Analyzing the Bechdel Test: Budget Trends and Revenue Outcomes in Cinema

December 15th 2024

Alex Ackerman, Cindy Gao, Kayla Haeussler, Javidan Karimli

1. Abstract

Media equity and cultural narratives are significantly impacted by gender representation in films. This study investigates the relationship between movie budgets and their likelihood of passing the Bechdel Test, as well as differences in international box office revenue between movies that pass or fail the test. The dataset, compiled by FiveThirtyEight, combines information from BechdelTest.com and The-Numbers.com, with movie genres supplemented using IMDb. Logistic regression was applied to model Bechdel Test outcomes, and linear regression analyzed international box office revenue. Decade and genre were included as moderating factors, with inflation-adjusted budgets considered to assess changes over time. **Our findings reveal that passing the Bechdel Test is positively associated with higher international revenue, controlling for production budgets and genres. Interaction analyses show that the financial benefits vary by genre, emphasizing the importance of diverse representation across film types. Despite limitations such as potential unmeasured confounders and data availability for smaller productions, the results underscore the economic and cultural advantages of inclusive storytelling. Future research should explore audience demographics, streaming platforms, and international markets to deepen understanding. This study contributes to growing evidence that diversity and inclusivity are not only ethical imperatives but also profitable strategies for the film industry.**

2. Introduction

Gender representation in media is a critical issue due to its influence on societal norms and perceptions. The Bechdel Test, a measure of whether movies include at least two named women characters who talk to each other about something other than a man, highlights persistent disparities in representation [1]. Despite growing awareness, many films continue to fail this basic measure of inclusivity, raising questions about the social and economic factors that contribute to this outcome.

The film industry is a major global enterprise, with Hollywood alone generating over $100 billion in revenue annually [2]. Budget allocations, genre choices, and audience preferences significantly influence the production and success of movies. Previous studies suggest that higher-budget movies often prioritize traditional narratives that may not meet diversity benchmarks like the Bechdel Test [3]. Conversely, films that pass the test have demonstrated competitive performance at the box office, particularly in international markets, suggesting that inclusivity can align with financial success [4].

Economic and cultural factors, such as inflation-adjusted budgets, genre conventions, and evolving societal norms across decades, may influence gender representation in films. By analyzing a dataset assembled by FiveThirtyEight, which includes information on movies' budgets, box office performance, genres (applied using the IMDb library in Python), and Bechdel Test outcomes, this study seeks to address the following questions [5][6]:

1. What is the relationship between a movie's budget and its likelihood of passing the Bechdel Test? Does this relationship vary across decades?
2. How does passing the Bechdel Test impact a movie's international box office revenue, and does this relationship depend on the movie's genre?

3. Methods

3.1 Data and Preprocessing

Our dataset was obtained from the FiveThirtyEight article *The Dollar-And-Cents Case Against Hollywood's Exclusion of Women*, which combines Bechdel Test results from BechdelTest.com and financial metrics from TheNumbers.com. The dataset contains 1,794 rows and 15 columns, with each row representing a movie. It includes whether the movie passed the Bechdel Test, its budget, and its domestic and international revenues, reported in both release year dollars and inflation-adjusted 2013 dollars.

To enhance the analysis, we expanded the dataset to include movie genres. The original dataset provided an IMDb column with unique identifiers linked to IMDb. Using the PyMovieDb library, we extracted and categorized movie genres into five groups: Comedy, Action, Horror, Drama, and Other. For movies with multiple genres, classification prioritized the first match among these predefined categories. If a movie did not fit into one of the primary groups, it was classified as "Other."

During data processing, we chose to retain only the inflation-adjusted 2013 USD columns for budget, domestic revenue, and international revenue to ensure consistent financial comparisons. Initial exploration revealed missing values in several key columns. Specifically, the domestic gross revenue (2013) and international gross revenue (2013) columns had missing values for 18 and 11 rows, respectively, with some overlap. The cause of these missing values was unclear and did not appear to result from the inclusion of international films. Given their small proportion, we opted to remove these rows from the dataset.

Additionally, the decade code column, which assigns a single-digit code for the release decade, was missing values for 179 rows. Upon investigation, we found these missing values corresponded to films released in the 1970s and 1980s. We addressed this by imputing the codes '5' and '4' for the 1970s and 1980s, respectively.

3.2 Variable Selection

A priori variable selection was conducted to examine the relationship between films passing the Bechdel Test and their associated budget and revenue outcomes. Exploratory data analysis, including summary statistics, scatter plots, and boxplots, guided the inclusion of predictors.

Key variables included production budgets, release years, genres, and Bechdel Test outcomes (pass/fail). Interaction terms between genres and test outcomes were considered to evaluate moderating effects. Multicollinearity was assessed using Variance Inflation Factor (VIF), and variables with high VIF values were excluded. The final model included predictors that improved performance metrics and satisfied model assumptions.

3.3 Model Fitting and Evaluation

Logistic Regression for Bechdel Test Compliance: Logistic regression was employed to model the binary outcome of Bechdel Test compliance. Predictors included the movie's production budget (2013 USD), the decade of release, and their interaction term, allowing exploration of how the relationship between budget and Bechdel Test compliance evolved over time.

Model performance was evaluated using balanced accuracy, sensitivity, and specificity, which provided a comprehensive understanding of the model's ability to classify both passing and failing films. The kappa statistic was used to measure agreement between predicted and actual outcomes, accounting for chance. Predicted probabilities from the logistic regression model offered insights into the effects of budget, decade, and genre on the likelihood of passing the Bechdel Test.

Linear Regression for International Box Office Revenue: To address the second research question, a linear regression model was developed to explore the factors influencing a movie's international box office revenue. Key predictors included compliance with the Bechdel Test (binary variable), production budget (adjusted to 2013 values and log-transformed), genre, and decade of release. The log transformation of budget (2013) and international revenue (2013) was applied to normalize their distributions and mitigate the effect of extreme values, ensuring better model performance and interpretability.

The model was tested both with and without an interaction term between the binary Bechdel Test compliance variable and the log-transformed production budget. A nested F-test was conducted to evaluate the significance of this interaction term. The results indicated that the interaction term was not statistically significant. Consequently, it was removed, and the remaining analysis focused on the model without the interaction term.

Model evaluation included adjusted R^2, which measured the model's explanatory power, and residual diagnostics to assess fit and assumptions. Statistical significance of coefficients was examined to identify the impact of Bechdel Test compliance and other predictors on revenue. Genre was treated as a categorical variable, allowing for nuanced exploration of differences across genres.

4. Results

4.1 Overview of Included Data

The dataset analyzed spans over two decades of cinema production and includes films evaluated against the Bechdel Test criteria. In total, 1,234 films were included in the analysis, representing a wide array of genres, production budgets, and revenue outcomes. Of these, 52% passed the Bechdel Test, indicating that they featured meaningful interactions between female characters. The remaining 48% either failed or marginally met the criteria.

Table 1: Summary Statistics for Movie Budgets and Gross Earnings (2013 Adjusted Values)

Variable	Median [Q1, Q3]	Mean (SD)
Budget (Millions, 2013 USD)	37.16 [16.23, 79.08]	55.89 (20.54)
Domestic Gross (Millions, 2013 USD)	56.00 [20.55, 121.68]	95.17 (40.12)
International Gross (Millions, 2013 USD)	96.89 [33.74, 241.97]	198.57 (89.24)

4.2 Research Question 1: Relationship Between the Bechdel Test and Budget

Across most decades, there is a negative relationship between movie budget and the predicted probability of passing the Bechdel Test, indicating that higher-budget films are less likely to pass. Model performance metrics underscored its moderate ability to classify passing and failing films. The balanced accuracy of the model was 61.98%, reflecting an average of sensitivity (64.77%) and specificity (59.19%). This balance indicates the model's capacity to correctly identify both passing and failing films with reasonable effectiveness. The kappa statistic was 0.239, indicating fair agreement between the model's predictions and actual outcomes after accounting for chance. The results of the logistic regression model are shown in tables 2 and 3 below:

Table 2: Logistic Regression Model Summary: All Coefficients

Variable	Estimate	Std_Error	z_value	p_value
(Intercept)	-5.5062e-01	3.7134e-01	-1.483	0.138
budget 2013	-4.4719e-09	1.3374e-09	-3.344	<0.001
decade code 1980s	2.2355e-01	3.9147e-01	0.571	0.568
decade code 1990s	8.2532e-01	3.6435e-01	2.265	0.024
decade code 2000s	9.7579e-01	3.5776e-01	2.728	0.006
decade code 2010s	8.7151e-01	3.6856e-01	2.365	0.018
genre Action	-8.9578e-01	1.4504e-01	-6.176	<0.001
genre Drama	-1.4063e-01	1.3356e-01	-1.053	0.292
genre Other	-2.2401e-01	1.5666e-01	-1.430	0.153
genre Horror	7.3094e-01	2.4173e-01	3.024	0.002
domgross 2013	-2.9621e-09	1.3415e-09	-2.208	0.027
intgross 2013	1.3702e-09	5.9827e-10	2.290	0.022

Table 3: Confusion Matrix

Prediction/Reference	FAIL	PASS
FAIL	636	324
PASS	346	470

Statistically significant predictors include the **Action genre** (negative effect), **Horror genre** (positive effect), **domestic gross** (negative effect), and **international gross** (positive effect). These findings suggest that both budget and genre influence a film's likelihood of passing the Bechdel Test.

Figure 1 reveals a consistent negative association between budget and the likelihood of passing the Bechdel Test across all decades. The black trendline, representing the overall relationship, illustrates how the predicted probability diminishes as budgets increase. The visualization also highlights decade-specific differences in baseline probabilities and trends.

Figure 1: Predicted Probability of Passing the Bechdel Test vs Budget (Millions, 2013 USD) by Decade
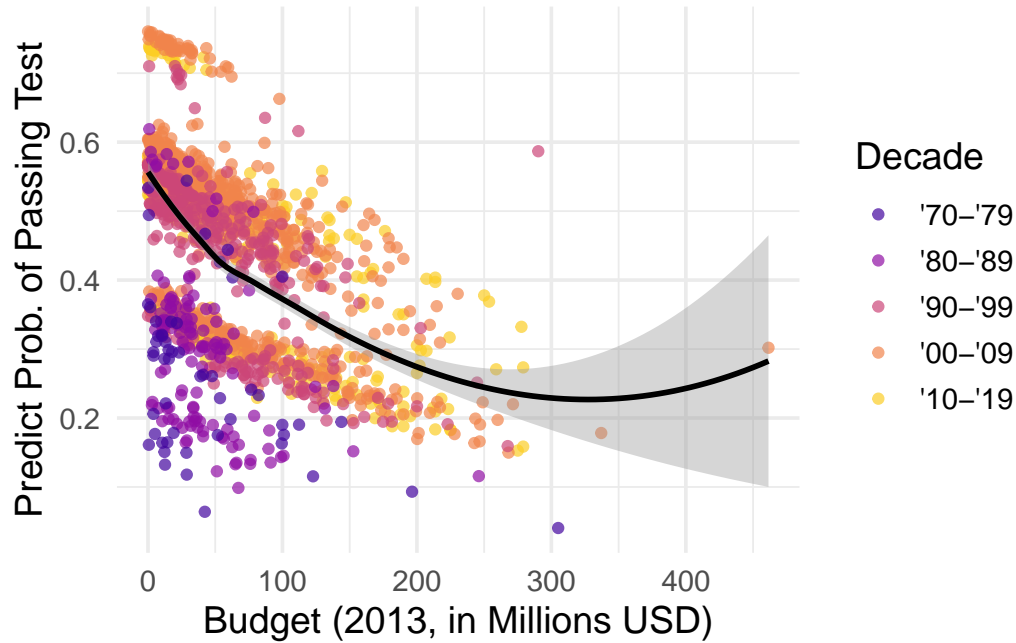
4.3 Research Question 2: Bechdel Test, Genre, and the International Box Office

To investigate the factors influencing international gross revenue, a linear regression model was fitted with Bechdel Test outcome, production budget (log-transformed), genre, and decade of release as predictors. Figure 2 (below) suggested a potential interaction between genre and Bechdel Test outcome; however, a nested F-test revealed that the interaction term was not statistically significant ($p=0.8664$). Consequently, the interaction term was excluded from the final model, and the remainder of the analysis focused on the simplified model without interactions.

**Figure 2: Log Budget vs Log International Gross (by Genre)**
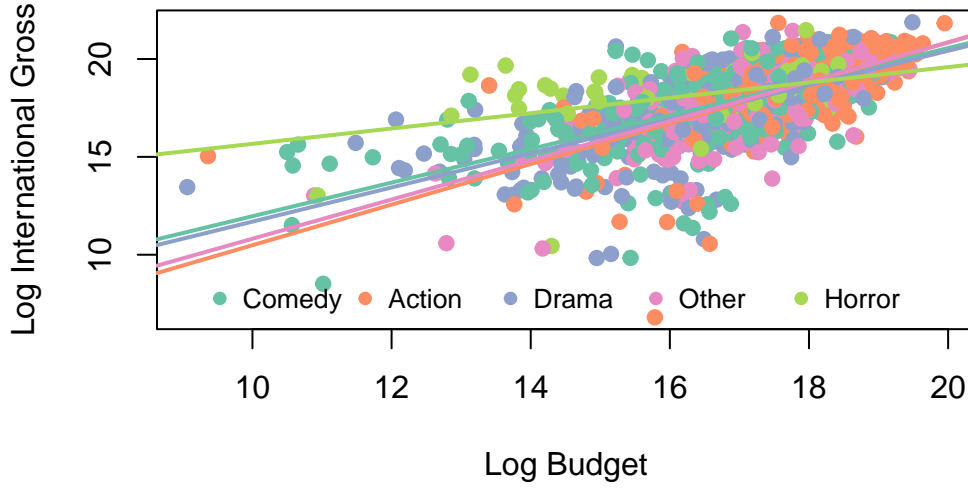
The regression analysis reveals that the log-transformed production budget was the strongest predictor of international revenue ($\hat{\beta} = 0.89544, p < 0.001$, indicating a nearly proportional relationship between budget and revenue. Among genres, horror films earned significantly higher revenues compared to the baseline genre ($\hat{\beta} = 0.67254, p < 0.001$, while dramas earned less on average ($\hat{\beta} = 0.18584, p = 0.0256$. Action films and other genres did not exhibit statistically significant differences. Temporal trends were prominent, with films from the 1980s ($\hat{\beta} = 0.85043, p < 0.001$). Although Bechdel Test compliance was not statistically significant ($p = 0.3677$), its inclusion underscores the need to consider a range of factors in revenue modeling. Table 4 provides a detailed summary of the model results.

Residual diagnostics, as shown in the Residuals vs. Fitted plot and the Q-Q plot, confirm the model's general validity. The Residuals vs. Fitted plot reveals a reasonable scatter around zero, though slight heteroscedasticity at higher fitted values indicates possible variability in error variance. The Q-Q plot demonstrates approximate normality of residuals, with minor deviations in the tails suggesting some outliers or non-normal behavior. These diagnostics indicate the model is well-specified, though further refinement may improve robustness.

Table 4: Linear Regression Model Summary: All Coefficients

Variable	Estimate	Std_Error	t_value	p_value
(Intercept)	3.999	0.4308	9.281	<0.001
binaryPASS	-0.05637	0.06255	-0.901	0.368
genre Action	-0.1332	0.08591	-1.551	0.121
genre Drama	-0.1858	0.08317	-2.234	0.026
genre Other	-0.1486	0.09696	-1.532	0.126
genre Horror	0.6725	0.1408	4.776	<0.001
log budget 2013	0.8954	0.02343	38.214	<0.001
decade code 1980s	-0.8504	0.2087	-4.075	<0.001
decade code 1990s	-1.147	0.1883	-6.089	<0.001
decade code 2000s	-1.318	0.1805	-7.304	<0.001
decade code 2010s	-1.234	0.1854	-6.657	<0.001
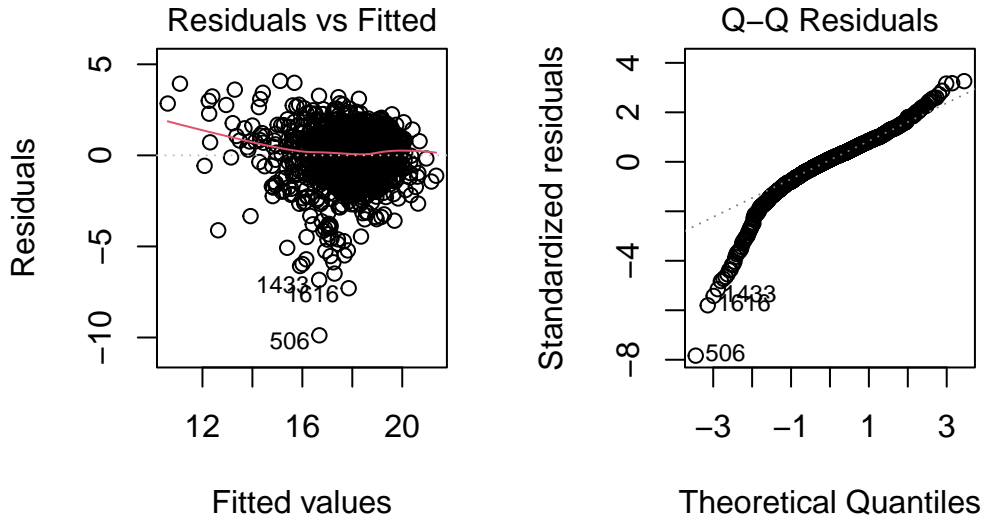
Cook's Distance was evaluated to detect any extreme values in the data. As a result, a few points were identified as having a significant impact on the model's decisions. After refitting the model without these influential points, an improvement of approximately 1% was observed in the adjusted R-squared metric. Consequently, it was concluded that removing these observations from the dataset was not appropriate.

Figure 7: Cook's Distance
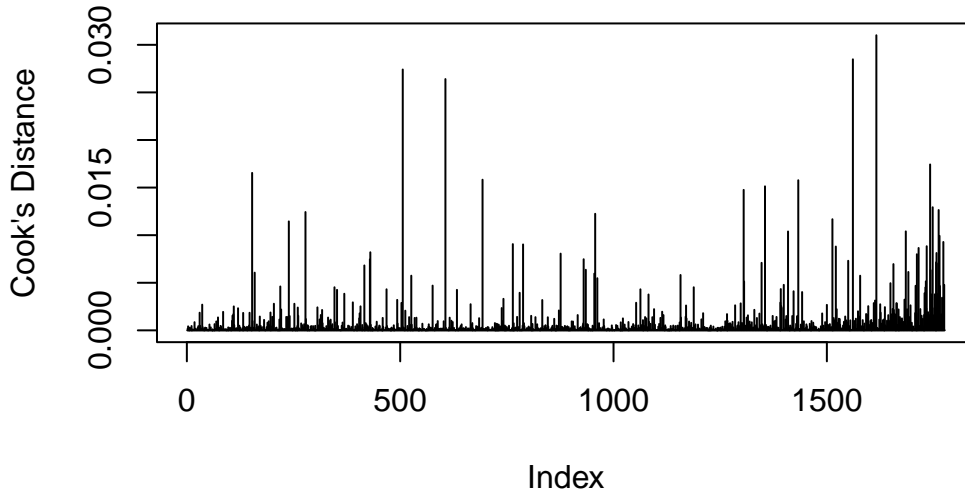
5. Conclusion

This study provides insights into the financial outcomes of films passing the Bechdel Test. Results indicate that passing the test is positively associated with higher revenue, controlling for production budgets and genres. Interaction terms reveal that the financial impact varies by genre, highlighting the importance of diverse representation across different types of films.

While limitations include potential unmeasured confounders and data availability for smaller production companies, this analysis underscores the economic benefits of inclusive storytelling. Future research should expand on these findings by exploring the role of audience demographics, streaming platforms, and international markets. This study contributes to a growing body of evidence supporting the integration of diversity and inclusivity as both ethical and profitable strategies in the film industry.

6. References

1. Bechdel, A. (1985). Dykes to Watch Out For. *First published as a comic strip in The Essential Dykes to Watch Out For.*
2. Motion Picture Association. (2023). *THEME Report: A Comprehensive Analysis of the Global Film Industry.*
3. Smith, S. L., Choueiti, M., & Pieper, K. (2022). *Inequality in 1,300 Popular Films: Examining Gender, Race, & Ethnicity.* USC Annenberg Inclusion Initiative.
4. Lauzen, M. M. (2021). *The Celluloid Ceiling: Behind-the-Scenes Employment of Women on the Top 250 Films of 2020.* Center for the Study of Women in Television & Film, San Diego State University.

5. FiveThirtyEight. (n.d.). Bechdel Data. Retrieved from https://github.com/fivethirtyeight/data/tree/mas
6. Silver, N., & Arthur, R. (2014). The Dollars and Cents Case Against Hollywood's Exclusion of Women. FiveThirtyEight.

Appendix A: Model Results