each other. For regression, it is necessary to have little/no correlation between the dependent and independent variable [14]. We can see a little correlation between the variables as the value range from -0.25 to +0.23 in Table II. The control variables i.e. Number of Votes (NOV) and Rating (RAT) have a modest level of correlation (0.59). This value is not essential as the correlation is between the control variables and not between the independent variables [14]. To cross-verify the correlation, Variance Inflation Test (VIF) is also performed for the independent variables [12]. The result shows, the VIF value below 5 for all the variables, hence we can conclude there is no correlation between the variables [12].

TABLE II
CORRELATION STATISTICS BETWEEN THE VARIABLES

	ROI	GER	CON	LAN	RAT	BTS	NOV
ROI	1						
GER	0.12	1					
CON	-0.18	-0.21	1				
LAN	-0.14	-0.25	0.12	1			
RAT	0.15	-0.17	0.07	0.09	1		
BTS	0.23	0.22	-0.04	-0.1	-0.1	1	
NOV	0.16	-0.24	0.09	0.15	0.6	-0.1	1

C. Multiple Linear Regression

The regression result is shown in Table III and indicates that the presence of women in a movie cast has a significant impact on the ROI of the movie (t = 3.63, p < 0.05). The coefficient of gender ratio GER (0.495) indicates a positive effect. So our hypothesis, H1 namely: The presence of women in a movie cast is positively related to movie the ROI stands true.

Interestingly, the dummy variable Bechdel Test Score (BTS) also has a significant p-value (0.04) and is positively related to the ROI (0.27). The results can be corroborated by simple observation of the raw data. Fig. 3 shows the films having high female characters (Bechdel Score 3) also have a high ROI of 2.04 for every dollar spent. While the average budget spent for films that do not pass the Bechdel Test is way higher than those that pass the test.

TABLE III
RESULT OF MULTIPLE LINEAR REGRESSION ANALYSIS

Variable	Coefficients	Standard Error	t Stat	P-value
Intercept	0	#N/A	#N/A	#N/A
GER	0.495	0.136	3.630	0.000
CON	-0.976	0.057	-1.932	0.054
LAN	-0.758	0.401	-1.889	0.060
RAT	-0.219	0.154	-1.423	0.156
BTS	0.270	0.157	1.721	0.046
NOV	1.311	0.421	3.115	0.002
Multiple R	0.542			
R Square	0.294			
Adjusted R Square	0.282			
Standard Error	2.925			
Observations	385			
F Statistic	26.304			

The variable Countries in which a movie is released (CON), Rating of the movie (RAT) and Languages in which a movie is released (LAN) do not have a significant p-value (p >0.05) considering 95% confidence interval. Hence, it indicates that the ROI of the movie does not depend on these factors. The Number of Votes (NOV) has a significant (p<0.05) and a positive impact (1.311) on the ROI of the movie. The movie becomes more popular as people start giving votes – the act of assigning a star rating [15]. Word of mouth or spread has a significant impact on the ROI of the movie [16]. The Adjusted R Square value for the model is 28.21%, which means only 28.21% of the variance in the dependent variable (ROI), can be explained by variances in the independent variables [13].

VI. CONCLUSION

A lot has been talked and discussed about having better parity between the sexes in society. News flashing the appointment of a female contender as the next CEO of a company or the next President of a country gathers accolades from all. These events, although very rare, do become a source of inspiration and women have started receiving their share of the pie, but is this really the truth?

We tried to investigate the current status of women in entertainment industry by understanding the male to female ratio across 385 Hollywood movies. The question addressed by the paper is "Whether the presence of more female characters in the movie has a direct impact on ROI". The results depict that the presence of more women in the cast, thus increasing the gender diversity, has a positive impact on the movie's performance. Therefore, the scriptwriters and directors should take this fact into consideration and involve more women's roles or scenes in their story.

The research has some limitations though. Gender diversity is measured by only considering the number of women present in the entire cast. This Gender Diversity does not give a clear picture of the duration of the role. There might be cases when the cast has too many women but still, the women do not have a significant role in the movie. Although Bechdel Test Score (BTS), which lays importance of at least one dialogue between women in the movie, is taken into consideration by the researcher. Still, there is a scope where we can consider the duration of the role which might give a different picture. The researchers are planning to consider this aspect for their future work. Also, we aim to carry this study for movies across several film industry and not restrict to the US film industry.
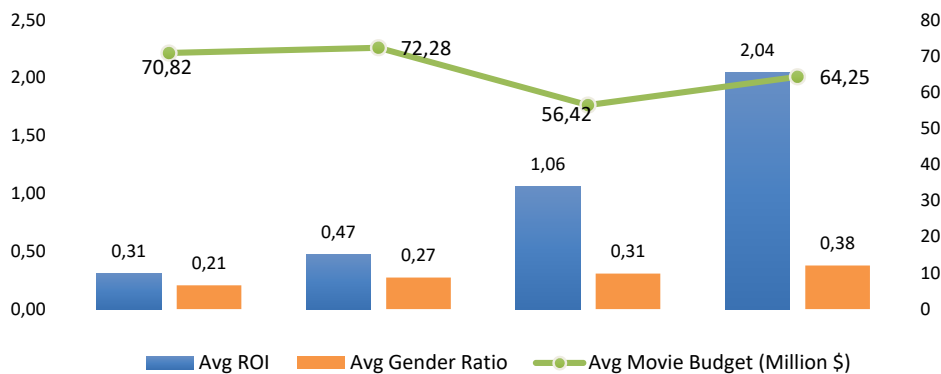

Fig. 3 Average ROI, Gender Ratio and Budget spent across films for different categories as per the Bechdel score

REFERENCES

[1] Statista – The Statistics Portal, Statistics and Studies from more than 22,500 Sources, Article: Film and Movie Industry – Statistics and Facts, https://www.statista.com/topics/964/film/.
[2] Dr Smith, L.S, Choueiti, M, Dr Pieper, K. (2017). Inequality in 900 Popular Films: Examining Portrayals of Gender, Race/Ethnicity, LGBT, and Disability from 2007-2016.
[3] Forbes Article – World's Highest-Paid Actors and Actresses 2017.
[4] Protasovs, I. (2015). Board Diversity and Firm's Financial Performance: Evidence from South-East Asia, 5th IBA Bachelor Thesis Conference, Netherlands.
[5] Lee-Kuen, Y.I, Sok-Gee, C, Zainudin, R. (2017). Gender Diversity and Firm's Financial Performance in Malaysia, Asian Academy of Management Journal of Accounting and Finance, Vol.13, No.1, 41-62.
[6] Sanan, K.N. (2016). Board Gender Diversity, Financial and Social Performance of Indian Firms, The Journal of Business Perspective, Vol 20, Issue 4, pp. 361-367.
[7] Internet Movie Database for Movies, TV and Celebrities - Information courtesy of IMDb – www.imdb.com.
[8] Bechdel Score and Test - www.bechdeltest.com.
[9] Trochim, W. M., & Donnelly, J. P. (2006). The research methods knowledge base (3rd ed.). Cincinnati, OH: Atomic Dog.
[10] Gravetter, F., & Wallnau, L. (2014). Essentials of statistics for the

behavioural sciences (8th ed.). Belmont, CA: Wadsworth.

[11] Field, A. (2000). Discovering statistics using SPSS for windows. London-Thousand Oaks-New Delhi: Sage publications.

[12] Field, A. (2009). Discovering statistics using SPSS. London: SAGE.

[13] Studenmund, A.H. (2014). Using Econometrics, A Practical Guide. Sixth Edition, Harlow:Person Education Limited.

[14] Bewick, V, Cheek, L and Ball, J. (2003). Statistics review 7: Correlation and regression. Crit Care 7(6):451-459.

[15] Ramos, M, Calvao A M, Anteneodo C. (2015). Statistical Pattern in Movie Rating Behavior: Plos One Journal.

[16] Ullah R, Zeb A, Kim W. (2015). The impact of Emotions on the Helpfulness of Movie Reviews: Journal of Applied Research and Technology.