Essays on the Economics of the Motion Picture Industry

Liang Xu

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2022

Reading Committee:

Yuya Takahashi, Chair

Patrick Bajari

Dong-Jae Eun

Jing Tao

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Essays on the Economics of the Motion Picture Industry

Liang Xu

Chair of the Supervisory Committee:
Yuya Takahashi
Department of Economics

My dissertation consists of three chapters that address important questions in the U.S. motion picture industry. In Chapter 1, I limit the attention to the theatrical market and use a linear model with fixed effects to examine the impact of online word of mouth (WOM) on box office revenue. The online word-of-mouth dataset is extracted from Internet Movie Database (IMDB), one of the most popular and informative movie websites in the U.S. Chapter 1 contributes to the literature by using basic natural language processing (NLP) techniques to construct text features including review polarity, subjectivity, readability, and similarity to the product description in a manner which has barely been discussed before. The evidence shows that 1) volume and valence of word-of-mouth communication in the last period are positively associated with the current movie sales; 2) reviews with extreme rating scores, no matter whether they are positive or negative, can attract attention and increase the movie demand; 3) the audience does read the reviews instead of relying only on summary statistics and subjective reviews with rich content, even at the cost of low readability, can potentially boost box office revenue; 4) the disadvantage of the insufficient marketing budgets for limited releases can be rectified through the impact of proper online word-of-mouth communication.

Chapter 2 expands the scope to include both theatrical and home video markets in the study. The home video market generates more gross revenue than the theatrical market but has received surprisingly less attention from economic scholars. Using market-level data, I

first conduct demand estimation separately for the two markets using logit and one-level nested logit models. This step allows me to quantify movie qualities, consumers' utility decay rates, seasonality in demand, and market expansion effect in order to understand how these two markets operate differently. Next, I pool all the sample movies from the two markets together and employ a two-level nested logit model to quantify consumers' substitution patterns between the two different viewership platforms. The results validate that the two-level nest structure is consistent with the maximization of a random utility function (McFadden 1978) and consumers do distinguish between watching movies in theaters and on home video.

Chapter 3 builds upon the demand estimation results from Chapter 2 and studies the optimal time to release a movie under the sequential distribution setup. Movie distributors use what is known as the windowing strategy — releasing a film in different venues over discrete periods of exclusivity to maximize consumption over the lifetime of a property. The trend toward shorter time lags in theatrical releases has caused controversy in the U.S. motion picture industry, necessitating the demystification of how distributors optimize release schedules for sequential distribution. I model distributors' windowing strategy as a one-shot sequential-move game with incomplete information. I follow the method of pseudo-backward induction proposed by Einav (2009) and further estimate the weekly cost of distribution and distributors' weighting coefficient on the two windows while taking the demand estimates as given. Eventually, by conducting counterfactual analyses in a two-player game, I find that higher cost of theatrical distribution, lower weights on the theatrical window, poorer movie quality with weak opening performance, or faster theatrical utility decay can all potentially rationalize the practice of shrinking theatrical window length.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to the chair of my supervisory committee, Professor Yuya Takahashi, for his support and insightful guidance throughout my Ph.D. journey. His great patience and encouragement inspired me and led me into the field of Industrial Organization. His valuable comments and resourcefulness have helped me shape and refine this dissertation. Besides, he introduced me to the data source and offered me the financial support in acquiring the data. I am also extremely thankful to my supervisory committee member, Professor Patrick Bajari, for inspiring me with new ideas and introducing me to plenty of research resources. His broad knowledge and network always led me to new perspectives in every one of our conversations. I am also truly thankful to my supervisory committee members, Professor Dong-Jae Eun, Professor Jing Tao, and Professor Stephen Groening. I thank Professor Dong-Jae Eun for having taught me a fantastic course in Industrial Organization which greatly expanded my vision and toolkit for my research. I also thank him for his constructive advice for my research which helped me be very clear about the research goal at the initial stage. I thank Professor Jing Tao for having taught me the important Econometrics class and for always being so kind and supportive during my studies at the University of Washington. I thank Professor Stephen Groening for being my GSR and having so many insightful and fruitful discussions with me on film-related topics. He always kindly wrote me long emails to answer my research questions, helped me ask around for data and research sources, and recommended me related articles and courses to study. Most importantly, he helped me understand how the movie industry works, which is crucial to my dissertation. The completion of the dissertation would not have been possible without their involvement.

I would like to extend my gratitude to many professors who once taught me at the University of Washington. These names include but are not limited to, Professor Jacques Lawarrée, Professor Fahad Khalil, Professor Alan Griffith, Professor Rachel Heath, Professor Xu Tan, Professor Yanqin Fan, Professor Quan Wen, Professor Eric Zivot, Professor Thomas Richardson, Professor Yu-chin Chen, and Professor Stephen Turnovsky. I would also like to send my sincere thanks to my peers, Zhe, Yue, Zhou, Sajid, Jackson, Chi-Hui, Minyan, Kwong-Yu, Kathy, Zihao, and Castiel, for sharing ideas and improving each other's research. Moreover, special thanks go to the staff at our economics department, Simon Reeve-Parker, Heidi Hannah, and Kim Lee.

Finally, I am particularly thankful for my parents. It is their unconditional love that allows me to explore my life freely and be the person I want myself to be.

DEDICATION

To my dear parents,

for their love,

and trust.

Chapter 1

WORD OF MOUTH AND MOTION PICTURES

1.1 Introduction

The media and entertainment industry is heavily comment-driven. Entertainment goods such as movies, television shows, music, games, and prints are often consumed collectively and feature in people's daily conversations. The experiential nature of these products makes word-of-mouth (WOM) communication typically critical to the formation of demand. Meanwhile, it is also recognized that WOM represents a challenging domain for researchers in a way that it is not an easy job to capture and measure WOM factors. Using the motion picture industry as an example, a series of papers, De Vany and Walls (1999 [23], 2004 [24]), De Vany and Lee (2001 [20]), and De Vany (2003 [19]), have showed with statistical evidence that the cascade of information (WOM) among the audience evolves so many paths that it is impossible to attribute the success of a movie to individual causal factors, which supports William Goldman's famous comment on the movie business, "Nobody knows anything[1]." Godes and Mayzlin (2004 [40]) have also pointed out that one of the major difficulties of working with WOM-related subjects is gathering the data and constructing meaningful features. Fortunately, the emergence of internet-based WOM communities where consumers share their opinions, rate products, and offer additional information about product quality helps alleviate the problem. Unlike "offline" WOM that "disappears into the thin air", online WOM provides persistent and easily accessible public records of everything that has been posted (Dellarocas, Awad, and Zhang (2007 [26])), which creates an excellent condition for research.

---

[1] It refers to the extreme uncertainties in the industry and executives' inability to foresee how the market would perform.

Volume and valence are among the most studied WOM attributes. Volume measures the total amount of WOM interactions and is often constructed by counting the number of reviews posted. Valence refers to the evaluative direction of the review, can be positive, neutral, or negative (Purnawirawan, De Pelsmacker, and Dens (2012 [71])) and is commonly measured by user-generated rating scores. Volume is often used to indicate awareness effect, that more WOM interactions may increase consumers' product awareness through dispersion, while valence is used for testing persuasive effect, that consumers can infer the quality of the goods and may change their purchase decisions accordingly. Here I list several well-cited studies that examine the effect of volume and valence of WOM on product sales. Chen, Wu, and Yoon (2004 [15]) studied the book market with cross-sectional data extracted from Amazon.com. They used a linear model with the sales rank of books as the dependent variable and showed that the number of reviews a book has received is positively related to its sales but the consumers' ratings of the book are not significantly related. Liu (2006 [57]) adopted a similar linear model as the one from Chen, Wu, and Yoon (2004 [15]) but applied it to the movie market using panel data from Yahoo Movies. They took the weekly box office revenue as the dependent variable, and lagged volume and valence of WOM messages as the key independent variables. Their results showed, again, that the volume but not the valence of WOM offers significant explanatory power for the box office revenue. Duan, Gu, and Whinston (2005 [28]) developed a two-equation system where one is with the daily box office revenue as the dependent variable and the other is with the daily number of posts (the volume of WOM). They estimated the two linear equations simultaneously with three-stage least-square (3SLS) and reached a similar conclusion that the volume but not the valence matters. Rather than modeling with linear regressions, Dellarocas, Awad, and Zhang (2007 [26]) tailored the classic Bass diffusion model to better fit the movie industry and found that 1) the early volume of online reviews can be used as a proxy of early movie sales, and 2) the average valence of user reviews is a significant explanatory variable for a movie's coefficient of internal influence (q in Bass model).

Besides volume and valence, several other WOM features have also once been brought in. For instance, the number of recommendations a product receives (Chen, Wu, and Yoon (2004 [15])), the dispersion of conversation which measures the extent to which product-related conversations have taken place across communities (Godes and Mayzlin (2004 [40])), the length of text reviews (Chevalier and Mayzlin (2006 [17])) and so on. In addition to the challenges of collecting data and building features, another difficulty comes from the endogeneity problem when modeling with WOM factors. Two sources of endogeneity are often mentioned. One is the omitted variable problems caused by unobserved heterogeneity, since products of which sales rely on the WOM effect are usually experience goods, and it is hard to include sufficient characteristics to capture their qualities. This is usually solved by introducing individual fixed effects, but data across time periods (panel data) would be required for implementation. The other is the simultaneity, that is, the causality between product sales and WOM works in both directions. Some previous studies only work on one direction, treating WOM as a driver of buyer behavior (Chen, Wu, and Yoon (2004 [15]), Liu (2006 [57]), and Dellarocas, Awad, and Zhang (2007 [26])), and others tackled the simultaneity problem by estimating a two-equation system (Duan, Gu, and Whinston (2005 [28])).

In this study, I use a linear model with fixed effects to examine how online WOM may affect movies' box office revenue. My work differs from and contributes to the current literature in the way that apart from the volume and valence of WOM, various features constructed from WOM text are investigated in detail. Basic natural language processing (NLP) techniques are used to help detect certain aspects of the text, including polarity, subjectivity, readability, and similarity to product descriptions. The motivation here is that consumers do read the text reviews. How the nuances in words affect consumers' purchase decisions is barely studied in the previous economic literature and the rich text review data I extracted can potentially help to fill the gap. The remaining paper is organized as follows. Section 1.2 describes the revenue data used in this paper and discusses the relationship between rank, revenue, and length of theatrical runs. Section 1.3 introduces the Bass model which has a built-in coefficient, q, measuring how the WOM communication shapes the revenue streams.

The results of fitting the Bass model with the revenue data are also presented in Section 1.3. Section 1.4 describes how the online WOM data are extracted from `IMDB.com`, offers descriptive statistics, and constructs text features for later modeling. Section 1.5 presents the linear fixed effects model along with its estimation results and interpretation. Section 1.6 concludes.

1.2 Weekly Box Office Revenue

The revenue data used in this paper are the OpusData Extracts from Nash Information Services[2]. The OpusData records the following information every week: the domestic box office revenue, the number of tickets sold, the number of theaters showing the movie[3], and the rank (by revenue) of the movie. The sample covers 11,613 titles released in theaters from 1997 to 2019 (Figure 1.1). This section is devoted to better understanding movies' theatrical life cycles. I first show examples of motion picture runs and revenue streams. Next, I explore the relationship between movies' rank and box office revenue. It turns out that revenue generated in the U.S. motion picture industry is concentrated on a few big hit movies that belong to a few major studios. Finally, I complement the section with an analysis of movies' rank and length of theatrical runs. It is observed that higher-ranked movies also have longer runs and longer-lived movies tend to reach and sustain high ranks.

1.2.1 Examples of Motion Picture Runs and Sales

To give some ideas about movies' theatrical runs in reality, I show in Figure 1.2 the weekly number of tickets sold in millions (orange curve, left axis), the weekly number of theaters showing the movie (blue curve, right axis), and the weekly box office revenue rankings of the movie (the labeled numbers) for four movies that were released in 2017: *Spider-Man:*

[2]OpusData Extracts from Nash Information Services, LLC (NIS): `https://www.opusdata.com/documentation/index.php/Database_Extracts`

[3]The number of screens showing the movie is not reported, and it is quite hard to know because theaters will chop and change which films are playing on which screens over the course of a weekend.

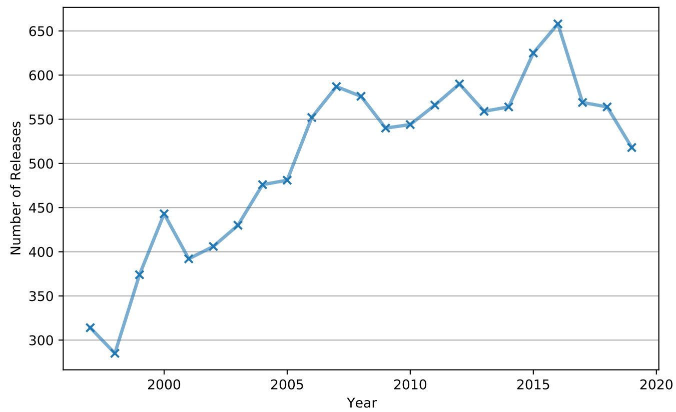


Figure 1.1: Number of Theatrical Releases in Each Sample Year

Homecoming, *Cars 3*, *The Big Sick*, and *Gifted*. The first two shown in the top two panels opened strongly, achieved rank 1 in their opening weeks, had relatively longer runs, and reached higher total revenue. In contrast, the other two opened moderately at low ranks and then began to climb in both rank and revenue.

Over the lifetime of a movie, it may or may not gain acceptance at its initial debut, but its rank and revenue must eventually decline because 1) the movie becomes vulnerable to newly released pictures as time passes, and 2) most people watch a movie only once and the movie's market potential shrinks over time. In addition, the shape of the weekly revenue in Figure 1.2 corresponds to the weekly supply pattern suggesting that theaters' and consumers' adoption are interdependent on each other. In practice, after the movie is released, its box office revenue is weekly reported and the theatrical supply is adapted accordingly.

1.2.2 Rank and Revenue

The four examples shown in Figure 1.2 can be considered as relatively successful runs since most of the movies follow none of these patterns. Figure 1.3 follows the idea of the Lorenz curve, showing that the box office revenue is highly unevenly distributed among all the movies; the top 20% of the market movies account for over 90% of the box office revenue. In

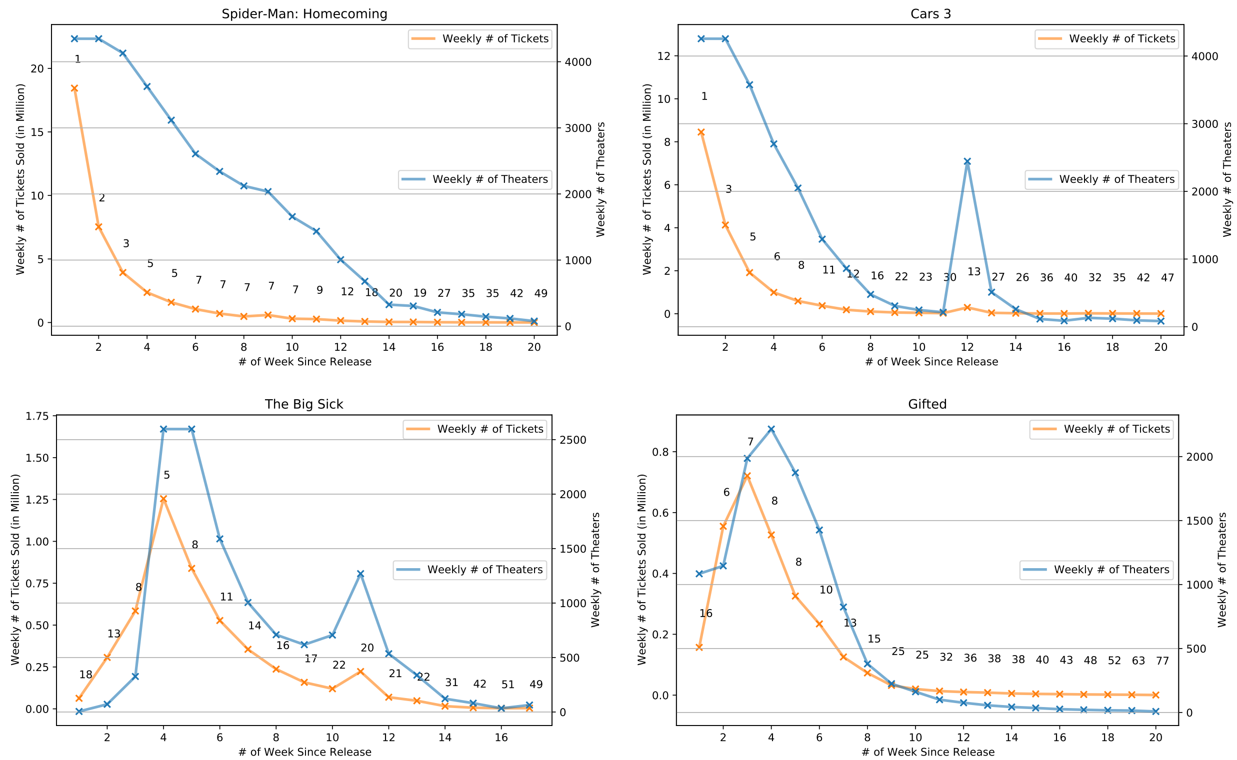


Figure 1.2: Examples of Theatrical Runs (labeled numbers are the weekly box office ranking.)

fact, most movies never have the chance to join the top tier and drop out of the market too quickly. Moreover, 84% of the 776 titles that have at least achieved the first rank once are distributed by the Major Six[4]. Table 1.1 reflects a similar pattern. Only 24% of the sample titles are distributed by the Major Six while they account for 78% of the total revenue aggregated across the sample years. This suggests the oligopoly market structure of the U.S. motion picture industry that revenues are concentrated on a few films and these hit movies are concentrated among a few major distributors[5].

---

[4]There are 640 distributors in the sample.

[5]Such pattern is consistent with the one shown in Table 3.1 using the online data from `the-numbers.com`, which indicates that our sample is representative of the real world.

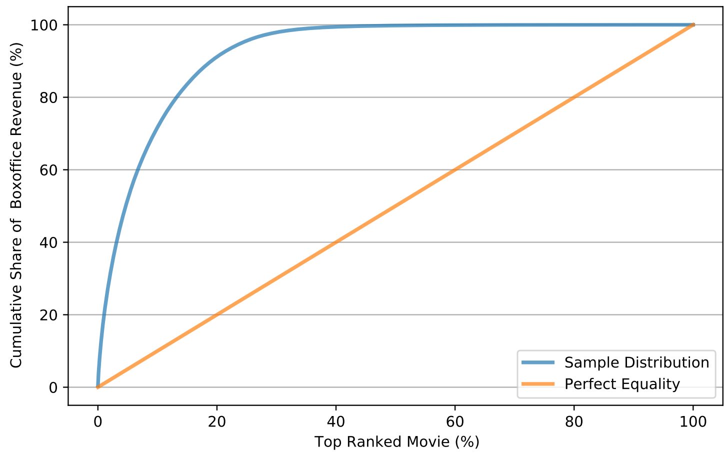


Figure 1.3: Rank-Revenue Distribution

Table 1.1: Box Office Revenue Distribution by Distributors (Year 1997 - 2019)

Distributors	% of Total Releases	% of Total Revenue
Warner Bros.	4.86 %	14.87%
Walt Disney	3.45 %	17.25%
20th Century Fox	3.68 %	11.53%
Sony Pictures	5.10 %	12.22%
Paramount Pictures	3.14 %	10.41%
Universal	3.66 %	11.76%
Other Studios	76.10 %	21.97%

1.2.3 Rank and Length of Theatrical Run

I track each film from its birth to death: birth is defined as the inception of a run and death is defined as being dropped from the theatrical market (De Vany and Walls (1997 [22])).

The variable "week in release" is the number of weeks that have passed since the movie's initial debut, representing the age of the movie. Table 1.2 shows that, on average, a film's rank at birth is 43 and its rank at death is 77. The decreasing number of observations in the second column suggests that movies are dropped out as they grow older and only about 50% of the sample titles can survive till the 10th week from their initial releases. Pooling all the sample titles together, the mean length of the theatrical runs is 12 weeks and the median is 10 weeks. Movies that reached the top rank at least once can run, on average, 17 weeks. To summarize, higher-ranked movies appear to also have longer runs (Figure A.4), and longer-lived movies tend to reach and sustain the high ranks (Figure A.3).

Table 1.2: Weeks in Release by Rank Position (Year 1997 - 2019)

		Rank Position				
Weeks in Release	Obs	Min	Max	Median	Mean	Std
1	10834	1	141	44	43.35	32.50
2	9682	1	139	40	42.23	33.08
3	8729	1	144	36	42.03	33.22
4	8057	1	140	36	43.07	32.82
5	7471	1	150	38	44.86	32.39
6	6950	1	144	41	47.17	31.56
7	6425	1	149	43	49.10	30.54
8	5927	1	144	45	51.10	30.02
9	5459	1	145	47	53.04	29.78
10	4984	1	154	49	54.71	29.57
Week of Death	11613	1	154	78	76.90	27.20

1.3 Measuring Word of Mouth with Bass Diffusion Model

Diffusion models are often constructed to describe and predict the development of a life cycle curve of new products in a social system. A key feature of the model is that it embeds a "contagion process" to characterize the spread of word-of-mouth between those who have adopted the innovation and those who have not yet. The well-known (first-purchase) Bass model (Bass (1969 [4])) brought the concept of diffusion to the marketing field and sparked a considerable amount of research extending the theoretical models and applying them to consumer behavior in various industries. The movie industry holds some features that are consistent with the assumptions[6] of the Bass model. For instance, 1) whether to watch the movie in a theater or not is a binary choice (binary diffusion process); 2) people most likely see each movie at most once in the cinema (no repeat buyers); 3) theatrical releases are new products with short life cycles and declining adoption over time. Therefore, in this section, I estimate the classic Bass model using the weekly theatrical revenue data (without any other factors) to examine how word-of-mouth (the coefficient q) is spread for each sample movie. I start with a brief introduction to the basic Bass model and then discuss the results.

1.3.1 The Bass Model

The Bass model divides the potential adopters of innovation into two groups, the "innovators" and the "imitators". Innovators are influenced only by mass media ("external" force) while imitators are influenced only by word-of-mouth ("internal" force) (Figure A.2). I begin to introduce the Bass model with a hazard function which is used to describe the probability

---

[6]The Bass model comes along with several simplifying assumptions (Mahajan, Muller, and Bass (1990 [60])). 1) The diffusion process is binary; the consumer either adopts or waits to adopt. 2) The maximum potential number of buyers (m) is constant over time. 3) No repeat purchase, or replacement purchase. 4) The impact of word-of-mouth (q) is independent of adoption time. 5) The diffusion process of one product is independent of all others (its substitutes or complements) in the market. 6) The diffusion process is not influenced by marketing strategies. 7) There are no supply restrictions and demand can always be met.

that an adoption will occur at time t given that it has not yet occurred.

$$h(t) = \frac{f(t)}{1 - F(t)} = p + qF(t) = p + \frac{q}{m}Y(t) \tag{1.1}$$

where $F(t)$ is the cumulative fraction of adopters at time t and $f(t)$ is the likelihood of purchase at time t[7]. The fundamental assumption of the Bass diffusion model is that the hazard function is assumed to be a *linear* function of the number of previous buyers. Bass (1969 [4]) further defines m as the total potential market size, $S(t)$ as the sales at time t, and $Y(t)$ as the total number purchasing in the $(0, t)$ time interval. Therefore, we have $S(t) = mf(t)$ and $Y(t) = mF(t)$. We are interested in estimating p, q, and m in Equation 1.1. p is the coefficient of innovation, capturing the intrinsic tendency to adopt as well as the effect of time invariant external influences. q is the coefficient of imitation, capturing the extent to which the probability that one adopts (given that one has not done so yet) increases with the proportion of eventual adopters that has already adopted.

Using Equation 1.1, we have

$$\begin{aligned} S(t) &= mf(t) \\ &= m(1 - F(t))h(t) \\ &= (m - Y(t))(p + \frac{q}{m}Y(t)) \end{aligned} \tag{1.2}$$

Bass (1969 [4]) has shown the solution to $F(t)$ and $f(t)$ are the following:

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}, \quad f(t) = \frac{\frac{(p+q)^2}{p}e^{-(p+q)t}}{\left(1 + \frac{q}{p}e^{-(p+q)t}\right)^2} \tag{1.3}$$

Based on Equations 1.3, $S(t)$ and $Y(t)$ can be further calculated by using $S(t) = mf(t)$ and $Y(t) = mF(t)$.

1.3.2 Estimation

Bass (1969 [4]) proposed to regress sales on cumulative sales and cumulative sales squared using ordinary least squares (OLS). Given that the Bass model contains three parameters,

[7] $F(T) = \int_0^T f(t)dt$

we need at least three time periods for the estimation purpose. To facilitate the empirical analysis, we can expand Equation 1.2 into the following discrete form and identify each term with the context from the movie industry (Dellarocas, Zhang, and Awad (2007 [26])).

$$Y_i(t) - Y_i(t-1) = m_i p_i + (q_i - p_i)Y_i(t-1) - \frac{q_i}{m_i} Y_i^2(t-1) \tag{1.4}$$

where $Y_i(t)$ is the cumulative box office revenue of movie i up until week t of its initial release; $Y_i(t) - Y_i(t-1)$ is the box office revenue of movie i during week t of its initial release; Using the following analogue: $\Delta Y_i(t) = a + bY_i(t-1) + cY_i^2(t-1)$ for $t = 2, 3, \ldots, T$, we can identify the parameters p, q, and m[8].

Schmittlein and Mahajan (1982 [76]) pointed out three shortcomings of using the OLS procedure: 1) sales and sales squared may cause multicollinearity; 2) OLS procedure cannot directly provide standard errors for the estimates; 3) time-interval bias may happen since discrete times series are used to estimate a continuous model. Hence, other estimation procedures are proposed to overcome these shortcomings. Schmittlein and Mahajan (1982 [76]) suggested using maximum likelihood estimation (MLE). Srinivasan and Mason (1986 [81]) proposed using nonlinear least squares (NLS) where the sum of squared deviations is minimized in Equation 1.2. In this paper, I show results from both OLS and NLS procedures and NLS does offer a significantly better fit.

1.3.3 Results and Discussion

I conduct the analyses separately for both consumers' and theaters' adoption. The weekly box office revenue is used to reflect consumers' adoption and the weekly number of theaters showing the movie is used for theaters' adoption. To fully estimate the model, we need at least three weekly data points for each movie. With this condition, we are left with 5790 titles for estimating consumers' adoption and 5753 for estimating theaters' adoption. The coefficients m, p, and q are estimated for each movie title and I show in Table 1.3 the results of the four examples that have been mentioned in Figure 1.2. Both ordinary

[8] $p = \frac{a}{m}$, $q = -mc$, $m = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c}$

least squares (OLS) and nonlinear least squares (NLS) are used for estimation. The OLS procedure does not directly provide standard errors for the estimated parameters and, hence, the statistical significance of these estimates cannot be assessed (Mahajan, Muller, and Bass (1990 [60])). Moreover, the empirical results under the NLS procedure use the OLS estimates as the starting values, although different starting solutions have a negligible effect on the final estimates (Srinivasan and Mason (1986 [81])). In addition to the estimation results, the periodic adoption (sales) function can be outlined as a graph where time is the X-axis and weekly adoption is the Y-axis. I show the graphs of consumers' adoption in Figure 1.4 and theaters' adoption in Figure 1.5. On each subplot, I plot the actual adoption, the predicted adoption by OLS procedure, and the predicted adoption by NLS procedure.

Now, I explain the estimation results in detail. First, the value of the market potential m indicates how much adoption one movie will eventually achieve. We expect m to be positive and 99.7% of the sample titles do yield positive estimates for m. Next, I discuss how the value of p and q determine the shape of the non-cumulative adoption curve ($S(t) = mf(t)$). Bass (1969 [4]) categorized the shape of the sales curves based on the value of $\frac{q}{p}$. If $\frac{q}{p} > 1$, then imitation effects dominate the innovation effects; the adoption of the new product experience growth and then decline due to saturation. This is the pattern of *The Big Sick* and *Gifted*. On the other hand, if $\frac{q}{p} < 1$, then innovation effects dominate; the sales start at a certain level and keep declining. This usually happens to blockbuster movies (Lilien, Rangaswamy, and De Bruyn (2000 [53])). This can also be seen from the last column of Table 1.4. The median market potential (m) of the cases where $p \geq q > 0$ is significantly larger than the one of the cases where $q \geq p > 0$. Movies with dominating innovation effects (large p) attract higher adoption. As a summary, Van den Bulte (2002 [85]) concluded that the parameters p and q reflect the speed of diffusion. A high value of p indicates that the diffusion has a quick start, but also tapers off quickly. A high value of q indicates that the diffusion is slow at first, but accelerates after a while.

Furthermore, Srinivasan and Mason (1986 [81]) noted that both p and q must be non-negative, which is not the case for the example *Spider-Man: Homecoming* and *Cars 3*. In fact, positive p combined with negative q happens quite frequently in our sample (Table 1.4 and Figure A.5), especially for consumers' adoption (27.6% of the sample under OLS and 23.6% under NLS). I follow the interpretation from Orbach (2016 [67]) that negative q is caused by a declining motivation to adopt as the number of adopters increases. One example can be the negative word-of-mouth effect (Moldovan and Goldenberg (2004 [63])) and it can fit a case where the benefit from a product declines as more people adopt.

Lastly, I finish this subsection by discussing the model fit. I calculate the Mean Squared Error[9] (MSE) for each movie title under both OLS and NLS procedures. Using all the available data, the model offers a much better fit under the NLS procedure.

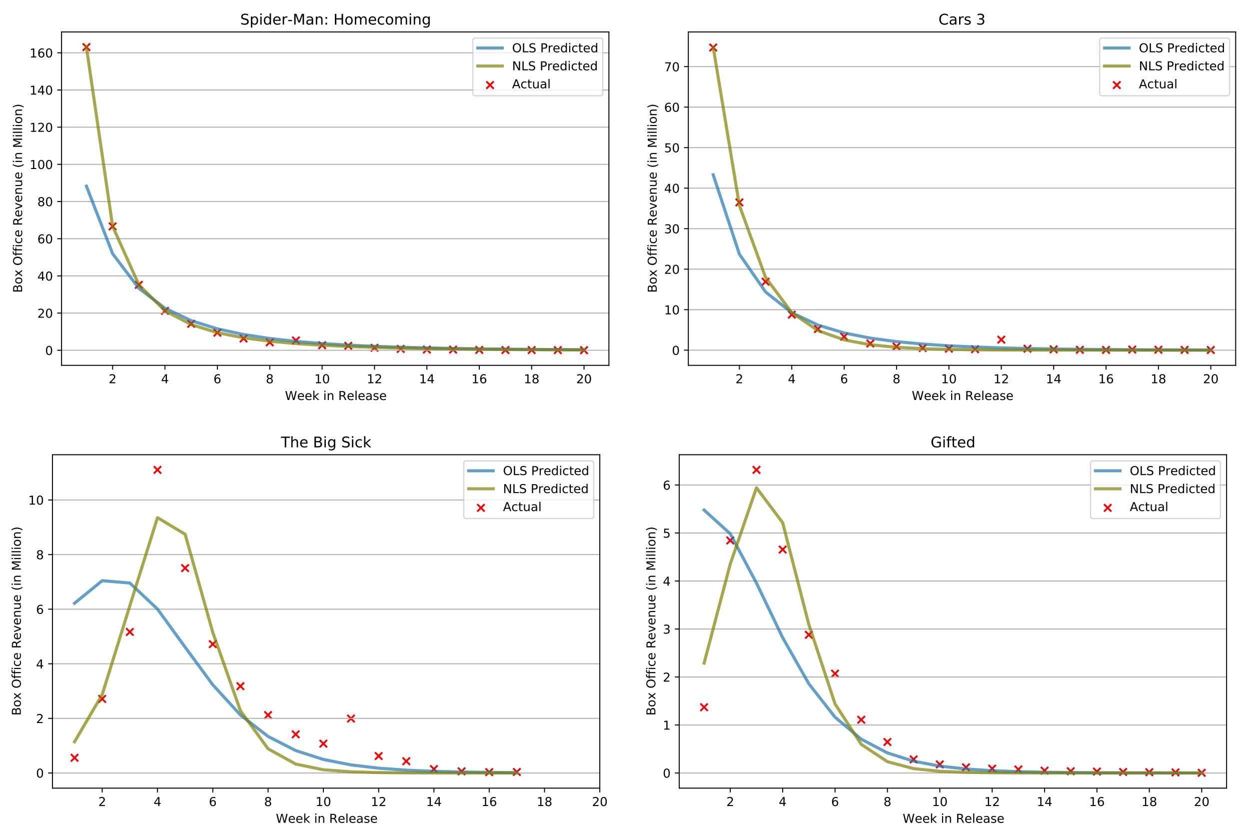


Figure 1.4: Consumers' Adoption: Regression Results of Fitting the Bass Model

[9]$\text{MSE}_\text{i} = \frac{1}{n}\sum_{t=1}^{n}\left(Y_{it} - \hat{Y}_{it}\right)^2$ where i is the index for the movie; t is the index for the week; n is the total number of weeks that the movie has been shown.

Table 1.3: Regression Results of The Bass Model: 4 Examples

Movie Title	OLS				NLS			
	m	p	q	MSE	m	p	q	MSE
Consumers' Adoption	(10^6)				(10^6)			
Spider-Man: Homecoming	335.188	0.520	-0.273	291.805	601.937	1.325	-1.077	0.231
					(12.120)	(0.052)	(0.075)	
Cars 3	152.196	0.627	-0.315	58.397	222.073	0.787	-0.152	0.481
					(7.643)	(0.065)	(0.122)	
Gifted	24.619	0.211	0.338	1.396	23.887	0.042	0.919	0.125
					(2.917)	(0.018)	(0.185)	
The Big Sick	42.125	0.115	0.414	5.659	37.318	0.011	1.016	0.875
					(2.950)	(0.004)	(0.109)	
Theaters' Adoption	(10^3)				(10^3)			
Spider-Man: Homecoming	35.503	0.125	0.100	0.059	38.016	0.116	0.127	0.027
					(5.034)	(0.028)	(0.105)	
Cars 3	24.435	0.236	-0.077	0.371	24.530	0.196	0.164	0.288
					(4.258)	(0.067)	(0.227)	
Gifted	11.574	0.089	0.506	0.066	11.713	0.041	0.656	0.007
					(3.354)	(0.038)	(0.330)	
The Big Sick	12.649	0.087	0.287	0.540	9.294	0.004	1.152	0.224
					(2.845)	(0.007)	(0.449)	

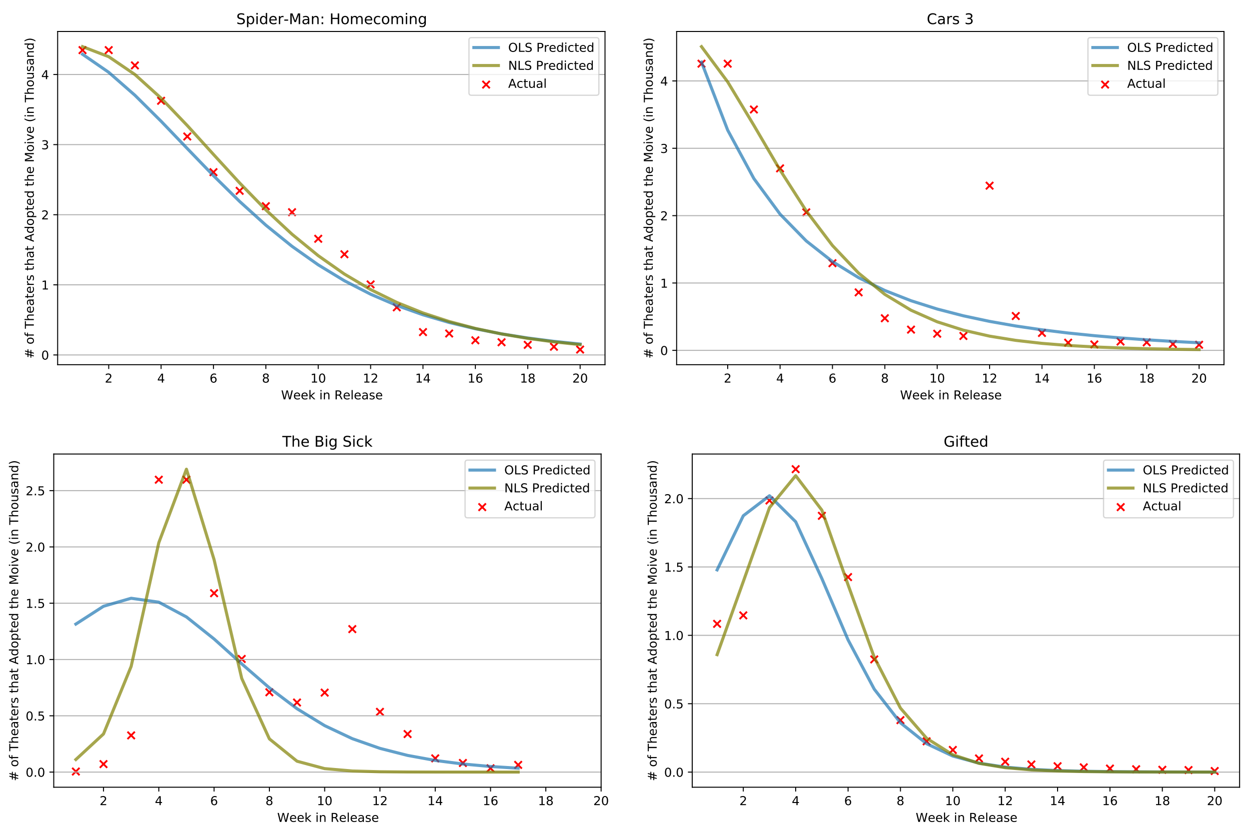


Figure 1.5: Theaters' Adoption: Regression Results of Fitting the Bass Model

Table 1.4: Regression Results of The Bass Model: An Overview

Range of p and q	# of the Titles	% of the Sample	Median m
Consumers' Adoption using OLS Procedure: 5762 Titles			(10^6)
$p \geq q > 0$	2133	37.0%	16.0
$q \geq p > 0$	1884	32.7%	2.3
$p > 0$ and $q < 0$	1592	27.6%	19.4
$p < 0$ and $q > 0$	153	2.7%	0.1
$p < 0$ and $q < 0$	0	0.0%	-
Consumers' Adoption using NLS Procedure: 5762 Titles			(10^6)
$p \geq q > 0$	1211	21.0%	40.9
$q \geq p > 0$	3194	55.4%	1.9
$p > 0$ and $q < 0$	1357	23.6%	81.7
$p < 0$ and $q > 0$	0	0.0%	-
$p < 0$ and $q < 0$	0	0.0%	-
Theaters' Adoption using OLS Procedure: 5732 Titles			(10^3)
$p \geq q > 0$	1500	26.2%	7.1
$q \geq p > 0$	3616	63.1%	2.9
$p > 0$ and $q < 0$	549	9.6%	3.9
$p < 0$ and $q > 0$	66	1.1%	0.1
$p < 0$ and $q < 0$	0	0.0%	-
Theaters' Adoption using NLS Procedure: 5732 Titles			(10^3)
$p \geq q > 0$	86	1.5%	15.2
$q \geq p > 0$	5550	96.8%	4.6
$p > 0$ and $q < 0$	96	1.7%	0.2
$p < 0$ and $q > 0$	0	0.0%	-
$p < 0$ and $q < 0$	0	0.0%	-

1.4 Online Word of Mouth from IMDB

Fitting the Bass model with the revenue data allows us to quantify the innovation and imitation effects for each sample movie. By comparing the values of p and q across different titles, one can roughly infer how word-of-mouth (WOM) communication shapes the revenue streams. However, no hypotheses can be tested here, and therefore, I collect an extra WOM data set from the user-review section of `IMDB.com` for further empirical analysis[10]. IMDB (Internet Movie Database) is an online database of information related to films, including cast, production crew and personal biographies, plot summaries, trivia, ratings, and critical reviews. As one of the most popular and informative movie websites in the U.S., IMDB can serve as a good source of movie WOM communication. I collect all the user review data for 9460 movie titles. Eventually, the data set includes 1,527,376 observations, each of which includes the full-text review, the reviewer's user identity number, the rating value out of a 10 scale, the date when the review is posted, if the review is a spoiler, the number of thumbs-up indicating the review is helpful, and the number of thumbs-down indicating the review is unhelpful. A scraped example is shown in Figure A.1. For the empirical analysis, I aggregate the WOM data to a weekly level for each title. In this section, I introduce how I construct the variables as well as provide summary statistics.

1.4.1 Volume and Valence

Volume and valence are the two most studied WOM attributes. Volume measures the total amount of WOM interaction and valence reflects the evaluative direction of the review. In this paper, I take the number of reviews posted as the indicator of the volume of WOM communication. Among the 9460 sample titles, each movie has received, on average, 161 reviews, much greater than the median, which suggests its distribution is heavily skewed to the right (Table 1.5). As for the valence of WOM, I adopt user-generated rating scores measured on a scale of 0 to 10. In contrast to the distribution of WOM volume, the valence

[10] The web scraping code in Python can be found here: `https://github.com/liaaaxu/IMDB-Scraping`.

is distributed more symmetrically with the sample mean equal to 5.97.

Table 1.5: Summary Statistics of Volume and Valence

Variables	Count	Mean	Std	Min	25%	50%	75%	Max
Review Volume	9460	161.46	372.53	1.0	15.0	25.0	154.0	8783.0
Review Rating	9460	5.97	1.48	0.0	5.0	5.96	6.96	10.0

Note: There are 1531262 sample reviews for 9460 sample movies in total. Volume is first aggregated to the movie-level. Valence is first averaged to the movie-level.

One important feature of online WOM communication is that it is dynamic. As time goes by, new reviews are added and both the volume and valence would change accordingly. I show in Figure 1.6 the weekly volume of reviews added (left panel) and the weekly change of the rating scores (right panel), averaging across all the sample titles[11]. I begin with the discussion on the dynamic volume pattern. WOM communication is most active during the pre-release period (week 0) and the opening week (from week 0 to 1). The volume decreases rapidly at first and then fluctuates at a lower level. Among our sample, 63.5% of the titles have the greatest weekly volume of reviews added before the release dates, and 22.2% of them have it during the opening week. The fact that movie WOM is most active during the very early period of the theatrical release when most of the moviegoers have not yet watched the movie suggests that WOM can be a useful measure to influence and explain box office sales (Liu (2006 [57])).

As for the dynamic valence pattern, the cumulative rating scores often experience drops at the beginning of the theatrical run. The right panel of Figure 1.6 indicates that, on average, movies start with more positive ratings in the pre-release period and then drop to lower levels after the opening week. One potential reason can be that the early audience who are the targeted consumers ("innovators" in the diffusion theory) are more likely to give positive

[11]The dashed lines stand for deviations of 1.96 standard errors.

comments; as the advertising generates awareness for the new products, more non-targeted consumers ("imitators" in the diffusion theory) who are less likely to give positive reviews are involved. Eventually, as the total volume of reviews increases, the review ratings tend to be more stable and the marginal changes bounce around zero.

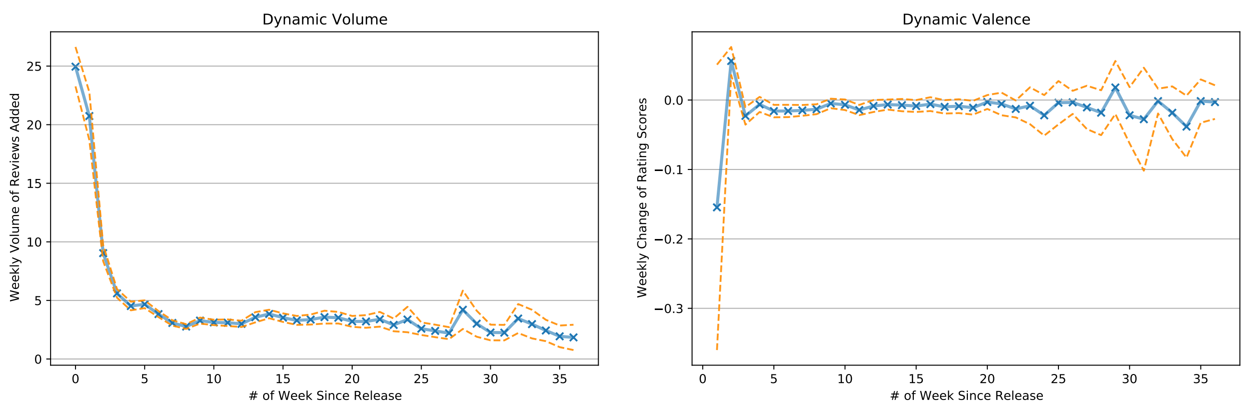


Figure 1.6: Dynamic Volume and Valence of Online WOM Communication

1.4.2 Features of Text Reviews

Besides the volume and valence of WOM, I also build features from the text reviews scraped from IMDB (Figure A.1). The raw text is processed through tokenization (breaking up each string into words and punctuation), removal of punctuation and stop words[12], and lemmatization[13] (Bird, Klein, and Loper (2009 [12])). With the help of Python packages, I create numerical indices to reflect certain aspects of the text data. I introduce these features in this subsection.

---

[12]Stop words are high-frequency words with little lexical content, such as *the*, *to*, and *also* that we sometimes want to filter out of a document before further processing.

[13]Lemmatization is a process that maps the various forms of a word (such as appeared, appears) to the canonical or citation form of the word, also known as the lexeme or lemma (e.g., appear).

Sentiment

Although the valence of WOM (user-generated rating scores) discussed previously can reflect users' attitudes towards the movie, it cannot fully express the mood and emotion in their words. Therefore, I conduct a simple sentiment analysis for each movie review and construct features using Python packages TextBlob (Loria (2018 [58])) and VADER (Valence Aware Dictionary and sEntiment Reasoner) (Hutto and Gilbert (2014 [49])). Table 1.6 lists the summary statistics of the sentiment features. Polarity and subjectivity are the two scores generated by TextBlob. Polarity score is a float that lies between $[-1, 1]$; -1 indicates negative sentiment and $+1$ indicates positive sentiment. Subjectivity score is a float that lies in the range of $[0, 1]$; 0 suggests very objective and 1 suggests very subjective. The sample average polarity and subjectivity scores are 0.15 and 0.53 respectively.

The compound score generated by VADER is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and then normalized to be between -1 (most extreme negative) and $+1$ (most extreme positive) (Hutto and Gilbert (2014 [49])). The typical thresholds for classifying sentences as either positive, neutral, or negative using the compound score is as follows: the sentiment is positive if compound score ≥ 0.05, negative if compound score ≤ -0.05, and neutral if compound score falls within $(-0.05, 0.05)$. Additionally, I show the correlation coefficients among user review rating, polarity, and sentiment compound score in Table 1.7. These three sentiment features are all positively correlated.

Cosine Similarity between User Reviews and Movie Plot Description

In Natural Language Processing (NLP), cosine similarity is one of the metrics to measure the text-similarity between two documents irrespective of their sizes[14]. In this paper, I use cosine similarity to measure the text-similarity between consumers' online movie reviews and the plot description shown on each movie's front page. I first adopt the method of bag-of-words

[14] Also see `https://studymachinelearning.com/cosine-similarity-text-similarity-metric/`

Table 1.6: Summary Statistics of Sentiment Features

Variables	Count	Mean	Std	Min	25%	50%	75%	Max
Using TextBlob								
Polarity Score	9460	0.15	0.07	-0.23	0.10	0.14	0.19	0.70
Subjectivity Score	9460	0.53	0.04	0.00	0.51	0.53	0.55	0.79
Using VADER								
Sentiment Compound Score	9460	0.50	0.34	-1.00	0.31	0.56	0.76	1.00

Note: There are 1531262 sample reviews for 9460 sample movies in total. All the scores are first averaged to the movie-level.

Table 1.7: Correlation Matrix of Sentiment Features

Variable	User Review Rating	Polarity	Sentiments Compound Score
User Review Rating	1.00		
Polarity	0.46	1.00	
Sentiment Compound Score	0.35	0.56	1.00

to vectorize the sample text reviews and plot description. Each document is represented by a vector that counts the occurrence of each word contained within the corpus using a document-term matrix. I calculate the cosine of the angle between the two vectors (the review vector and the plot vector) to quantify how similar the two documents are[15]. The value of the metric ranges from 0 to 1 where the value closer to 0 suggests less similarity.

[15]Cosine Similarity $= \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$

The average cosine similarity value of our sample reviews is 0.18.

Readability

Readability measures the ease with which a reader can understand a written text. In this paper, the readability of each movie review is computed in an approximate manner based on the complexity of the vocabulary and syntax used. Table 1.8 shows the summary statistics of readability features. I first count the number of words, sentences, syllables, and polysyllables in each movie review. It is expected that reviews with greater length or/and a larger number of complex words (e.g. polysyllables) may lower the readability but increase the content richness. Moreover, based on the basic counts, I calculate the commonly used readability indices for English writing using the formula listed in the footnote. Figure 1.7 shows the pairwise Pearson correlation matrix of all the readability indices. As expected, the count of words, sentences, syllables, and polysyllables are all highly correlated to one another, and so do the four readability indices.

Helpfulness

Lastly, I define review helpfulness as follows.

$$\text{Helpfulness} = \frac{\text{Number of Thumbs-up}}{\text{Total Number of Votes}} \tag{1.5}$$

where total number of votes is the sum of thumbs-up and thumbs-down. Helpfulness measures the quality of the reviews and we expect that movies with more "helpful" reviews tend to attract more consumers.

[16]Flesch Reading Ease Score $= 206.835 - 1.015 \times (\frac{\text{total words}}{\text{total sentences}}) - 84.6 \times (\frac{\text{total syllables}}{\text{total words}})$. Higher scores indicate that the material is easier to read.

[17]Flesch–Kincaid Grade Level $= 0.39 \times (\frac{\text{total words}}{\text{total sentences}}) + 11.8 \times (\frac{\text{total syllables}}{\text{total words}}) - 15.59$. The result is a number that corresponds with a U.S. grade level.

[18]Gunning Fog Index $= 0.4 \times [(\frac{\text{total words}}{\text{total sentences}}) + 100 \times (\frac{\text{total polysyllables}}{\text{total words}})]$. The index estimates the years of formal education a person needs to understand the text on the first reading.

[19]Smog Index $= 1.0430 \times \sqrt{\text{total polysyllables} \times \frac{30}{\text{total sentences}}} + 3.1291$. The score estimates the years of education needed to understand a piece of writing.

Table 1.8: Summary Statistics of Readability Features

Variables	Count	Mean	Std	Min	25%	50%	75%	Max
Basic Counts								
Word Count	9460	253.01	75.63	10.00	207.44	251.35	294.56	1234.00
Sentence Count	9460	11.15	2.83	1.00	9.51	11.20	12.69	49.00
Syllable Count	9460	350.85	109.30	18.00	283.87	346.22	408.52	1902.00
Polysyllable Count	9460	23.65	9.09	0.00	17.80	22.50	27.86	170.00
Syllables Per Word	9460	1.38	0.05	1.17	1.35	1.38	1.41	2.83
Words Per Sentence	9460	25.75	6.68	7.80	22.50	24.91	27.57	173.00
Polysyllables Per Sentence	9460	2.40	0.89	0.00	1.91	2.25	2.71	27.00
Readability Indices for English Writing								
Flesch Reading Ease Score[16]	9460	63.90	8.63	-79.77	60.38	65.07	68.80	94.79
Flesch–Kincaid Grade Level[17]	9460	10.74	2.76	1.98	9.33	10.37	11.58	67.36
Gunning Fog Index[18]	9460	14.00	2.85	5.17	12.53	13.61	14.87	71.97
Smog Index[19]	9460	11.43	1.31	3.13	10.61	11.29	12.11	32.81

Note: There are 1531262 sample reviews for 9460 sample movies in total. All the indices are first averaged to the movie-level.

1.5 Empirical Analysis

I list all the variables along with their descriptions in Table 1.9. Natural logarithms are taken for variables that could take unbounded positive values. All the variables are aggregated to a weekly level to match the weekly box office report dates. Besides the features that have been mentioned in Section 1.4, I add $RivalAge_{it}$ to control for the effects of competition. $RivalAge_{it}$ is the average age of all rivals of movie i in week t, measured in weeks. We expect higher rivals' age may help the movie's own demand. Additionally, I generate one more set of

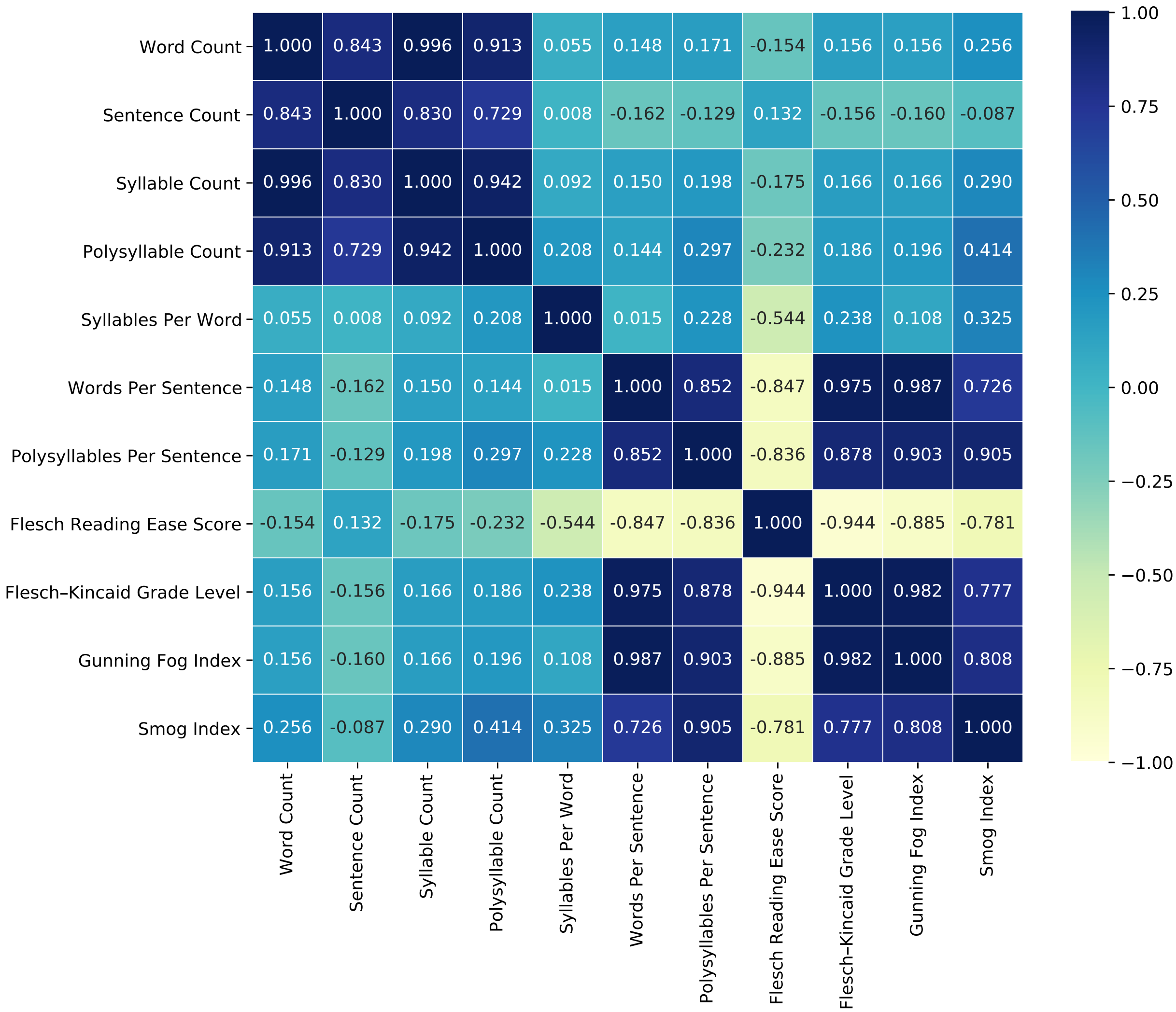


Figure 1.7: Pearson Correlation of Readability Indices

variables to represent the valence of WOM: the fraction of the reviews that are n-star where $n = 0, 1, \ldots, 10$. These features are computed by dividing the number of n-star reviews by the number of total reviews posted each week for each movie. I specifically include $FracStar10_{it}$ and $FracStar0_{it}$ to examine if the fraction of extreme positive or negative WOM would affect the movie demand.

The remaining part of this section presents the benchmark and extended models, shows the results, and discusses how various WOM attributes help explain the weekly theatrical revenue.

1.5.1 The Benchmark Model

To examine the weekly effect of WOM attributes, volume and valence, on weekly movie demand, I estimate the following benchmark model.

$$\begin{aligned}\log(Revenue_{it}) = \alpha_0 + \alpha_1 \log(Theater_{it}) + \alpha_2 RivalAge_{it} \\ + \alpha_3 \log(ReviewVol_{i,t-1}) + \alpha_4 ReviewRating_{i,t-1} \\ + v_i + \epsilon_{it} \qquad (1.6)\end{aligned}$$

It is recognized that the unobserved movie fixed effects are correlated with independent variables. To avoid omitted variable problems, I include v_i as movie fixed effects to capture all unobserved, time-constant factors that affect $\log(Revenue_{it})$. The estimation results are displayed in Table 1.10. All the regression results have achieved high levels of model fit with the adjusted R^2 greater than 0.7.

Theatrical Supply and Competition Effect

Table 1.10, Column 1 shows the results for a fixed effects regression in which no WOM attributes are included, only the theatrical supply control, $\log(Theater_{it})$, and the competition effect, $RivalAge_{it}$. Both coefficients are positive and significant. The coefficient of $\log(Theater_{it})$ tells that a 1% increase in the number of theaters showing the movie would lead to a 1.231% increase in its box office revenue, suggesting the importance of the supply availability. The coefficient of $RivalAge_{it}$ shows that if the movie's rivals are on average 1 week older in the market, the movie's own theatrical revenue can rise 2.64%. This is as expected since consumers generally prefer new releases to old ones. Table 1.10, Column 2 includes the weekly dummies to control the seasonality. The motivation is that movie demand and market competition fluctuates at certain weeks of the year, for example, holiday seasons

Table 1.9: Variables, Descriptions, and Measures

Variable	Description
Basic Features	
$\log(Revenue_{it})$	Box office revenue of movie i in week t (in natural log)
$\log(Theater_{it})$	Number of theaters showing movie i in week t (in natural log)
$\log(ReviewVol_{it})$	Number of reviews posted in week t for movie i (in natural log)
$ReviewRating_{it}$	Average rating score of movie i's reviews posted in week t, ranging from 0 to 10
Additional WOM Valence Features	
$FracStar10_{it}$	Fraction of movie i's reviews posted in week t that are rated 10-star
$FracStar0_{it}$	Fraction of movie i's reviews posted in week t that are rated 0-star
Text Features	
$FracSpoiler_{i,t}$	Fraction of movie i's reviews posted in week t that are marked as spoiler warning
$FracPos_{it}$	Fraction of movie i's reviews posted in week t that are positive (sentiment compound score ≥ 0.05)
$FracNeg_{it}$	Fraction of movie i's reviews posted in week t that are negative (sentiment compound score ≤ -0.05)
$Polarity_{it}$	Average polarity score of movie i's reviews posted in week t, falls within $[-1, 1]$
$Subjectivity_{it}$	Average subjectivity score of movie i's reviews posted in week t, falls within $[0, 1]$
$CosSim_{it}$	Average cosine similarity (between reviews and plot description) of movie i in week t, falls within $[0, 1]$
$\log(WordCount_{it})$	Average length (word count) of movie i's reviews posted in week t (in natural log)
$\log(Readability_{it})$	Average Flesch reading ease score / Flesch-Kincaid Grade Level / Gunning Fog Index / Smog Index / syllables per word / words per sentence / polysyllables per sentence of movie i's reviews in week t (in natural log)
$Helpfulness_{it}$	Average helpfulness of movie i's reviews posted in week t, falls within $[0, 1]$
Competition Features	
$RivalAge_{it}$	Average age of movie i's rivals in week t

Note: t is the week number in each movie's life span (i.e. the number of weeks since the movie's initial theatrical release).

represent higher demand and more intense competition. After adding the weekly dummies, the two coefficients still remain positive and significant.

Volume and Valence of WOM

Table 1.10, Column 3 adds the two lagged WOM attributes: the total number of reviews posted in the last week and the average rating score of the reviews posted in the last week. The positive and significant coefficient of $\log(ReviewVol_{i,t-1})$ shows that a 1% increase in the total number of reviews added during the last week would lead to a 0.424% increase in the theatrical revenue of this period. This result is consistent with what previous studies have found that consumers are influenced by the awareness effect and WOM volume can explain future movie sales. As for WOM valence, unlike the insignificant coefficients presented in Liu (2006 [57]), my results offer a positive and significant coefficient of $ReveiwRating_{i,t-1}$. It suggests that a higher average rating of the reviews posted in the last week can boost the box office revenue in this period.

Table 1.10, Columns 4 - 7 use alternative ways of measuring WOM valence. In Column 4, in place of the average rating score, the fraction of 0-star reviews and the fraction of 10-star reviews are included to examine if extreme positive or negative WOM would affect the movie demand. The two coefficients are both positive and significant, indicating that reviews with extreme rating scores, no matter whether they are positive or negative, can all attract attention and increase the demand. And in terms of the magnitude, the coefficient of $FracStar10_{i,t-1}$ is larger than the one of $FracStar0_{i,t-1}$, thus extremely positively rated reviews can boost the sales more. In Column 5, we further break it down to include the fraction of n-star reviews where $n = 0, 1, \ldots, 10$. It shows that a higher fraction of 1-star or 2-star reviews would hurt sales, although insignificantly. Lastly, for Columns 6 and 7, we adopt the valence features generated from the text data instead of directly using the ratings assigned by the users. In this way, we can further capture the emotions in the words. In Column 6, the fraction of positive and negative reviews are included based on the sentiment compound scores generated by VADER. A similar conclusion can be drawn that both positive and negative reviews can increase future sales and positive ones can increase more in magnitude. In Column 7, we use the polarity score generated by Textblob, and again, the coefficient remains positive and significant.

1.5.2 The Extended Model

The benchmark results confirm the importance of both WOM volume and valence. Now we extend the model to include more text features and the results are shown in Table 1.11.

Helpfulness Table 1.11, Column 1 includes the average helpfulness (defined in Equation 1.5) of the reviews posted in the last week. Helpfulness is an indicator of review quality and informativeness. As expected, the positive and significant coefficient suggests that if the reviews posted in the last period were voted with higher average helpfulness, more audiences would be attracted and theatrical sales in this period would increase.

Spoiler Warning Table 1.11, Column 2 adds the fraction of the reviews that are marked with spoiler warnings in the last period. The tag spoiler warning is used when the reviewer would like to warn the reader that a plot spoiler is about to be revealed, which also hints that the review contains solid information about the movie. The positive and significant coefficient of $FracSpoiler_{i,t-1}$ tells that the audience does read the reviews even if they are marked as spoilers. And more consumers would be attracted to purchase if a higher fraction of the reviews posted in the last period were tagged as spoilers.

Cosine Similarity Table 1.11, Column 3 includes the cosine similarity measure between each review and the plot description. A higher value of the measure means that the review shows more common words with the plot description, and therefore, it is more likely that the review discusses the story plot with relevant information. The coefficient of cosine similarity is positive and significant, which shows that this week's theatrical sales would be boosted if the reviews posted during the last week, on average, overlapped the plot description more.

Subjectivity Table 1.11, Column 4 shows how the sales can be affected by the subjectivity of the reviews. The subjectivity score measures how much the review is based on or influenced by personal feelings, tastes, or opinions. The positive and significant coefficient suggests that more subjective reviews are associated with higher box office revenue. This result is in line with the intuition since films are one of the typical experience goods and movie reviews are expected to reflect personal preferences.

Readability Table 1.11, Columns 4 and 5 are designed to study the impact of text readability on movie sales. The two measures, length of the review (total word count) and readability index Flesch-Kincaid Grade Level are included. Longer reviews represent richer content and more effort on the part of the reviewers (Chevalier and Mayzlin (2006 [17])), although the readability may be lowered. Similarly, a higher value of Flesch-Kincaid Grade Level reflects lower readability but also indicates that the review is with longer sentences and more complex words. The coefficients of both measures are positive and significant. We can infer that 1) the audience does read the reviews instead of relying only on summary statistics and 2) content richness outweighs readability; more consumers would make the purchase if reviews posted in the last period are, on average, longer and more difficult to read.

1.5.3 Wide and Limited Releases

The previous results reflect the average level in the industry, however, movies are assigned to different release patterns before they enter the theatrical market. In general, there are two types of releases, wide and limited. Wide releases are most likely from major studios with higher production and advertising budgets while limited releases are relatively small movies belonging to a different segment of the industry (Einav (2007 [29])). In our data set, a movie released in 600 theaters or more is considered a wide release, otherwise, a limited release. In this section, I re-estimate the model separately for wide and limited releases and the results are shown in Table 1.12 and 1.13.

The regressions displayed in Table 1.10, Columns 2, 3, 5, and 6 are re-estimated and the results are shown in Table 1.12. I discuss the key findings here. First, comparing the coefficients of $\log(Theater_{it})$ in Columns 1 and 2, we can infer that, on average, a 1% increase in the number of theaters showing the wide-released movie would lead to a 1.369% increase in its box office revenue while for the case of limited release, the revenue increase would be lower, 1.010%. Moreover, the adjusted R^2 is much greater for wide releases (0.89) than for limited ones (0.54), suggesting that the theatrical supply availability offers a much higher explanatory power for the demand for wide-released movies. This also means that other factors can potentially be brought in to explain the sales of limited releases. Second, the coefficients of $RivalAge_{it}$ are positive as we expected in all columns of Table 1.10, but only significant for limited releases. This suggests that the competition effect only works on limited releases and the reason could be that limited releases are small in production and more fragile when facing rivals.

Third, the coefficients of the volume of WOM, $\log(ReviewVol_{i,t-1})$, are much larger for limited releases. Specifically, in Table 1.10, Columns 3 and 4, a 1% increase in the total number of reviews posted during the last week would lead to a 0.288% increase in the current revenue of a wide release, while the increase would be almost doubled, 0.528%, for a limited-released movie. The results suggest that limited releases can potentially benefit more from the online awareness effect. To explain intuitively, limited releases are usually associated with less marketing budget; fewer consumers would know the movie title before the actual release. Thus, online WOM can create a larger marginal impact on movie sales. In contrast, wide releases are more likely to have received a large awareness effect from the advertising campaigns and the marginal awareness effect from online WOM would be relatively small.

Lastly, Table 1.10, Columns 5 - 7 show that the impact of the valence of WOM on the movie demand is also much greater in magnitude for limited-released movies. Interestingly, the coefficient of $FracNeg_{i,t-1}$ is negative and significant for wide releases (Column 7) but positive and significant for limited releases (Column 8). Similarly, the coefficients of $FracStar1_{i,t-1}$, $FracStar2_{i,t-1}$, $FracStar3_{i,t-1}$, $FracStar4_{i,t-1}$, $FracStar5_{i,t-1}$, and $FracStar6_{i,t-1}$ are all negative and mostly significant for wide releases (Column 5) but mostly positive and insignificant for limited releases (Column 6). This evidence suggests that compared to limited releases, wide-released movies are more sensitive to negative reviews (but not extremely negative reviews) and more insensitive to positive reviews. One explanation could be that wide releases that commonly carry a highly positive reputation originating from studios' marketing campaigns are weaker to the impact of online negative tones since they tend to bring more disappointment to the potential audience. Meanwhile, limited releases work the other way around; the audience tends to expect less from them but would give a strong response once positive WOM emerges.

The regressions displayed in Table 1.11 are re-estimated and the results are shown in Table 1.13. Overall, the coefficients of the WOM text features are all larger in magnitude for limited releases which is consistent with the evidence in Table 1.10. Limited releases could benefit more on the margin if studios manage the online WOM effect properly. Specifically, subjective reviews with rich content even at the cost of low readability can help boost the box office revenue. I now discuss the new evidence for wide releases. The negative and significant coefficient of $FracSpoiler_{i,t-1}$ for wide releases indicates that fewer consumers would be attracted to purchase if a higher fraction of the reviews posted in the last period were tagged as spoilers. This is the exact opposite of limited releases. One explanation can be that consumers tend to be excited about big hit movies due to the promotion and spoilers would make them lose interest, but limited releases with insufficient publicity can take advantage of the spoilers to get the audience attracted. Furthermore, the coefficients of $CosSim_{i,t-1}$, $Subjectivity_{i,t-1}$, $\log(WordCount_{i,t-1})$, and $\log(FleschKincaid_{i,t-1})$ are all insignificant and mostly with the opposite signs compared to the limited releases, which suggests that these text features do not matter to wide releases. The comparison between wide and limited releases suggests the substitution between online WOM and marketing spending; less insufficient advertising funds can be made up through online WOM.

Table 1.10: The Effect of WOM on Theatrical Sales: Benchmark Model

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\log(Theater_{it})$	1.231***	1.234***	1.133***	1.132***	1.132***	1.132***	1.132***
	(227.70)	(238.31)	(228.01)	(227.84)	(228.07)	(227.49)	(227.81)
$RivalAge_{it}$	0.0264**	0.0391***	0.0277**	0.0282***	0.0277**	0.0280***	0.0283***
	(3.01)	(4.13)	(3.28)	(3.34)	(3.29)	(3.33)	(3.35)
$\log(ReviewVol_{i,t-1})$			0.424***	0.437***	0.418***	0.419***	0.432***
			(85.97)	(89.25)	(84.09)	(84.22)	(88.95)
$ReviewRating_{i,t-1}$			0.0231***				
			(20.75)				
$FracStar0_{i,t-1}$				0.156***	0.202***		
				(9.50)	(12.11)		
$FracStar10_{i,t-1}$				0.213***	0.262***		
				(13.10)	(15.81)		
$FracStar1_{i,t-1}$					-0.00883		
					(-0.40)		
$FracStar2_{i,t-1}$					-0.00504		
					(-0.17)		
$FracStar3_{i,t-1}$					0.0460		
					(1.61)		
$FracStar4_{i,t-1}$					0.0322		
					(1.11)		
$FracStar5_{i,t-1}$					0.0525*		
					(2.03)		
$FracStar6_{i,t-1}$					0.111***		
					(5.08)		
$FracStar7_{i,t-1}$					0.193***		
					(10.52)		
$FracStar8_{i,t-1}$					0.202***		
					(11.70)		
$FracStar9_{i,t-1}$					0.240***		
					(12.96)		
$FracPos_{i,t-1}$						0.191***	
						(21.32)	
$FracNeg_{i,t-1}$						0.0824***	
						(5.70)	
$Polarity_{i,t-1}$							0.581***
							(17.74)
N	83725	83725	83725	83725	83725	83725	83725
adj. R^2	0.749	0.754	0.797	0.796	0.797	0.797	0.796
Movie Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seasonality Fixed Effect	No	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.11: The Effect of WOM on Theatrical Sales: Extended Model

	(1)	(2)	(3)	(4)	(5)	(6)
$\log(Theater_{it})$	1.133***	1.133***	1.133***	1.133***	1.133***	1.133***
	(228.33)	(228.04)	(227.92)	(227.89)	(227.85)	(227.96)
$RivalAge_{it}$	0.0279***	0.0278***	0.0277**	0.0278***	0.0277**	0.0277**
	(3.32)	(3.30)	(3.28)	(3.30)	(3.29)	(3.28)
$\log(ReviewVol_{i,t-1})$	0.410***	0.422***	0.420***	0.419***	0.416***	0.416***
	(81.60)	(85.66)	(84.95)	(84.50)	(83.20)	(83.43)
$ReviewRating_{i,t-1}$	0.0101***	0.0221***	0.0150***	0.0161***	0.0141***	0.0140***
	(6.80)	(19.50)	(10.82)	(9.79)	(8.54)	(8.68)
$Helpfulness_{i,t-1}$	0.217***					
	(12.63)					
$FracSpoiler_{i,t-1}$		0.0848***				
		(4.21)				
$CosSim_{i,t-1}$			0.422***			
			(9.07)			
$Subjectivity_{i,t-1}$				0.132***		
				(5.67)		
$\log(WordCount_{i,t-1})$					0.0168***	
					(7.29)	
$\log(FleschKincaid_{i,t-1})$						0.0407***
						(7.80)
N	83725	83725	83725	83725	83725	83725
adj. R^2	0.797	0.797	0.797	0.797	0.797	0.797
Movie Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Seasonality Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.12: The Effect of WOM on Theatrical Sales: Wide and Limited Releases

	(1) Wide	(2) Limited	(3) Wide	(4) Limited	(5) Wide	(6) Limited	(7) Wide	(8) Limited
$\log(Theater_{it})$	1.369***	1.010***	1.263***	1.006***	1.263***	1.004***	1.263***	1.004***
	(268.49)	(117.64)	(211.11)	(124.94)	(211.69)	(124.92)	(211.27)	(124.61)
$RivalAge_{it}$	0.0169	0.0533**	0.0143	0.0354*	0.0142	0.0360*	0.0142	0.0364*
	(1.85)	(3.18)	(1.72)	(2.41)	(1.72)	(2.46)	(1.71)	(2.48)
$\log(ReviewVol_{i,t-1})$			0.288***	0.528***	0.288***	0.514***	0.289***	0.515***
			(49.68)	(57.52)	(49.70)	(54.99)	(49.65)	(54.96)
$ReviewRating_{i,t-1}$			0.00610***	0.0291***				
			(4.29)	(17.44)				
$FracStar0_{i,t-1}$					0.0850***	0.316***		
					(4.84)	(10.50)		
$FracStar1_{i,t-1}$					-0.0806***	-0.000447		
					(-3.69)	(-0.01)		
$FracStar2_{i,t-1}$					-0.119***	0.0600		
					(-4.00)	(1.05)		
$FracStar3_{i,t-1}$					-0.0495	0.0956		
					(-1.80)	(1.82)		
$FracStar4_{i,t-1}$					-0.0528*	0.0595		
					(-2.11)	(1.02)		
$FracStar5_{i,t-1}$					-0.0518*	0.0993*		
					(-2.00)	(2.22)		
$FracStar6_{i,t-1}$					-0.0526*	0.214***		
					(-2.46)	(5.71)		
$FracStar7_{i,t-1}$					0.00254	0.294***		
					(0.14)	(9.71)		
$FracStar8_{i,t-1}$					0.0334	0.257***		
					(1.57)	(10.25)		
$FracStar9_{i,t-1}$					0.0716**	0.295***		
					(3.07)	(11.23)		
$FracStar10_{i,t-1}$					0.126***	0.297***		
					(5.82)	(12.19)		
$FracPos_{i,t-1}$							0.0452***	0.262***
							(4.03)	(18.85)
$FracNeg_{i,t-1}$							-0.0571***	0.151***
							(-3.38)	(6.62)
N	40862	42863	40862	42863	40862	42863	40862	42863
adj. R^2	0.890	0.543	0.905	0.620	0.906	0.621	0.905	0.620
Movie Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seasonality Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1.13: The Effect of WOM on Theatrical Sales: Extended Model (Wide and Limited Releases)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Wide	Limited	Wide	Limited	Wide	Limited	Wide	Limited	Wide	Limited	Wide	Limited
$\log(Theater_{it})$	1.263***	1.005***	1.263***	1.005***	1.263***	1.005***	1.263***	1.005***	1.263***	1.004***	1.263***	1.004***
	(211.33)	(125.28)	(211.14)	(125.03)	(211.09)	(124.84)	(211.17)	(124.73)	(211.07)	(124.69)	(211.28)	(124.92)
$RivalAge_{it}$	0.0143	0.0363*	0.0141	0.0359*	0.0143	0.0357*	0.0143	0.0360*	0.0143	0.0359*	0.0143	0.0358*
	(1.73)	(2.49)	(1.70)	(2.44)	(1.72)	(2.43)	(1.72)	(2.45)	(1.72)	(2.45)	(1.72)	(2.44)
$\log(ReviewVol_{i,t-1})$	0.283***	0.504***	0.289***	0.525***	0.287***	0.521***	0.289***	0.515***	0.290***	0.511***	0.289***	0.511***
	(48.52)	(52.81)	(49.84)	(57.09)	(49.47)	(56.35)	(49.78)	(55.11)	(49.52)	(53.83)	(49.44)	(54.15)
$ReviewRating_{i,t-1}$	0.00247	0.00698**	0.00684***	0.0272***	0.00470**	0.0167***	0.00714***	0.0139***	0.00824***	0.0110***	0.00696***	0.0119***
	(1.47)	(2.71)	(4.79)	(15.79)	(2.96)	(7.46)	(3.92)	(5.01)	(4.59)	(3.97)	(4.05)	(4.35)
$Helpfulness_{i,t-1}$	0.0746***	0.335***										
	(3.69)	(11.19)										
$FracSpoiler_{i,t-1}$			-0.127***	0.130***								
			(-5.70)	(4.71)								
$CosSim_{i,t-1}$					0.0888	0.583***						
					(1.71)	(7.80)						
$Subjectivity_{i,t-1}$							-0.0206	0.282***				
							(-0.83)	(6.89)				
$\log(WordCount_{i,t-1})$									-0.00450	0.0320***		
									(-1.78)	(8.21)		
$\log(FleschKincaid_{i,t-1})$											-0.00434	0.0721***
											(-0.76)	(8.10)
N	40862	42863	40862	42863	40862	42863	40862	42863	40862	42863	40862	42863
adj. R^2	0.905	0.621	0.905	0.620	0.905	0.620	0.905	0.620	0.905	0.620	0.905	0.620
Movie Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Seasonality Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

1.6 Conclusion

In this paper, I employ a linear fixed effects model to examine the impact of online WOM on box office revenue using text review data extracted from IMDB. Following previous literature, I first estimate the benchmark model focusing on testing how the volume and valance of WOM affect movie sales. I count the total number of reviews posted as the index for the volume of WOM and take the average of user-generated rating scores as the measure for the valence of WOM. The results show that higher volume and valence of WOM in the last period are positively associated with current movie sales. Moreover, by replacing the benchmark measure of the valence of WOM with the fraction of 0-star reviews and the fraction of 10-star reviews, I found that reviews with extreme ratings, no matter whether they are positive or negative, can all attract attention and increase the movie demand.

This study contributes to the current literature by further constructing text features using basic natural language processing techniques. I extend the benchmark model to include review helpfulness, the fraction of spoilers, cosine similarity between reviews and plot descriptions, subjectivity score of the reviews, length of the reviews, and readability measures. The evidence shows that the audience does read the reviews instead of relying only on summary statistics. Plus, subjective reviews with rich content even at the cost of low readability can potentially boost the box office revenue.

Lastly, I re-estimated the model separately for wide and limited releases. The results show that the impact of online WOM on movie demand is larger for limited releases. Wide releases that frequently carry a positive reputation before the actual release due to intense marketing campaigns are more sensitive to negative reviews and less sensitive to positive ones while limited releases show the exact opposite pattern. The disadvantage of the insufficient marketing budget for limited releases could potentially be compensated by managing the online WOM effect properly.

I conclude this chapter by discussing some opportunities for future research. First, the unresolved simultaneity problem deserves further investigation that the causality between product sales and WOM effects works in both directions. In this paper, I treat WOM of the last period as a determinant of current movie sales, but at the same time, WOM can also be the outcome of past purchasing behavior. Moreover, with the large number of reviews posted online, each website may have different sorting algorithms to order the reviews. For example, IMDB offers the option of sorting user reviews by helpfulness, review date, total votes, prolific reviewer, or review rating. The order can decide which reviews have more chance to be seen and thus higher weights should be assigned when modeling (position bias). Future studies can examine how the order of the reviews displayed online may affect product sales.

Chapter 2

MOVIE DEMAND ESTIMATION: EVIDENCE FROM THEATRICAL AND HOME VIDEO MARKETS

2.1 Introduction

Starting in 1986, the gross revenue generated from home videos (about $2 billion) exceeds the one from theaters ($1.6 billion), which altered the fundamental structure of the U.S. movie business (Vogel (2020 [86])). However, compared to the large bulk of literature on theatrical demand estimation[1], the home video demand has received surprisingly less attention from researchers. This can be explained by the data availability that video revenue data is difficult to achieve while box office revenues are open to the public (Eliashberg, Weinberg, and Hui (2008 [36])). Given the nature of sequential distribution in the motion picture industry[2], the two markets are intercorrelated and many important business decisions require distributors to consider the two markets together. Therefore, it is worthwhile to bring the home video market in and investigate how consumers' purchasing behavior differs across the two markets, which is the gap in the economics literature that this paper aims to fill in.

In this paper, I first conduct demand estimation separately for theatrical and home video markets using logit and one-level nested logit models. This allows me to quantify movie qualities, consumers' utility decay rates, seasonality in demand, and market expansion effect. By comparing the results across the two markets, we can better understand how these two markets are operated differently. Next, I pool all the sample movies from the two markets together and implement a two-level nested logit to study consumers' substitution pattern between the two watching modes. I follow the well-established literature on demand

[1] Also discussed in Chapter 1.

[2] Movies are commonly first released into theaters, then to the home video market after, on average, 12 weeks. Discussed in Chapter 3.

estimation with discrete choice models in the industrial organization (IO) field, especially the ones regarding using aggregate data including Berry (1994 [11]), Berry, Levinsohn and Pakes (1995 [10]), Einav (2007 [29]), and Chiou (2007 [18]).

The remaining paper is organized as follows. Section 2.2 offers the data sources and descriptive statistics of the product features I adopt. Section 2.3 presents the demand structural model. Section 2.4 discusses the results and interpretation. Section 2.5 concludes.

2.2 Data

The data for this paper consists of two parts: 1) film features and 2) panel data on weekly revenue of theatrical and home video markets. The film features are collected mainly from Internet Movie Database (IMDB). The weekly revenue data are requested from Nash Information Services[3]. Additionally, US Census data on population is adopted as the market size. In the following subsections, I present the two main datasets in detail.

2.2.1 Panel Data on Weekly Demand

The panel data on weekly box office revenue cover 109536 entries of 10985 titles released between 1997 and 2019. The panel data on weekly home video revenue (DVD + Bluray) cover 258211 entries of 5760 titles released between 2006 and 2019. After merging the two datasets, we have 144029 entries of 2489 titles.

Figure 2.1 plots the distribution of movies' (log) total revenue in theatrical (left panel) and home video (right panel) markets; total revenue is calculated by aggregating all weekly revenue streams. The figure emphasizes that the release pattern matters to the sales. In practice, limited releases are relatively small movies belonging to a different segment of the industry (Einav (2007 [29])) and they, on average, earn much less revenue. In this paper, I restrict the attention to wide releases only.

---

[3]OpusData Extracts from Nash Information Services, LLC (NIS): `https://www.opusdata.com/documentation/index.php/Database_Extracts`

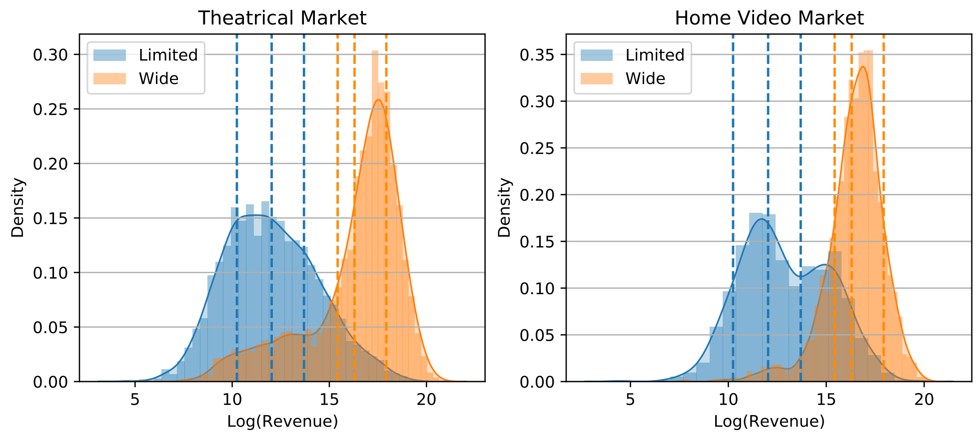


Figure 2.1: Total Revenue Distribution in Theatrical and Home Video Markets

Table 2.1 provides the descriptive statistics of demand variables for the domestic wide releases in our sample. Most of the demand variables are self-explanatory in terms of what they measure and how they are constructed. I highlight several here to be more clear. 1) Theatrical and video market shares are both calculated by dividing the domestic revenue by unit price and by U.S. population. In this way, we can eliminate the impact of population growth and price inflation to reflect the total number of units sold per capita. 2) Theatrical engagement, measuring the theatrical market supply, is calculated as the sum of the weekly number of theaters that the movie has been played in. One unit of theatrical engagement means that the movie has been played in one theater for one week throughout its entire theatrical run. 3) IMDB number of votes records the number of votes one title has received on `IMDB.com`. 4) Domestic video revenue (or units sold) in this paper is computed by summing up the revenue (or units sold) of DVD and Bluray. Almost all of the demand variables have the mean much higher than the median, suggesting the distribution is heavily skewed to the right. The two exceptions are opening weekend theaters and theatrical engagement of which the distributions are relatively more symmetric.

Figure 2.2 displays the pairwise Pearson correlation coefficients of the demand variables listed in Table 2.1. All the variables show a positive correlation with one another. Movies

Table 2.1: Descriptive Statistics (Domestic Wide Releases)

Variables	Count	Mean	Median	Std
Box Office Revenue ($M)	3876	50.4	25.2	76.8
Theatrical Market Share (%)	3876	2.4	1.2	3.6
Opening Weekend Revenue ($M)	3876	15.2	8.3	23.7
Opening Weekend Theaters (K)	3876	2.0	2.3	1.3
Theatrical Engagements (K)	3876	10.4	9.1	8.9
IMDB Number of Votes (K)	3876	104.2	43.5	172.2
Video Revenue ($M)	1810	30.5	16.9	44.7
Video Units (M)	1811	1.8	1.0	2.5
Video Market Share (%)	1811	0.6	0.3	0.8
DVD Revenue ($M)	1780	23.4	12.4	38.1
DVD Units (M)	1780	1.5	0.8	2.2
Bluray Revenue ($M)	1314	10.3	3.9	16.9
Bluray Units (M)	1319	0.5	0.2	0.8

with a higher theatrical engagement also have a higher chance to gain better theatrical performance (the coefficient equals 0.831), which reflects the interaction between consumers' demand and retailers' adoption. Movies with higher opening revenue also have higher total box office revenue (the coefficient equals 0.930), suggesting the opening theatrical performance can be a good indicator for the movie's overall sales. Moreover, movies with better theatrical performance also have higher video sales (the coefficient is 0.509 for DVD and 0.712 for Bluray), which is consistent with the nature of sequential distribution. Among these demand variables, I use market shares (of theatrical and home video markets) as the dependent variables for the modeling purpose later.

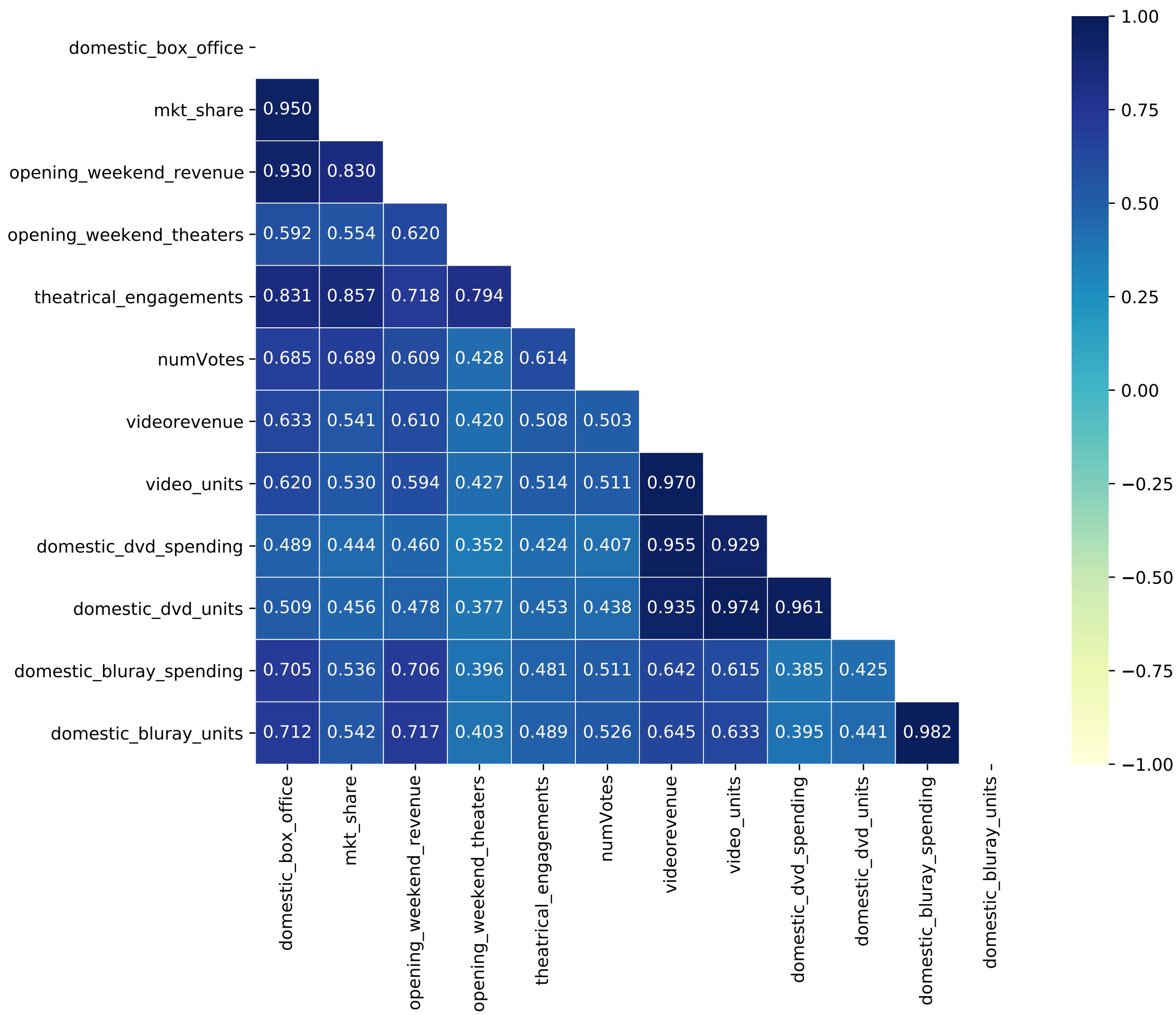


Figure 2.2: Pearson Correlation of Revenue Variables

2.2.2 Movie Characteristics

The movie feature data are collected from `IMDB.com`. It covers 8977 titles domestically released between January 1, 1997 and December 31, 2019. Movie features are in general time-invariant and mostly categorical; they are the movie characteristics that the audience can observe *before* purchasing. In this subsection, I introduce several key features that have been frequently adopted in the past literature along with their sample statistics.

Sequel Most of the previous studies found a significant and positive effect of sequels on movie performance despite using different estimation methods and data samples. In our sample, 11.4% of the titles are sequel movies, and they, on average, have a higher theatrical (upper left) and video (upper right) demand as well as theatrical engagements (bottom left) compared to the non-sequels (Figure 2.3). To explain intuitively, a sequel can attract more theaters to show it in the first place since most likely its original has been tested successful. Plus, moviegoers who like the original would also be more likely to see the sequel. However, given the positive effects on movie sales, we do not observe increasing trends in the number of sequels produced in the industry. Dhar, Sun, and Weinberg (2012 [27]) offered one potential reason that sequels are more costly to make. Our sample shows a consistent pattern (bottom right panel) that sequel movies on average require a higher production budget.

Genre Movies are commonly associated with more-than-one genres, for example, George Lucas' Star Wars (1999) is labeled as Action, Adventure, Fantasy, and Sci-Fi on `IMDB.com`. To simplify the problem and avoid the situation where a film falls into multiple genres, I use only the first genre label listed on `IMDB.com` which is also the closest genre that describes a movie. In this paper, all the movies are categorized into one of the following genres: Action, Adventure, Animation, Biography, Comedy, Crime, Documentary, Drama, Horror, and Other Genres. "Other Genres" represents the following genres jointly: Family, Fantasy, History, Music, Musical, Mystery, Romance, Sci-Fi, Sport, and Thriller. Of our sample wide releases, 29% are Comedies, 24% are Actions and 18% are Dramas, meanwhile, only 0.7% are assigned to "Other Genres" (Figure 2.4).

Genre, as an important set of categorical features, is frequently adopted in past literature, although researchers disagree on what effect it casts on the movie performance. Some do not find (Liu (2006 [57])) movie genres to be informative about theatrical performance; others

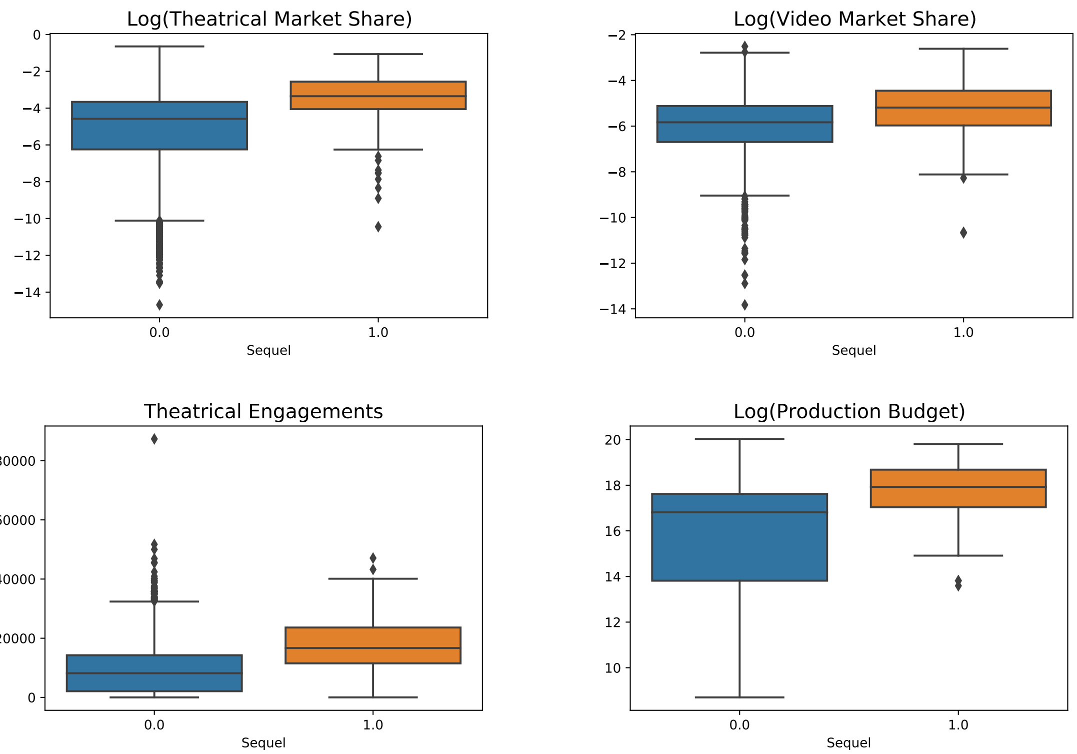


Figure 2.3: Theatrical and Video Demand, Theatrical Engagements, and Budget by Sequel

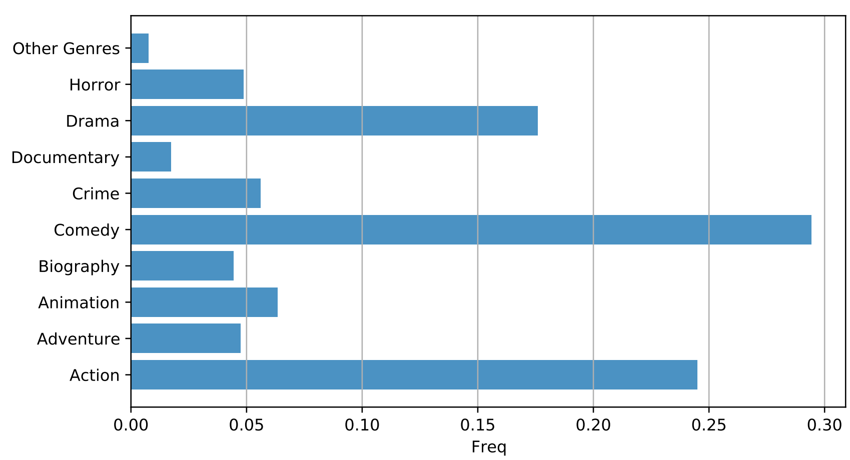


Figure 2.4: Frequency Plot of Genres

find action, adventure, comedy, science fiction, horror have a significant positive effect on the box office revenue. In our sample, Figure 2.5, the boxplot[4] showing the distribution of market share with respect to genres, tells that action, adventure, and animation movies relatively attract more audience (in both theatrical and home video markets) on the median.

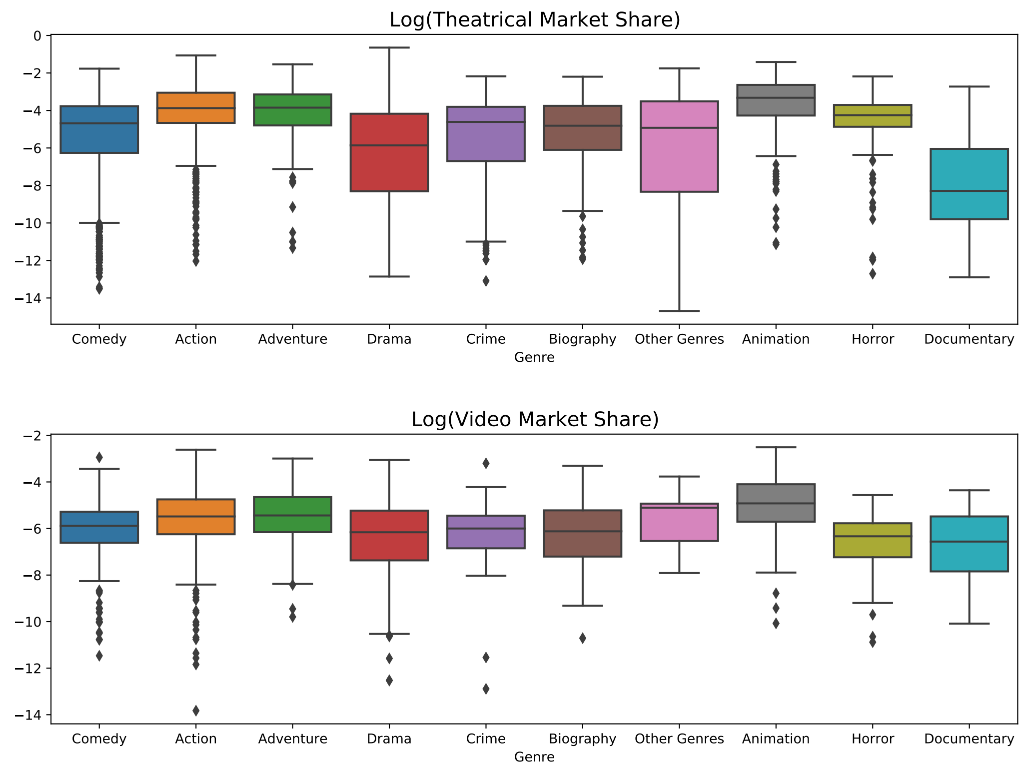


Figure 2.5: Theatrical and Video Demand by Genre

Star, Director, and Writer Star power is another variable that is commonly used in previous studies. Researchers are most interested in the relationship between star power and a film's financial success. Some report that the presence of a star has a positive and significant

[4]The box represents the interquartile range (IQR) with the horizontal line inside showing the median, and the points beyond the maximum and minimum lines are outliers.

effect on theatrical revenue while others find no relationship between the two. Despite the disagreement, star power is usually introduced into regressions as a dummy variable which equals one if any "top star" participates in the movie. However, the definition of "top star" varies. In some studies, "top star" refers to Oscar-nominated best (supporting) actor or actress; or in other cases, it is defined as the stars that have made top ten movies in terms of box office revenues. These two methods may greatly limit the universe of stars (Nelson and Glotfelty (2012 [66])) and other researchers adopt continuous measures of star power. For instance, Elberse (2007 [32]) used the average box office revenue of a star's most recent five movies to measure her commercial success; and the total number of Oscar nominations received before a particular movie to measure her artistic success.

In this paper, the scraped data from `IMDB.com` offer us three names of the leading stars for each movie title. For instance, James Cameron's Titanic (1997) has Leonardo DiCaprio, Kate Winslet, and Billy Zane listed. We certainly can turn each star name into a dummy variable, but it may cause overfitting and dimension problems. Therefore, I count the number of "top stars" for each title rather than adding all the name dummies. Before I define "top star", I first rank the sample stars by their productivity which is measured by the number of films the star has participated in as leading actor or actress. There are in total 14746 stars in our sample and I listed the top 10 productive ones in Table B.1. After ranking, I define "top star" to be the top 10% productive stars. Thus, for each movie title, the maximum possible value of the variable "top star" is 3, indicating that all the three leading actors and actresses are in the top-star list; and the minimum possible value is 0, indicating none of the three leading actors and actresses are in the top-star list. Eventually, 45% of the sample wide releases are with 3 top stars, 26% are with 2 top stars, 15% are with 1 top star, and the rest 14% have no top stars. I also create similar indicators for "top directors" (Table B.2) and "top writers" (Table B.3).

Distributor I categorize the distributors in our sample to be the Major Six and "Other Studios". Although 46% of the sample wide releases are from 'Other Studios" (Figure 2.6), these movies account for only 19% of the total box office revenue and 18% of the total video revenue. Figure 2.7 is the boxplots to show the distribution of market share with respect to distributors. It suggests that wide releases from Non-Major-Six studios have greater variability with a significantly lower median market share in both theatrical and video markets.

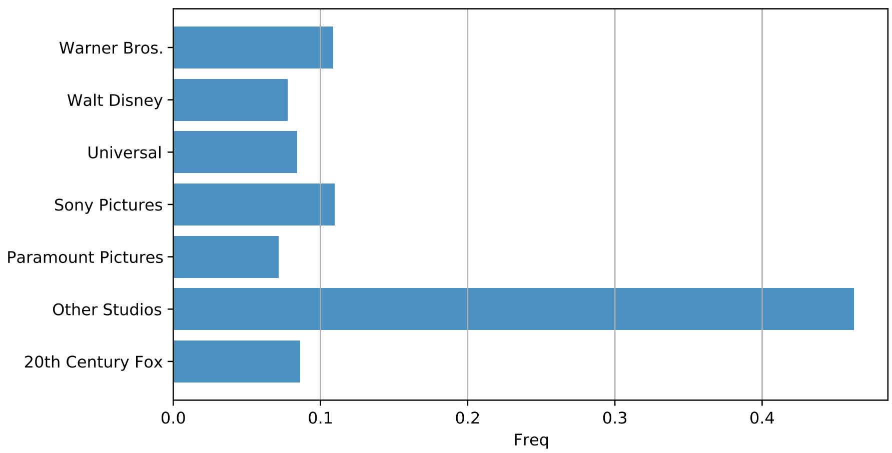


Figure 2.6: Frequency Plot of Distributors

In addition to the features mentioned above, I also include movies' country of origin, language, color info, and runtime in minutes. Among our sample wide releases, 75% of the pictures are originated from the U.S., 91% are in English, and 97% are in color. I also show the sample distribution of runtime in minutes in Figure B.1; the average runtime is 107 minutes and the median is 104 minutes.

2.3 The Structural Model

I introduce the demand structural model in this section. I first employ a logit discrete-choice model as the benchmark. I estimate the theatrical and video markets separately; at this stage, the goal is to understand what movie features matter to consumers and how their utility decays as time passes. Next, I adopt a one-level nested logit discrete-choice model

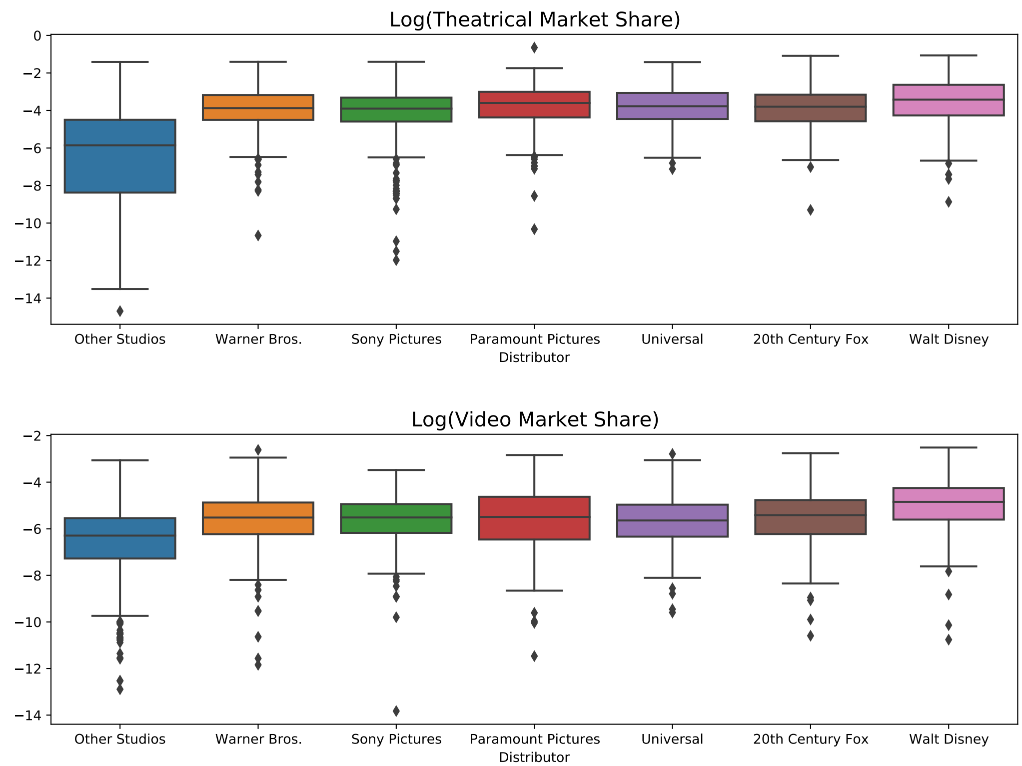


Figure 2.7: Theatrical and Video Demand by Distributor

to estimate the market expansion effect in the two markets. Lastly, to further examine the substitution pattern of the two watching modes (watch in theaters or purchasing videos), I pool all the sample movies together and use a two-level nested logit discrete-choice model.

2.3.1 Logit

Berry (1994 [11]) and Berry, Levinsohn, and Pakes (1995 [10]) developed a method to empirically analyze demand using only product-level and aggregate consumer-level data, which benefit our case of the motion picture industry where individual-level data is not available. In this section, I set up the model following Einav (2007 [29]). Assume we observe $t = 1, \dots, T$ markets, each with $i = 1, \dots, I_t$ consumers. Given that we have weekly revenue data, I define a market t as a week-year combination, with $t = 1, \dots, 722$ for the U.S. theatrical and home video markets respectively from the year 2006 to 2019.

The utility of consumer i from watching movie j in week t is given by

$$\begin{aligned} u_{ijt} &= \delta_{jt} + \epsilon_{ijt} \\ &= X_j\beta + \xi_{jt} + \lambda(t - r_j) + \tau_t + \epsilon_{ijt} \\ &\quad i = 1, \dots, I_t, \quad j = 1, \dots, J, \quad t = 1, \dots, T \end{aligned} \tag{2.1}$$

where δ_{jt} is the mean utility from movie j in week t and ϵ_{ijt} is the consumer-specific shock. The mean utility δ_{jt} depends on observed time-invariant movie characteristics X_j, unobserved (by the researchers) movie characteristics ξ_j, the number of weeks that have passed since the movie's initial release $t - r_j$ where r_j is the initial release week of movie j, note that r_j is not the same for theatrical and video markets given the nature of sequential distribution, and week fixed effect τ_t representing the underlying seasonality in the market. The utility specification does not include the price factor because of the non-price competition in the U.S. motion picture industry where consumers pay the same price for differentiated films in theaters or the video market. Additionally, β is the coefficient of movie characteristics and it indicates which features matter more to the audience. λ is expected to be negative showing how much consumers' utility decays as time passes. The decay happens because consumers enjoy watching the movie as early as possible to be able to lead the market trend and join the word-of-mouth effect.

I define the outside good to be not watching any movies included in our sample and I normalize the utility of all consumers from the outside good to be zero. For simplicity, I further assume that ϵ_{ijt} follows Type 1 extreme value distribution, which allows me to integrate analytically over consumers and obtain a closed-form solution like the following.

$$s_{jt} = \frac{\exp(\delta_{jt})}{\sum_{k=0}^{J_t} \exp(\delta_{kt})}, \quad s_{0t} = \frac{1}{\sum_{k=0}^{J_t} \exp(\delta_{kt})}$$

where s_{jt} is the market share of movie j in week t and J_t is the set of all movies that are available on the market in week t. We compute s_{jt}/s_{0t} to have Equation 2.2 which can be estimated with ordinary least squares (OLS).

$$\begin{aligned}\log(s_{jt}) - \log(s_{0t}) &= \delta_{jt} \\ &= X_j\beta + \lambda(t - r_j) + \tau_t + \xi_{jt} \qquad (2.2)\end{aligned}$$

2.3.2 Nested Logit

The logit model implies that the choice probability between any two alternatives does not depend on a third one (the IIA restriction) and thus, the substitution between products is driven completely by market shares and not by how similar the products are. To resolve the restrictions, I also implement a nested logit Model[5] where the products are grouped into exhaustive and mutually exclusive groups, $g = 0, 1, \ldots, G$. In particular, Equation 2.1 becomes

$$\begin{aligned}u_{ijt} &= \delta_{jt} + \zeta_{it} + (1-\sigma)\epsilon_{ijt} \\ &= X_j\beta + \xi_{jt} + \lambda(t - r_j) + \tau_t + \zeta_{it} + (1-\sigma)\epsilon_{ijt} \qquad (2.3) \\ & i = 1, \ldots, I_t, \quad j = 1, \ldots, J, \quad t = 1, \ldots, T, \quad 0 \le \sigma \le 1\end{aligned}$$

where ϵ_{ijt} follows Type 1 extreme value distribution and $\zeta_{it} + (1-\sigma)\epsilon_{ijt}$ is also an extreme value random variable. Additionally, ζ_{it} is idiosyncratic group preference and σ is a parameter measuring how substitution patterns differ within and across nests; it characterizes the correlation of utilities that a consumer experiences among the products within and across the groups. As σ goes to 1, the within-group correlation of utilities increases, and consumers perceive products of the same group as perfect substitutes relative to products outside the group. As σ goes to zero, the within-group correlation goes to zero, and the model reduces to the standard logit. After solving the math and rearranging (See Berry (1994 [11])), we

[5] For a more detailed discussion of the nested logit Model, see Ben-Akiva and Lerman (1985 Ch.10 [8]).

have:

$$\log s_{jt} - \log s_{0t} = \delta_{jt} + \sigma \log(s_{j|g})$$
$$= X_j\beta + \lambda(t - r_j) + \tau_t + \sigma \log(s_{j|g}) + \xi_{jt} \tag{2.4}$$

where $s_{j|g}$ is the probability that product j is selected condition on its group g being selected. It measures product j's within-group market share which, by construction, is endogenous in the model and instruments will be needed for estimation.

As for designing the nests, I start with following Einav (2007 [29]) and partitioning the market into inside and outside goods (Figure 2.8). I conduct this step for the theatrical and video markets separately. This is for identifying the market expansion effect based on the estimated value of σ in Equation 2.4. If σ is close to 1, then consumers perceive the market films as close substitutes and there is little substitution between the inside and the outside goods, thus no market expansion effect. In other words, consumers will not be driven to purchase in the market even if more films are released in the industry. The closer to 0 the σ gets, the bigger the market expansion is.

Equation 2.3 is referred to as one-level or single-level nested logit. If we add one more level, that is, the products are partitioned into G groups, $g = 0, 1, \ldots, G$, and each group g is further partitioned into H_g subgroups, $h = 1, 2, \ldots, H_g$, then Equation 2.4 is extended to be the following:

$$\log s_{jt} - \log s_{0t} = \delta_{jt} + \sigma_1 \log(s_{j|h}) + \sigma_2 \log(s_{h|g})$$
$$= X_j\beta + \lambda(t - r_j) + \tau_t + \sigma_1 \log(s_{j|h}) + \sigma_2 \log(s_{h|g}) + \xi_{jt} \tag{2.5}$$

where $s_{j|h}$ is the selection probability of product j condition on its subgroup h being selected, that is, product j's within-subgroup market share; $s_{h|g}$ is the probability that product j's subgroup h is selected condition on its group g being selected, that is, subgroup h's within-group market share. Both $s_{j|h}$ and $s_{h|g}$ are endogenous and we need at least two instruments for estimation.

I adopt the two-level nested logit to check the correlation of utilities that a consumer experiences across the two watching modes, I pool all the sample films together, treat the two markets as a whole, and further segment the movies by the watching mode into theatrical and video nests. In this way, we have two levels of nests shown in Figure 2.8 and the task is to estimate Equation 2.5.

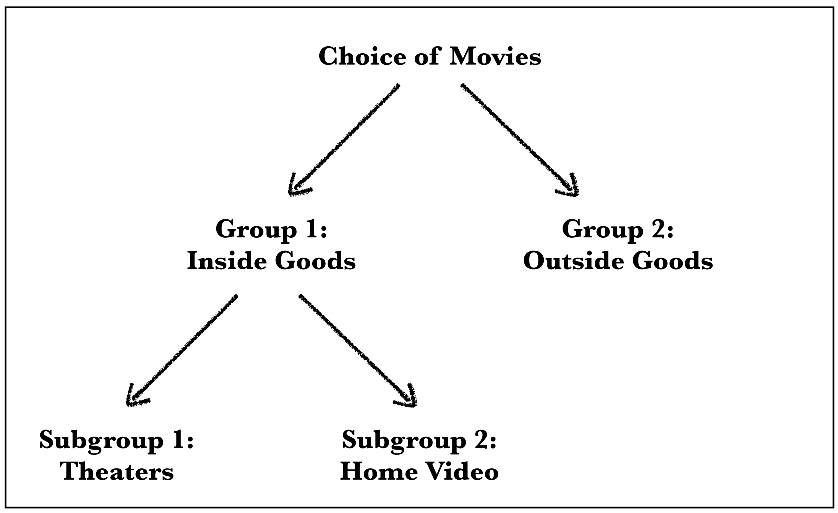


Figure 2.8: Nested Logit

McFadden (1978, pp. 86-87) proves that Equation 2.5 is consistent with the maximization of a random utility function if $1 \geq \sigma_1 \geq \sigma_2 \geq 0$. The interpretation of this condition originates from the idea that the variance of the random utilities is the smallest at the lowest level of the nest tree (or intuitively the alternatives within the subset are closer substitutes), and it cannot decrease as we move from a low to a higher level. In other words, the heterogeneity of preferences at the higher level (groups) is at least as great as at the lower level (subgroups)[6].

In Equation 2.5, the parameter σ_1 captures the correlation of utilities that consumers experience among products in the same subgroup, and similarly, the parameter σ_2 captures

[6]See Anderson, S. P., De Palma, A., & Thisse, J. F. (1992, Ch 2.7.1). Discrete choice theory of product differentiation. MIT press [2].

the correlation of utilities that consumers experience among subgroups in the same group. The model collapses to a one-level nested logit with groups as nests if $\sigma_1 = \sigma_2$, in other words, the second-level partition of subgroups is not necessary. In our case, $\sigma_1 = \sigma_2$ means consumers do not distinguish between watching movies in theaters or on video at home. The model collapses to a one-level nested logit with subgroups as nests if $\sigma_2 = 0$[7], that is, the market should be directly partitioned into theatrical goods, video goods, and outside goods.

2.3.3 Instruments

Now I discuss the choice of instruments for the endogenous variables $s_{j|g}$ in Equation 2.4 and $s_{j|h}$, $s_{h|g}$ in Equation 2.5. The ideal instruments should be correlated with the within-group or within-subgroup market share, but not correlated with the error term. I use two instruments here. One is the number of movies in a given group. For instance, if we partition the market to be inside and outside goods, then the instrument for $s_{j|g}$ would be the number of movies in the market. More movies available in the group will lead to more competition, and therefore, it should be negatively related to the within-group market share $s_{j|g}$. Plus, the number of movies in the market does not directly affect the demand if we control the movie-fixed effect. Thus, considering these two points, the number of movies can be a good choice. However, one may argue that higher-demand seasons will attract more movies to be released, then the weekly dummies in the model can alleviate such concern (Einav (2007 [29])).

The other set of instruments I adopt is the average characteristics of all the rivals within the same group, including movies' rating score, number of top stars, and freshness (how long the movie has been introduced to the market). This choice follows Berry (1994 [11]) and Berry, Levinsohn and Pakes (1995 [10]). The rating score (from IMDB) and the number of top stars are both indicators of how attractive the movie appears, capturing the intensity of competition from the rivals. In general, consumers prefer movies with higher ratings and

[7]Also see *Notes on the Nested Logit Demand Model* by Ryan Mansley, Nathan Miller, Conor Ryan, and Matt Weinberg.

more top casting. If the same-group rivals of movie j, on average, have higher rating scores or more top stars, the within-group market share of movie j should be lower. Thus, these two instruments should be negatively related to the endogenous variables. Similarly, consumers prefer new releases. Hence, if the rivals of movie j are, on average, newer in the market, the within-group market share of j is expected to be lower. We use "the number of weeks passed from the initial release" to measure the freshness, and therefore, the instruments should be positively related to the endogenous variables. Lastly, to validate this set of instruments, we also need to check if they are uncorrelated with the error term, and the rivals' characteristics should not affect the own demand.

2.4 Results

2.4.1 Parameter Estimates

In this subsection, I present the parameter estimates of our demand models. I first estimate the theatrical and home video markets separately. The key results are shown in Table 2.2 and 2.3[8]. Then, I pool two markets together and estimate them jointly. The key results are displayed in Table 2.4. Each table includes the results of both logit and one-level nested logit models with and without the movie fixed effects. I partition the market into inside and outside goods when using a one-level nested logit model (Figure 2.8) and further partition the inside goods into theatrical goods and video market goods when using a two-level nested logit model.

Utility Decay, $\boldsymbol{\lambda}$: The results of the benchmark logit model (Column 1 of Table 2.2 and 2.3) show that the estimated decay coefficients, λ's, are -0.317 for the theatrical market and -0.075 for the video market. Modeling with the one-level nested logit (Column 3 of Table 2.2 and 2.3), the decay effects become smaller in both markets and the estimated λ's now are -0.147 for the theatrical market and -0.024 for the video market. Consumers' utility decays

[8]The full results are shown in Table B.4 and B.5 in the Appendices.

Table 2.2: Demand Estimation: Theatrical (2006 - 2019)

	Logit		**One-Level Nested Logit**	
	(1)	(2)	(3)	(4)
	OLS	OLS	IV-2SLS	IV-2SLS
weeks in release	-0.317***	-0.407***	-0.147***	-0.204***
	(-39.54)	(-52.01)	(-4.04)	(-7.23)
log($s_{j\|g}$)			0.538***	0.474***
			(16.89)	(6.75)
Observations	22320	22320	22320	22320
Adjusted R^2	0.580	0.792		
Movie Fixed Effect	No	Yes	No	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.3: Demand Estimation: Home Video (2006 - 2019)

	Logit		**One-Level Nested Logit**	
	(1)	(2)	(3)	(4)
	OLS	OLS	IV-2SLS	IV-2SLS
weeks in release	-0.0750***	-0.0727***	-0.0237	-0.0183***
	(-28.70)	(-55.80)	(-1.60)	(-4.04)
log($s_{j\|g}$)			0.686***	0.751***
			(3.29)	(12.22)
Observations	69790	69790	69790	69790
Adjusted R^2	0.494	0.627		
Movie Fixed Effect	No	Yes	No	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

faster in the theatrical window, which can also be reflected by Figure 3.3 that theatrical demand is more concentrated in the opening weeks than the video demand. Intuitively, the earlier the window is in the sequential distribution channel, the higher the decay rate would be; the theatrical window is the initial window and it is expected to associate with the highest decay rate.

Equation 2.2 and 2.4 are based on the assumption of linear utility decay. To complement, I employ an alternative demand model where I add a quadratic term $\lambda'(t - r_j)^2$ to allow for non-linear utility decay. Table B.6 and B.7 show the results of the alternative demand models for theatrical and home video markets. The values of the utility decay rates in the alternative demand models, λ, decrease from -0.317 to -0.593 for the theatrical market and from -0.075 to -0.221 for the home video market. The coefficients of the squared "weeks in release", $(t - r_j)^2$, are positive and significant in both markets. The negative λ combined with positive λ' suggest that consumers' utility decays faster in the beginning, then slows down towards the end of the runs.

Movie Features, $\boldsymbol{\beta}$: Next, Column 1 and 3 of Table B.4 and B.5 in the Appendices allow us to answer which movie features matter to the audience. Most of the estimates are significantly different from zero and the signs of the estimates are consistent with our intuition. For example, in both theatrical and video markets, sequel movies attract more consumers than non-sequel ones; movies in English sell better than non-English ones; movies with more top stars/directors/writers are significantly more popular. To examine the coefficients on genre dummies, I set "action movies" to be the baseline group. The results show that adventure and animation movies are significantly more attractive compared to the baseline while documentaries are significantly less favored in both markets. Lastly, by including dummies on studios, we find that wide releases from Walt Disney significantly attract more audience compared to the rest of the Major Six.

Substitution between Inside and Outside Goods (Market Expansion), $\boldsymbol{\sigma}$: Another key finding from the one-level nested logit model is the substitution pattern between the inside and outside goods represented by the coefficient σ in Equation 2.4. Einav (2007 [29]) called it the market expansion effect since how consumers substitute between watching market movies and pursuing outside options can indicate if the market is expanding. By studying the market expansion effect, we can estimate whether more or better movies becoming available in the market can draw the audience from other movies or options other than watching movies. Column 3 of Table 2.2 shows that σ equals 0.538 for the theatrical market. As mentioned in the previous section, if σ is close to 1, consumers perceive the market films as close substitutes and there is little substitution between the inside and the outside good. The closer to 0 the σ gets, the bigger the market expansion effect is. In our case, $\sigma = 0.538$ suggests that consumers' substitution patterns differ significantly in the choice among different theatrical movies and the choice whether to go to theaters[9]. Similarly, σ equals 0.686 (Column 3 of Table 2.3) for the home video market and the market expansion effect is smaller compared to the theatrical market. Furthermore, when I pool two markets together, σ decreases to 0.157 (Column 1 of Table 2.4). Given that we are enlarging the market scope, it is reasonable to have a bigger market expansion effect.

Weekly Dummy Variables, $\boldsymbol{\tau}$: The estimated coefficients on the weekly dummy variables (both with and without the movie fixed effects) are plotted in Figure 2.9 and 2.10. The dashed lines stand for deviations of 1.96 standard errors. Compared to the seasonal trends in the theatrical and video markets presented in Figure 3.4 and 3.5, the estimated seasonal pattern successfully captures the two peak seasons in the theatrical market: Summer and Christmas periods, and one peak season in the home video market: Thanksgiving-Christmas holiday. Especially, the estimates from the nested logit model fit the shape more accurately.

[9] In Einav (2007 [29]), the market expansion effect σ was estimated to be 0.577.

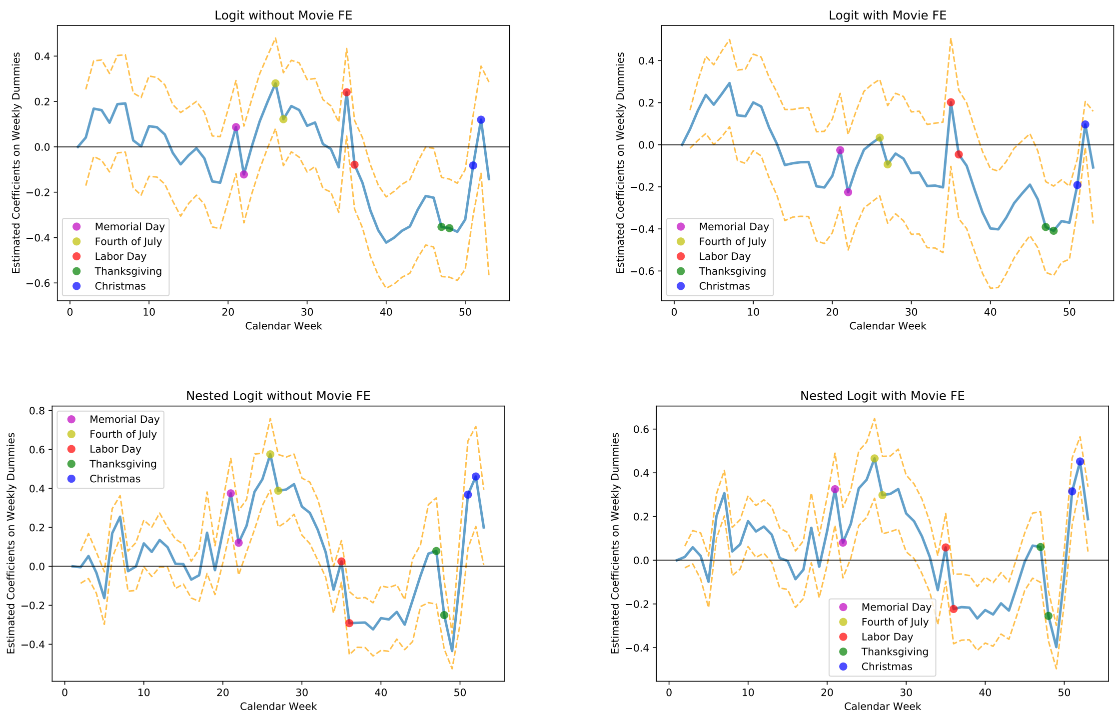


Figure 2.9: Estimated Coefficients on Weekly Dummies: Theatrical

Substitution between Watching Movies at Theaters and Purchasing Home Videos, $\boldsymbol{\sigma_1}$ & $\boldsymbol{\sigma_2}$: Lastly, I pool all the sample movies from the theatrical and home video markets together and estimate the demand using the two-level nested logit model as Equation 2.5. I design the nest structure as shown in Figure 2.8. This step allows me to conclude how consumers distinguish between watching movies at theaters and purchasing home videos. The main results are shown in Column 3 to 5 of Table 2.4. First, we have mentioned previously that if $\sigma_1 = \sigma_2$ holds, then the subgroup level partition is redundant and consumers do not distinguish between watching movies in theaters or on video at home. Now based on the estimated results, that the two coefficients are tested to be significantly different from each other confirms the opposite. Second, if $\sigma_2 = 0$, the group level partition should be abandoned. And the estimated σ_2 equals 0.531, indicating our nest structure is valid. Third, our results are consistent with the maximization of a random utility function since the coefficient

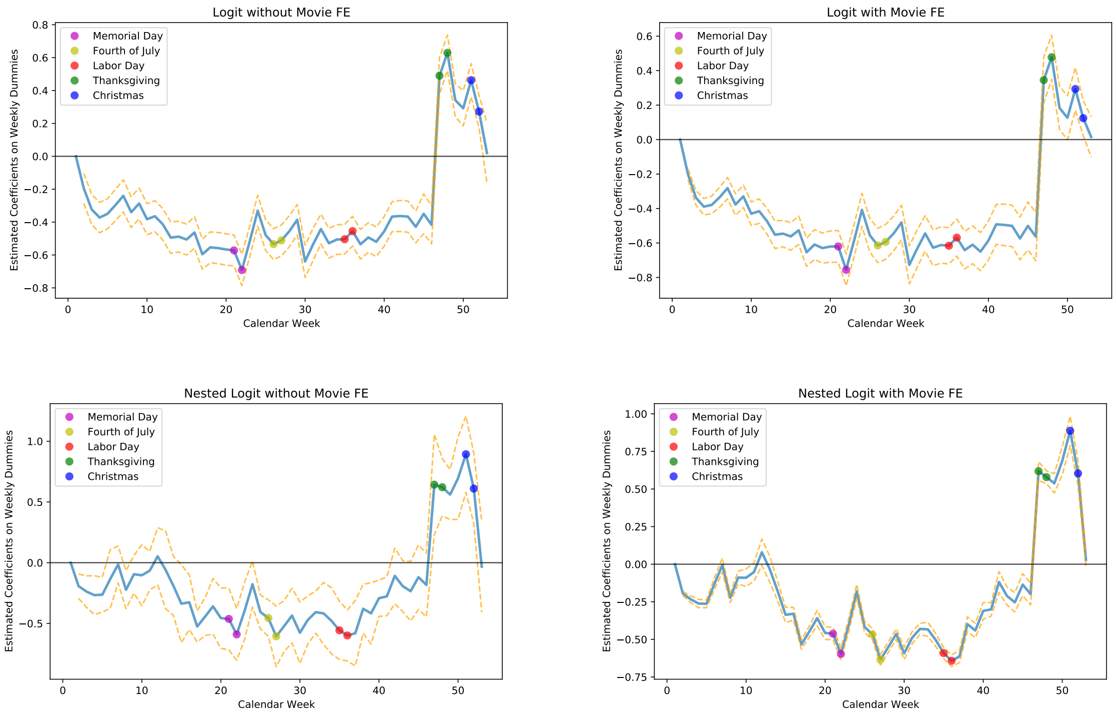


Figure 2.10: Estimated Coefficients on Weekly Dummies: Home Video

of $\boldsymbol{\log(s_{j|h})}$ ($\sigma_1 = 0.587$) is greater than the one of $\boldsymbol{\log(s_{h|g})}$ ($\sigma_2 = 0.531$) and both of them are between 0 and 1. This, again, confirms that our nest structure makes sense because the movies in the sub-nests are closer substitutes than the ones in the higher nest level.

2.4.2 Model Fit

In this section, I briefly evaluate the model fit for the demand estimation. Figure 2.11 and 2.12 present the observed and predicted distributions of the weekly market shares in theatrical and video markets respectively. The predicted distributions are graphed based on both logit and nested logit models. Our model prediction is reasonably good at the average level; the mean values of the observed and predicted log(weekly market share) overlap with each other. The distributional fit looks better under the nested logit model, especially for the theatrical market that the nested logit can even capture the “bimodal-shape” roughly.

Table 2.4: Demand Estimation: Theatrical and Video Pooled (2006 - 2019)

	One-Level Nested Logit		**Two-Level Nested Logit**		
	(1)	(2)	(1)	(2)	(3)
	IV-2SLS	IV-2SLS	IV-2SLS	IV-2SLS	Iterated GMM
weeks in release	-0.0715***	-0.0803***	-0.0350***	-0.0281***	-0.0368***
	(-13.49)	(-4.32)	(-13.02)	(-5.59)	(-13.12)
log($s_{j\|g}$)	0.157*	0.0965			
	(2.51)	(0.45)			
log($s_{j\|h}$)			0.587***	0.698***	0.566***
			(18.39)	(12.14)	(16.98)
log($s_{h\|g}$)			0.531***	0.597***	0.520***
			(19.28)	(9.28)	(18.27)
Observations	92110	92110	92110	92110	92110
Adjusted R^2	0.722		0.928		0.922
Movei Fixed Effect	No	Yes	No	Yes	No
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

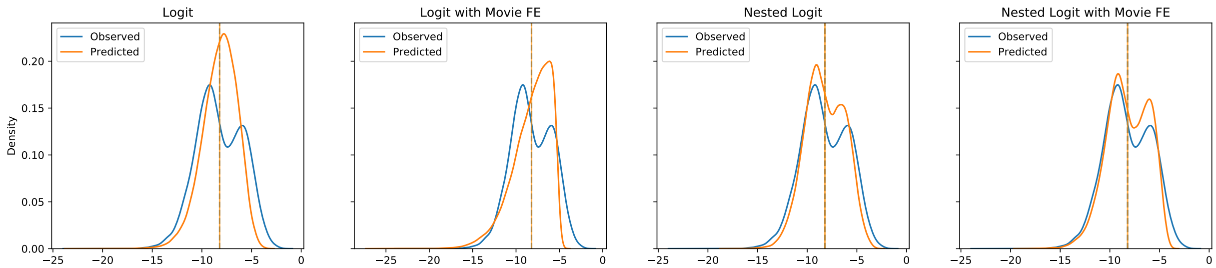


Figure 2.11: Model Fit (Theatrical): Log(Weekly Market Share)

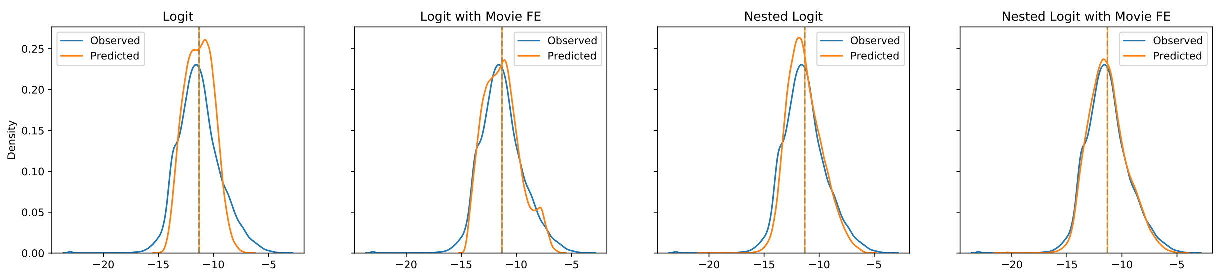


Figure 2.12: Model Fit (Home Video): Log(Weekly Market Share)

Figure 2.13 shows the observed and predicted distributions of the total market share (aggregated through weeks and sequential markets) of the sample movies. The predicted total market share tends to be smaller than the actual one, but the gap shrinks under the nested logit model.

2.5 Conclusion

This paper aims to fill in the gaps in the economics literature on home video market research. By conducting demand estimation for theatrical and home video markets together, I investigate how consumers' purchasing behavior differs across the two markets and how much the audience distinguishes between the two watching modes.

I start with a logit discrete-choice model to estimate the demand separately for the theatrical and home video markets. First, the results show that the theatrical market where the movies are first released is associated with a higher utility decay rate (-0.317) compared

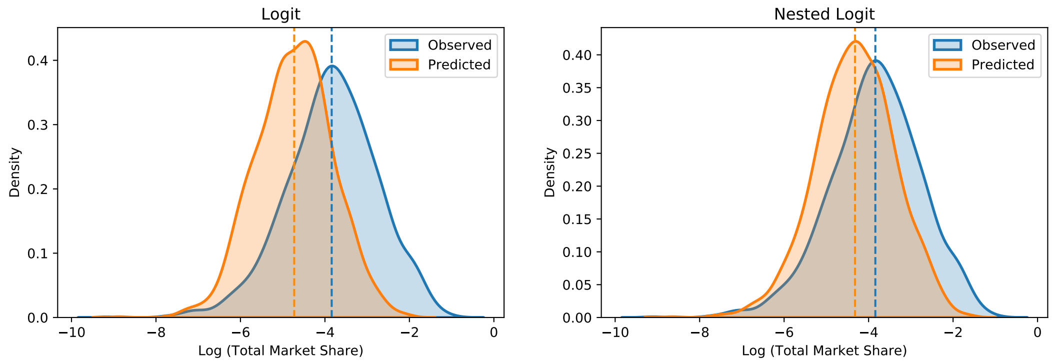


Figure 2.13: Total Market Share in Sequential Distribution

to the video market (-0.075) where the movies are sequentially released. This pattern is consistent with the nature of sequential distribution in the movie industry. Moreover, by allowing non-linear utility decay and adding a quadratic term of movies' age, we can confirm that, in both markets, the consumers' utility decays faster in the beginning, then slow down towards the end of the movie runs. Second, the logit model allows us to quantify movie qualities and identify the features that have a significant impact on demand. For example, sequels, movies in English, or movies with more top stars/directors/writers can attract more consumers.

To relax the IIA restriction implied by the logit model, I next partition each market into inside and outside goods and adopt a one-level nested logit to estimate the market expansion effect. The coefficient equals 0.538 for the theatrical market and 0.686 for the video market, which means that, in both markets, consumers' substitution patterns differ significantly in the choice among different market movies and the choice of whether to watch movies or not. If we add more movies into the markets, we can potentially drive more purchases from the consumers. Additionally, the one-level nested logit model allows us to estimate the seasonality in demand. The estimated coefficients of weekly dummies successfully capture the peak seasons in the two markets.

To study how consumers distinguish between the two watching modes, watching at theaters or purchasing home videos, I pool all the sample movies together and further partition the insides goods into theatrical goods and video market goods. Using a two-level nested logit model, I validate that the nest structure I adopt is consistent with McFadden's theory on the maximization of a random utility function and consumers do distinguish between watching movies in theaters or on video at home.

Chapter 3

OPTIMAL RELEASE TIME IN SEQUENTIAL DISTRIBUTION: THE WINDOWING STRATEGY OF THE U.S. MOTION PICTURE INDUSTRY

3.1 Introduction

> The goals of distribution in the film industry are generally to maximize value and revenue streams over the life of a property. One of the primary and most profitable ways this maximization is achieved is through ***windowing*** – the strategy of releasing a film in different venues and on different platforms over a period of time, usually with discreet periods of exclusivity and variable pricing.
>
> – Holt and Sanson (2013 [47])

The U.S. motion picture industry has undergone waves of technology shocks over the past decades, from the introduction of DVDs in 1997 to the rise of streaming giants more recently, which tremendously affects distributors' release timing decisions. In those years, the industry was committed to prioritizing the theatrical market and left at least six months for the theatrical window before the products enter any sequential markets. Up until now, the length of the theatrical window has dropped 50%, from 24 weeks in 2000 to 12 weeks in 2018 (Figure 3.6), and the streaming firms fasten the trends advocating the practice of skipping theatrical screening directly. Debates over the issue of shrinking theatrical windows and early streaming or digital releases have never stopped. For example, the movie *Roma* by Alfonso Cuarón sponsored by Netflix earning 10 nominations of the 91st Academy Awards 2019 was banned from theatrical exhibitors' annual Best-Picture Oscars showcases since it broke the rule of the traditional 90-day window and stayed in theaters for only three weeks. Steven Spielberg, a member of the Academy's board of governors and one of the most powerful people in Hollywood, said that films that forgo a theater run or have a limited theatrical release should not be in contention for Oscars while Netflix claims “These things are not mutually exclusive”.

This paper is motivated by the controversial practice in the industry and aims to demystify how distributors optimize release time decisions under sequential distribution. What factors may impact the series of release time decisions if we rationalize distributors’ choices with profit-maximization assumption? I tackle the problem through two stages. I first adopt the logit and nested logit discrete-choice models for demand estimation. I quantify movie qualities, consumers’ utility decay rates, seasonality in demand, market expansion effect, and consumers’ substitution pattern between the two different watching modes[1]. This stage is designed to understand consumers’ behavior from distributors’ perspectives before making the release time choices. Next, I model distributors’ windowing strategy as a one-shot sequential-move game with incomplete information. I take the demand estimates from the first stage as given and follow the method of pseudo-backward induction proposed by Einav (2009 [30]) to further estimate two coefficients in the game model: the weekly variable cost of theatrical distribution and distributors’ weight on the theatrical window. After solving the simplified two-player game, I conduct counterfactual analyses where I hypothetically change the value of one of the key coefficients holding other factors constant to investigate how the release time decisions may change. I found that higher cost of theatrical distribution, lower weights on the theatrical window, poorer movie quality with weak opening performance, or faster theatrical utility decay can all potentially rationalize the shrinking theatrical window.

The study of release time choices in sequential distribution applies to more than just the motion picture industry. It is frequently observed that many products evolve through generations and versions. Films are re-released on videos for home consumption after theatrical runs; paperback versions of books are released after about a year of their hardback release (Wilson and Norton (1989 [87])); fashion houses introduce designs sequentially into lower

[1]The detailed modeling and estimation results are presented in Chapter 2.

price outlets after every fashion season (Pesendorfer (1995 [68])). The most essential characteristic of sequential distribution is that the windows are complements but also substitutes to each other at the same time. One stage of the distribution system may provide a good indication of sales in the next stage of the sequential distribution channel (Lehmann and Weinberg (2000 [52])), meanwhile, early opening of the sequential channel may crowd out the demand from the previous one. The trade-off makes release time decisions crucial.

This paper complements and contributes to a few branches of the economics literature. First, it builds on the prior research related to timing decisions under sequential distribution in the motion picture industry. Frank (1994 [39]) and Lehmann and Weinberg (2000 [52]) both modeled theatrical and video revenue as linear functions of time and derived a closed-form solution for determining the optimal time to open the sequential video window. They both theoretically validated the fact that the growing video market of the last decade was a good reason for the industry to shorten the theatrical window. Lehmann and Weinberg (2000 [52]) further suggested that the total profits would increase if movies were released to video sooner than the current practice. Prasad, Bronnenberg, and Mahajan (2004 [70]) also drew a similar conclusion that attempts to maximize profits with respect to the video introduction time would force distributors to release the video earlier and earlier till the video is ultimately released simultaneously with the theatrical version. In addition, they pointed out that the determinants of entry time include the discounting of future profits, the foresight of the firm, customers' expectations, and the possibility of cannibalization.

Until this point in time, most of the research in this literature area adopts reduced forms and offers insights based on closed-form expressions; little empirical evidence is offered. More recent research has improved these two points. In the working paper from Luan and Sudhir (2006 [59]), the authors built a structural demand model for forward-looking consumers with rational, adaptive expectations who dynamically seek to optimize the utilities inter-temporally, and found that given the consumer preferences and industry cost structure, the theater-to-DVD window that maximizes the revenue is about 2.5 months for an average movie instead of the current industrial practice, 5.5 months. Berbeglia, Derdenger, and Tayur

(2020 [9]) (working paper) developed a dynamic discrete choice model for action movies in theaters and home videos. The authors found that shortening the theater-to-video window increases the video demand by increasing freshness and the advertising spillover from the theatrical run. August, Dao, and Shin (2015 [3]) also looked at the consumer-side with a normative approach and limited the attention on the impact of congestion externalities[2]. They showed that during seasons of peak congestion, day-and-date strategies[3] are optimal for high-quality films. Moreover, lower congestion effects provide studios with incentives to delay the video release. To summarize, most of the structural models in this branch are developed from consumers' perspective. The researchers agree that shrinking the theater-to-video window is optimal given consumers' forward-looking purchasing behavior. My paper adds to this emerging body of literature by modeling from the supply side. Using a strategic timing game, I emphasize the distributors' competition effect which is barely discussed in the previous studies.

Second, my paper is also related to topics regarding modeling and estimating strategic games where firms compete with endogenous product-type choices. The competition can be on various business decisions, including choosing prices, promotion, placement, positioning, time of entry, and so on. The structure of the game depends on what is being modeled: 1) decisions may be discrete or continuous, 2) payoffs may be observed or latent, 3) information may be complete or incomplete, and 4) the game may be one-shot or continuing for many periods. When solving the game, economists are generally concerned with the possibility of having multiple equilibria (Bresnahan and Reiss (1991 [13])), and four main approaches proposed in the previous literature are 1) looking at an aggregate level (e.g. estimating the number of entrants instead of who enters), 2) modeling sequential moves instead of simultaneous ones, 3) providing a selection rule to select amongst many equilibria, and 4) employing

[2]Anticipating sold-out screenings, longer waits at ticketing, poorer seat choices in auditoriums, and other undesirable crowd externalities (e.g., crying babies, noise from conversations, and temperature problems), some consumers may prefer to delay their viewing.

[3]A day-and-date strategy typically means that a product is released across two or more distinct channels on the same day.

a method that can handle non-uniqueness (Ellickson and Misra (2011 [37])). Another major concern in the literature of game estimation is the computational burden, which is eased by introducing incomplete information and players no longer observe everything about their rivals. In this paper, I follow these insights and construct a sequential-move timing game with incomplete information. Here I list several papers that are most relevant to mine.

Krider and Weinberg (1998 [51]) used a game-theoretic model to analyze the high-season release timing of two motion pictures with different drawing power that are competing directly for the same target audience. They conducted an equilibrium analysis of the product introduction timing game and concluded that the weaker movie may be unable to commit to the desirable early release date and forced to delay opening. They emphasized the rationale of "avoiding the competition". Mazzeo (2002 [62]), on the other hand, predicted that firms will voluntarily offer products unlike those of their competitors since the negative effect that a competitor has on firm payoffs is up to twice as large if that competitor is the same product type. Seim (2006 [77]) focused on the strategic importance of product positioning within a market. The author applied a simultaneous entry game of incomplete information to geographic location choices in the video retail industry and found that firms use spatial differentiation to shield themselves from the competition. Similarly, Einav (2010 [30]) studied the choice of location in time-space. Einav developed an empirical model for discrete sequential games with private information and applied it to the release date timing game played by movie distributors. The game was solved by a pseudo-backward induction algorithm and the results suggested that theatrical release dates of movies are too clustered around big holiday weekends. I follow Einav (2010 [30]) and expand the dimension of the choice set to handle the setup of sequential distribution. I also discuss wider portfolio strategies when distributors' payoff function is not only based on the individual product.

The remaining paper is organized as follows. Section 3.2 introduces the sequencing problem and explains why the release time decision matters to the industry. Section 3.3 provides the industry background including the traditional distribution practices and the recent trend. Section 3.4 describes the data sources and descriptive statistics. Section 3.5 presents the model. Section 3.6 discusses the results and the interpretation. Section 3.7 presents the counterfactual analyses. Section 3.8 concludes.

3.2 Sequencing Problem and Windowing Strategy

Intellectual property rights are infinitely divisible, and distributing a film is the art of maximizing consumption and corresponding revenues across exploitation options over the life of a property (Ulin (2013 [84])). One of the primary and most profitable ways this maximization is achieved is through *windowing*, the strategy of releasing a film in different venues and on different platforms over a period of time, usually with discrete periods of exclusivity and variable pricing (Holt and Sanson (2013 [47])). From Vogel (2020 [86]), "films are normally first distributed to the market that generates the highest marginal revenue over the least amount of time. They then "cascade" in order of mariginal-revenue contribution down to markets that return the lowest revenue per unit time."

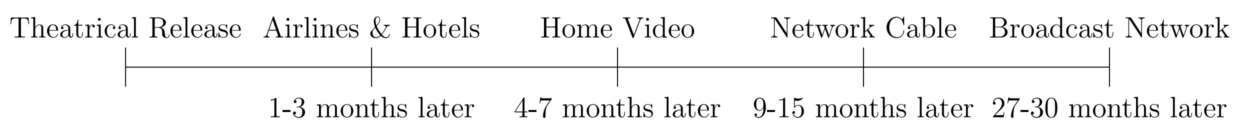


Figure 3.1: Traditional Release Pattern for a Studio Film in the U.S.

Historically, the sequence of the windows starts from theatrical (Figure 3.1). The distributor writes contracts with exhibitors, the owners of the theaters, on a movie-by-movie and theater-by-theater basis. Most commonly, the theatrical window lasts for 1 to 3 months. When the theatrical run ends, a "black" or rest period may follow in which the film wouldn't be available in any market except for on airlines and in hotels. The home video window, including EST, DVD, Blu-ray, VHS, LaserDisc, and VOD, would open approximately 4 to 7 months later. For major studio films, the *same* distributor again writes contracts with video retailers who would buy units of movies from the studio distributor, and then rent or sell the cassettes out to customers. The pay television window comes next and it would be 9 to 15 months after the initial theatrical release. It is also possible for a film to be on free television or broadcast network, but that generally happens 27 to 30 months after the initial theatrical release.

In this paper, I use theatrical and home video windows to study the release time decision in sequencing distribution taking the variable pricing as given, which fits the current industry setting where the inflation-adjusted movie ticket price is quite stable over the years (Figure C.1) and the unit price of one DVD or Blu-ray fluctuate within the range of $15 to $20 (excluding the special edition releases for collection purpose). Thus, the window structure in our setting can be reduced to Figure 3.2. For each movie, the distributor needs to decide when to release the film into the theatrical market (r_1) and when to exit the theatrical market and enter the sequential home video one (r_2).

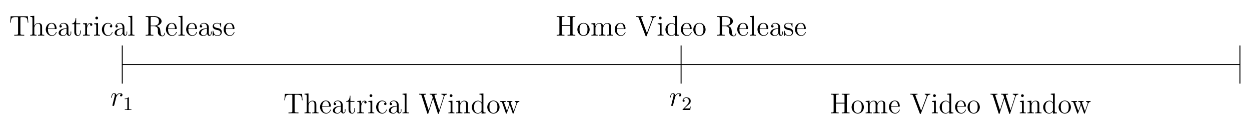


Figure 3.2: A Theatrical - Video Window

Why does the release time decision matter in the motion picture industry? Time matters due to the fact that the audience cannot wait. Once the movie is released, no matter in theaters, on video, or streaming, the highest demand usually appears within the first week; after that, the weekly demand will decline significantly as time passes. The opening week performance somehow determines or predicts the overall one and the revenue curve commonly has a longtail. Figure 3.3 shows that the theatrical revenue of the first week after the release reaches almost 40% of the eventual revenue. Moreover, 80% of the theatrical revenue is earned within the first month after the release; similarly, 40% of the video revenue is earned within the first month after the video release date. Therefore, release time is such a crucial decision faced by the distributors; they aim for entering the market at the correct time point to extract as much profit as possible since soon later the chance for the movie to be successful gets smaller and smaller. Additionally, high demand at the first window (theatrical) can create information and word-of-mouth effect among the consumers; it can decide the performance of sequential window (home video) and induce more repeat consumption.

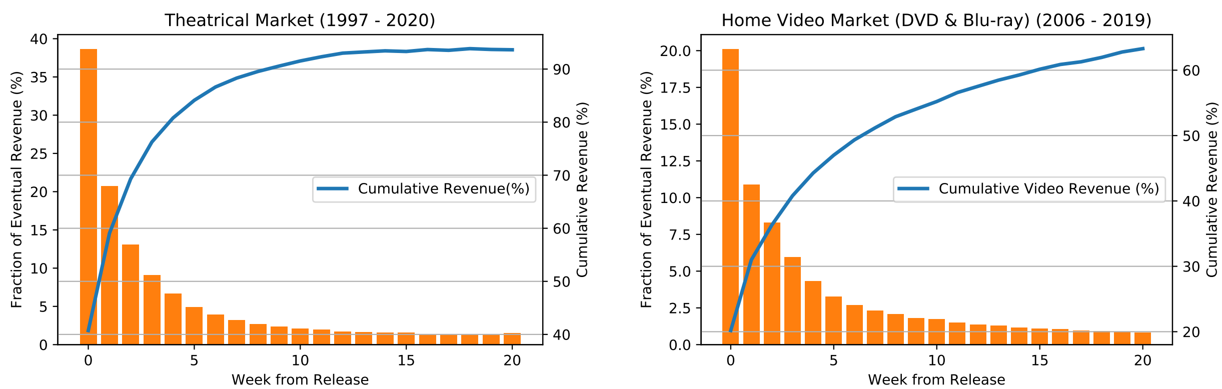


Figure 3.3: Movie Revenue Lifecycle

3.3 Industry Background

The U.S. Motion Picture industry (Hollywood) is considered to be the oldest film industry in the world, in the sense of being the place where the earliest film studios and production companies emerged[4]. Over the years, the industry, in general, keeps expanding, producing more products (left panel of Figure C.2) and generating higher revenue (right panel of Figure C.2). In this section, I introduce several aspects of the motion pictures industry including 1) its history and development 2) how a movie typically is made through three stages of production, distribution, and exhibition, 3) how concentrated the market is and which major studios are dominating, 4) how theatrical and home video distribution works in practice, and 5) how much the theatrical window has been shrunk over the years.

[4]Also see `https://en.wikipedia.org/wiki/Cinema_of_the_United_States`

3.3.1 History and Development

By the early 1890s, Thomas Edison, together with his main assistant William Dickson invented the "Kinetoscope", the first camera that was capable of photographing objects in motion, which marked the beginning of the film industry[5]. Soon after the invention of the "Kinetoscope", the technological evolution of cameras, films, and projection equipment accelerated, and the movie business, therefore, became lucrative attracting entrepreneurs all over the world. To standardize the market, the Motion Picture Patents Company (MPPC), also known as the "Trust", was founded in 1908 by ten major studios. They often engaged in aggressive business tactics and monopolistic control to eliminate the competition in the market. Eventutally, the Trust dissolved in 1917 due to the increasing market power of the independent studios which later integrated both vertically and horizontally into Hollywood's giants.

By the 1930s, Hollywood thrived to dominate the world cinema taking the advantage of the well-developed industrial organization and the diversified immigrant cultures (See Trumpbour (2002) [83]). A "Golden Age" of cinema came during the late 1930s and early 1940s; it was a time of unparalleled success for the motion picture industry, during which two-thirds of Americans were attending the theater at least once a week. The most influential studios by that time were all vertically integrated, that is, they controlled every part of the system as it related to their films, production, distribution, and exhibition, all of the three crucial stages to operate in the market. Those companies were Warner Brothers., RKO, Twentieth Century Fox, Paramount, and MGM. Their overwhelming control over the market caused complaints from the harmed. The leading studios were charged with illegally conspiring to restrain trade and the U.S. Department of Justice eventually stepped in. By 1948, ten years after the initial suit was filed, the defendants finally signed the decree to disintegrate and let go the control over their theater chains. Since then, production and distribution run separately with the exhibition. Moreover, a series of antitrust actions against various segments of

[5] This subsection is partially adapted from Vogel, H. L. (2020). Entertainment industry economics: A guide for financial analysis. Cambridge University Press [86].

the industry followed (including violations of the consent decree, price fixing, block booking[6], product splitting[7], and other anti-competitive activities), which further changed the market structure and regulated the industry. Now, contracts between distributors and exhibitors are commonly written on a movie-by-movie and theater-by-theater basis; even theaters owned by the same exhibitor write separate contracts for the same movie (Einav (2007 [29])).

Apart from the federal antitrust action, the advent and proliferation of television in the late 1940s, the rise of the home video market during the 1980s to 1990s, and the emergence of streaming services in the 2010s are the other major forces that affect the traditional theatrical movie business. Unquestionably, technology improvement has brought challenges to the industry (competing against other forms of entertainment), as well as new opportunities (more efficient ways to distribute and easier ways to watch).

3.3.2 Production, Distribution, and Exhibition

The three major players in the motion pictures industry are producers, distributors, and exhibitors. Producers are in charge of generating ideas, composing scripts (with the help of writers), and screening and identifying the most promising projects. Once the film is initiated, the producer will then seek external funding, which usually can be from a film studio. After the anticipated budget is allocated, the project enters the production phase where the actual shooting of the movie takes place. After the shooting has been completed, the movie is assembled during the post-production stage. During this phase, visual effects, music, sounds, and voice-overs are added; the raw film footage is edited and reduced into the final movie.

Once the movie has been completed, it will be handed over to distributors. Distributors normally would first design the marketing campaigns for the targeted audience of the movie. They will align their releases with the most demographically suitable theaters, subject to the screen availability and previously established relationships with the theater chains. The

---

[6] A practice in which an exhibitor who wanted any of a distributor's pictures had to take all of them.

[7] A practice in which distributors or exhibitors explicitly cooperated in splitting the market.

final release strategy (including the release dates, locations, number of screens, etc.) can provide the best possible marketing mix for the picture.

When the negotiation is settled, distributors and exhibitors are ready to sign the contracts[8]. In practice, the exact formulas for calculating films' revenue flows attributable to the different parties involved are the result of negotiation and thus can vary from project to project (Hofmann (2012 [46])). Traditional contractual arrangements between the distributor and the exhibitor commonly include a minimum playing time and an agreement as to how the box-office receipts are to be shared[9]. A sliding-scale agreement with an after house allowance ("nut") split and a guaranteed minimum ("floor") is frequently used (Eliashberg, Elberse, and Leenders (2006 [34])). The house allowance includes the expenses incurred by the exhibitors running the facility (e.g., rent, insurance, and maintenance). The split is typically in favor of the distributor but tends to decline as the movie's run proceeds so that the exhibitor has the incentive to keep showing the movie. The specific values of the split, floor, and the rate at which they slide are determined by the relative bargaining power of the two parties, for example, independent distributors claim much lower percentages than majors or mini-majors.

After the movie is released, distributors and exhibitors use box office reports to match supply to demand over time, i.e. supply adaptation. If the movie has performed well in the early weeks, more screens may be added and new exhibitors may reach out to book the film. On the other hand, if the movie receives poor box office revenue, exhibitors can drop the film or even end its entire theatrical run in order to cut their losses, although the signed contracts may limit such adjustment.

---

[8]In the industry, movies are normally first distributed to the market that generates the highest marginal revenue over the least amount of time, then "cascade" in the order of marginal-revenue contribution down to markets that return the lowest revenue per unit time (Vogel, H. L. (2020 [86])). This has historically meant theatrical market first, followed by home video, network cable, broadcast network, etc. Therefore, the distributor(s) of a movie probably would sign contracts with multiple media platforms. In this subsection, I focus on the contracting between movie distributors and *theatrical exhibitors*, which is also the most studied in the literature.

[9]A traditional way of splitting the box office revenue between the distributor and the theatrical exhibitor has long been 90 - 10 split in favor of the distributor after the exhibitor's expenses are deducted.

3.3.3 The Major Six

Table 3.1 shows that the U.S. motion picture industry is quite concentrated over the recent years and dominated by six major studios that have integrated production and distribution arms. They are Walt Disney Pictures, Warner Bros., Universal Pictures, Paramount Pictures, 20th Century Fox, and Sony Pictures, however, the rank of the top studios varies from year to year[10]. From 1995 to 2020, the Major Six have occupied an average of 77.52% of the U.S. theatrical market share. Additionally, the market is filled with a number of independent producers who do not have integrated production and distribution. They use either the major studios' distribution division or one of the independent distributors in the market.

Table 3.1: Industry Concentration (1995 - 2019)

	Theatrical Market Share (%)	Home Video Market Share (%)
Walt Disney	16.94	16.37
Warner Bros.	15.21	14.20
Sony Pictures	12.28	11.68
Universal	11.72	14.48
20th Century Fox	11.03	13.53
Paramount Pictures	10.34	9.15
C6	77.52	79.41

Data Source: `the-numbers.com/market/distributors`

[10] One note to add here, Disney purchased 20th Century Fox through its acquisition in 2019 and renamed the studio as 20th Century Studios. In the same year, Netflix became the newest member of the Motion Picture Association, joining Disney, Paramount, Sony Pictures, Universal, and Warner Bros. as the "new" Major Six. The addition of Netflix reflected the association's increased focus on streaming services as the industry adapted to the viewing practices of modern audiences.

3.3.4 Theatrical Distribution

In the U.S. movie market, films are released on a weekly basis, mostly on Fridays. Hence, for a given year, there are fifty-three slots open to choose from. But distributors, in general, do not choose one out of the fifty-three, instead, they would first target a specific season they prefer, for example, "Summer 2022" or "Late Summer 2022". This is because certain genres align with certain seasons. For instance, "Oscars Season" often starts in November. Prestige films are often released around that time to generate critical attention and publicity so that they stay fresh in the minds of award voters. Blockbusters used to aim for Memorial Day to take advantage of the beginning of summers when teenagers are out of school and ready for a break. In recent years, big-hit movies, like films from the Marvel Studios, are being released earlier and earlier and push the summer release season to begin in early May. The months between Oscar Season and summer (February to April) are considered "dumping grounds" for low-quality films. Films released during this period are generally thought of as films that studios have to release because of their contracts but do not expect a return on.

The seasonal release pattern is consistent with the seasonal demand in the market. The left panel of Figure 3.4 shows the theatrical market share, averaged from 1997 to 2019[11]. The market share is calculated using the number of weekly tickets sold, normalized by the U.S. population. It can be thought of as the per capita number of movies seen each week, for example, a market share of 0.1 implies that 10% of the U.S. population goes to the movies in the corresponding week. The curve depicts two strong periods as mentioned in the seasonal release pattern, one is the summertime from Memorial Day to Labor Day, and the other is the Oscars Season from Thanksgiving to Christmas.

After the distributor pins down the release season for a given title, she would later announce the exact release week within the season. When doing so, she is wary of the competition effect from rivals in the market. The competition effect can be from other products released by the same distributor, which is easy to tackle. The distributor can strategically arrange

[11]Figure 3.4 follows the first panel of Figure 2 in Einav (2007). The dashed lines stand for deviations of 1.96 standard errors.

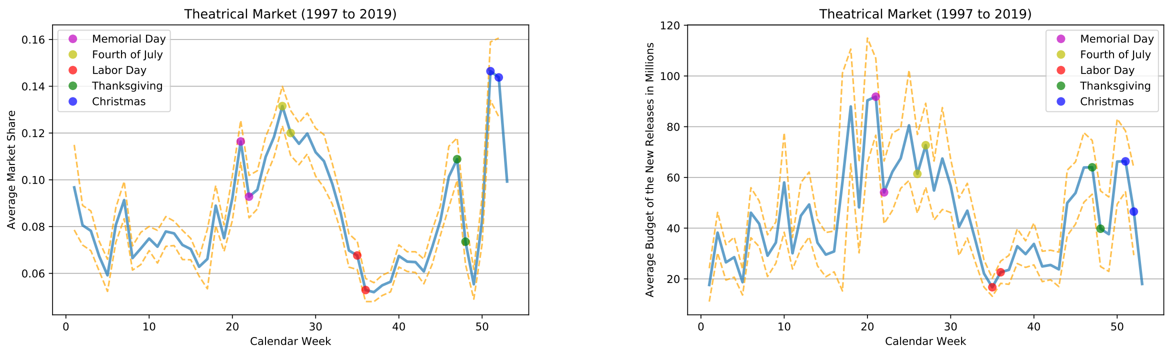


Figure 3.4: Seasonal Demand and Competition in Theatrical Window

the release time for a portfolio of films instead of looking at them one by one individually. More complicatedly, the competition is brought from products released by other players in the market. In general, distributors would like to avoid direct competition with movies with similar target audiences, unless in an attempt to crush the competition and sustain an upper hand in the market. One of the commonly practiced strategies is to announce the movie's release date early on in the hope that this preemption will result in other distributors avoiding the announced release date (Einav (2009 [30])). This strategy is often adopted for blockbusters from the major studios, for instance, James Cameron's Avatar 2 and Avatar 3 distributed by 20th Century Studios (the subsidiary of Walt Disney after the acquisition) have already been announced to be released on December 16th, 2022 and December 20th, 2024[12]. Big-hit films distributed by major studios tend to have a high production budget, which does not guarantee the movie quality but does somehow allow them to have high bargaining power and early-mover advantage in the industry. The right panel of Figure 3.4 shows the budget of theatrical releases for given weeks of the year, averaged from 1997 to 2019. It emphasizes that the two popular seasons are also filled with high-budget movies.

Lastly, even if the release dates are set, distributors often change them in response to new

[12]It is the year 2020 at this point.

information concerning the release dates of similar movies. The timing game of the theatrical stage alone thus can be approximated by a repeated announcement game. Here I use Josh Boone's The New Mutants, which is an extreme example, to illustrate this point (Table 3.2), although most of the observed changes in practice do not shift the original release date by more than a month[13].

Table 3.2: The New Mutants: A Timeline of the Release Schedule

Report Date	Release Schedule Changes
Apr 22nd, 2017	The New Mutants was originally set to be released on Apr 13th, 2018.
Jan 11th, 2018	First delay to Feb 22nd, 2019, to avoid David Leitch's Deadpool 2 from the same distributor
Mar 26th, 2018	Second delay to Aug 2nd, 2019, to avoid Simon Kinberg's Dark Phoenix from the same distributor
Mid-2018	The New Mutants will undergo reshoots.
May 7th, 2019	Third delay to April 3rd, 2020 due to screen rights shift after Disney's acquisition of 21st Century Fox
Mar 9th, 2020	The New Mutants is delayed indefinitely due to the COVID-19 pandemic.
May 13th, 2020	The New Mutants gets its fifth release date: Aug 28th, 2020.
Aug 28th, 2020	The New Mutants opens in theaters and drive-ins.

3.3.5 Home Video Distribution

The home video sector can be divided into physical and digital; the former includes rental and purchase of DVD, Blu-ray, VHS, LaserDisc., and the latter includes Electronic Sell-Through

[13]Einav (2009 [30]) found that over 60% of his sample movies changed their release dates at least once, but 75% of the observed changes do not shift the release date by more than a month.

(EST), Video-on-Demand (VOD), and paid subscription streaming. Despite the fact that the majority of movie revenues are earned in the Home Entertainment sector (digital and physical) rather than the box office (Figure C.3), not much attention has been paid to the releasing strategy of the home video market. Unlike the theatrical market where consumers only have a short period of time to attend, the home video window is a continuous one where products rarely truly exit the market. This allows the video market to be associated with mild variations in demand and competition throughout the whole year. From Figure 3.5, we observe flat demand in most of the year and one peak during the Christmas season; as for the competition pattern, we do not see a clear pattern indicating high-budget movies clustering in any particular season. From the industrial perspective, the video release time matters at a different level on which distributors care more about the relationship between the two sequential windows rather than merely capture the seasonality of the market. On one hand, theatrical and home video markets are complementary in the way that the theatrical performance and the fresh market sensation generated will largely benefit the video window and induce repeat consumption. On the other hand, the two markets are competitive with each other; releasing the video version too early will sabotage the box office revenue, despite the early arrived return on investment.

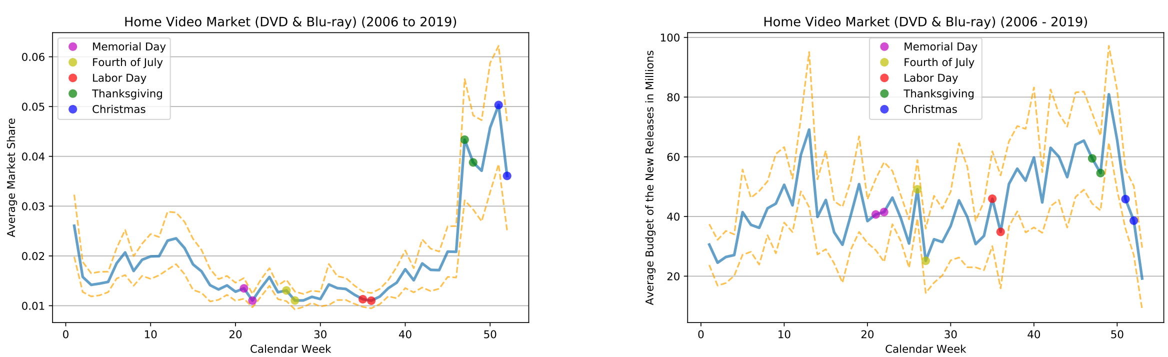


Figure 3.5: Seasonal Demand and Competition in Home Video Window

In practice, the video release date is often announced after the theatrical run, but the waiting time of the announcement varies case by case. I use data from The National Association of Theatre Owners (NATO) to offer an overview of the waiting time of video release announcements in Table 3.3. The sample covers 3511 titles released during the year 2000 to the year 2019, of which 2742 are distributed by major studios. Averaging the sample titles across all major distributors, video release announcements are made in the 11th week and videos are out in the 19th week after the theatrical release.

Table 3.3: Announcement & Video Release Overview (Major Studios 2000 - 2019)

Distributor	Number of Titles Released	Number of Weeks before Announcement	Number of Weeks before Video Release
Sony	697	10.9	18.3
Fox	456	10.2	17.8
Universal	448	9.2	16.6
Warner Bros.	414	10.3	18.1
Disney	402	12.2	21.9
Paramount	325	11.3	18.8
Total	2742	10.6	18.5

3.3.6 The Shrinking Theatrical Window

Leaving the 6-month window between theatrical and post-theatrical releases used to be an unwritten rule in the U.S. motion picture industry. However, that films are built upon large amounts of capital invested has pressured distributors to open all windows earlier for faster returns. Additionally, digitalization has sped up this process; from physical disc to electronic sell-through (two to four weeks earlier than the traditional disc version since the early 2010s), now to the streaming services (skipping theatrical release became possible), distributors have more ways to design the windows and let go the long-lasting tradition of protecting theatrical window. This can be reflected by Figure 3.6, depicting the average number of weeks between theatrical release and video release. The theatrical window length has been decreasing from 24 weeks in the early 2000 to 12 weeks more recently.

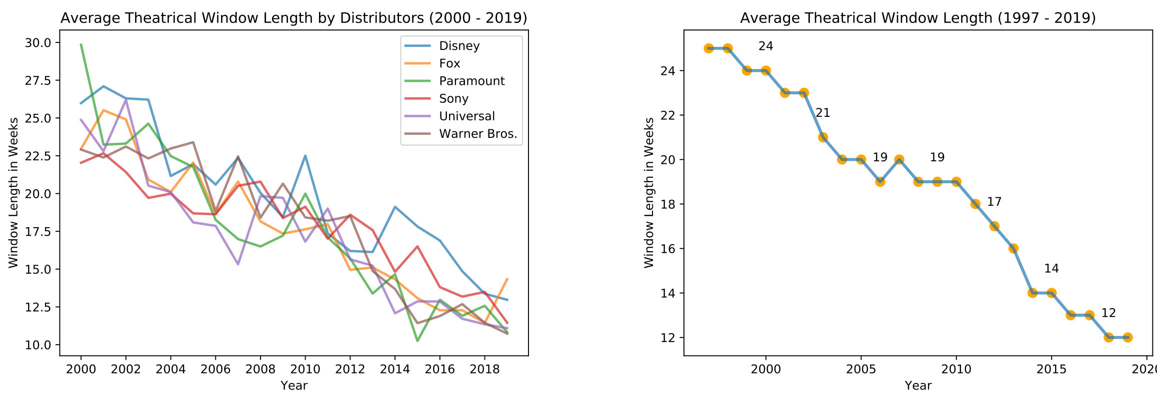


Figure 3.6: Average Theatrical Window Length in the U.S.

Todd Wanger, the co-owner of 2929 Entertainment, who supported the first experiment on the "day-and-date" releasing strategy (i.e. simultaneous release) have once publicly cast doubts on the inefficiencies in the market: why spend marketing money twice for theatrical and video markets separately? why let those who cannot afford to take the family to theaters or dislike watching films in theaters wait several months to see it on videos? why not adopt the "day-and-date" release strategy (Ulin (2013 [84]))? Fifteen years after, Ted Sarandos, Netflix's Chief Content Officer, pointed out the same idea but with a leading position in the industry:

> The current model for distribution of movies is pretty antiquated — waiting ten months or more for home video exploitation...We want to accelerate the model and do it day and date in theaters and on Netflix and we will fund the movies to make it work.

3.4 Data

The data for this paper consists of two parts: the panel data on weekly revenue of theatrical and home video markets and distributors' observed strategic timing choices. In this section, I first use the revenue data to conduct the reduced form analysis regarding the video release timing decision. Next, I offer detailed descriptive statistics for the timing data.

3.4.1 Reduced Form Analysis

As Figure 3.3 shows, the revenue declines exponentially as time goes, thus in this section, I fit the demand using an exponential function following Lehmann & Weinberg (2000)[52]. I derive the optimal condition for choosing the window length $(r_2 - r_1)$ and also present the reduced form results.

The theatrical demand $D^c(t, r_2)$ (c stands for cinema) starts from m_1^c at $t = 0$ and then declines at the rate of m_2^c. Since the movie exits the theatrical market before its video release date at $t = r_2$, the theatrical demand is not identified after r_2.

$$D^c(t, r_2) = m_1^c e^{-m_2^c t} \quad \text{where } t \in [0, r_2) \text{ and } m_2^c > 0 \tag{3.1}$$

The video demand opens from $m_1^v(r_2)$ and then declines at the rate of m_{2A}^v. $m_1^v(r_2)$ is modeled as a function of video release time r_2, more specifically, $m_1^v(r_2) = m_1^v e^{-m_{2B}^v r_2}$ where m_1^v is the opening video demand if the video is released simultaneously in two markets ($t = r_2 = 0$) and m_{2B}^v represents the video's implicit decay rate before the actual release date. Additionally, we have $\frac{dm_1^v(r_2)}{dr_2} < 0$, that is, the longer the video release is delayed, the lower the opening video demand. This happens because as time passes, consumers either have already watched the movie in theaters or forgotten to purchase due to the declining marketing effect.

$$\begin{aligned} D^v(t, r_2) &= m_1^v(r_2) e^{-m_{2A}^v (t - r_2)} \\ &= m_1^v e^{-m_{2B}^v r_2} e^{-m_{2A}^v (t - r_2)} \quad \text{where } t \in [r_2, \infty) \end{aligned} \tag{3.2}$$

Eventually, the objective function is defined by summing up the demand from two markets.

$$\pi(r_2) = \int_{t=0}^{r_2} D^c(t, r_2)dt + \int_{t=r_2}^{\infty} D^v(t, r_2)dt$$
$$= \frac{m_1^c}{m_2^c} - \frac{m_1^c}{m_2^c} e^{-m_2^c r_2} + \frac{m_1^v}{m_{2A}^v} e^{-m_{2B}^v r_2} \tag{3.3}$$

We choose the optimal r_2^* to maximize the cumulative demand; mathematically, taking the derivative of $\pi(r_2)$ with respect to r_2, setting it equal to 0, and solving gives us the following optimal condition.

$$r_2^* = \frac{\log(\frac{m_1^c}{m_1^v}) + \log(\frac{m_{2A}^v}{m_{2B}^v})}{m_2^c - m_{2B}^v} \tag{3.4}$$

First, $\frac{m_1^c}{m_1^v}$ represents the ratio of opening demand of theatrical and video markets. The more demand the theatrical market can generate relative to the video market, the greater the optimal delay in the video's release. Meanwhile, m_1^c is expected to be positively correlated with m_1^v since movies with strong theatrical opening should also have strong video opening. Second, the more rapid the video's decay rate m_{2A}^v after the video release date is relative to the video's implicit decay rate m_{2B}^v before the video release date, the greater the optimal delay in the video's release. Lastly, the more the theatrical decay rate exceeds the implicit video decay rate $m_2^c - m_{2B}^v$, the earlier the video should be released.

Now I present the reduced form results of Equation 3.1 and 3.2. Taking log for both Equation 3.1 and 3.2, we have the following linear regressions.

$$\log(D^c(t, r_2)) = \log(m_1^c) - m_2^c t$$
$$\log(D^v(t, r_2)) = \log(m_1^v) + (m_{2A}^v - m_{2B}^v) r_2 - m_{2A}^v t$$

The market demand (share) is calculated using the number of weekly tickets or units sold, normalized by the U.S. population. I pool all the sample movies together and estimate the regressions using the OLS method. Table 3.4 shows that all the decay rates, m_2^c, m_{2B}^v and m_{2A}^v, are positive as expected and the optimal video release time is 21 weeks after the theatrical release. The reduced form helps to understand the shape and the trend of the sample demand data, but it oversimplifies distributors' decision process, which will be solved using structural models.

Table 3.4: Reduced Form Results

Coefficients	Interpretation	Estimates OLS	Estimates OLS
$\log(m_1^c)$	Theatrical Opening Power	-7.5662***	-7.5856***
		(-315.75)	(-297.29)
m_2^c	Theatrical Revenue Decay Rate	0.2011***	0.1987***
		(-85.71)	(-79.66)
$\log(m_1^v)$	Implicit Video Opening Power	-10.1498***	-10.3906***
		(-522.15)	(-457.33)
m_{2B}^v	Implicit Video Revenue Decay Rate	0.1099***	0.0934***
		(-60.26)	(-28.89)
m_{2A}^v	Video Revenue Decay Rate	0.0559***	0.0611***
		(-137.98)	(-129.25)
Year Fixed Effect		Yes	No
Optimal r_2^*		20.9326	22.6226

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The word "implicit" represents the case of simultaneous release.

3.4.2 Timing Data

Before taking a look at the release dates, I begin this section by introducing three different theatrical release patterns: wide release, limited release, and exclusive release. In our dataset, a movie released in 600 theaters or more is considered a wide release. Wide releases are the most common with the major studios; the movie screening availability goes nationwide with full capacity from day one. A limited release is any release in less than 600 theaters, which is often adopted by low-budget films targeting several cities rather than the whole country. And lastly, an exclusive release is a release in a very limited number of theaters (often one in Los Angeles and one in New York) to create word-of-mouth prior to the nationwide release of the movie, and the expansion from the exclusive release into a wide one may come one to four weeks after. If a movie starts with an exclusive release and then expands to a wide release, I use the date of expand-to-wide as its actual theatrical release date and I label it as a wide release. For example, Toy Story 2 (1999) first had an exclusive release on November 19th, 1999. It expanded to the wide release one week after on November 26th, 1999. I record the second date, November 26th, 1999, as its observed theatrical release date. Therefore, all the sample movies are labeled as either wide releases or limited releases.

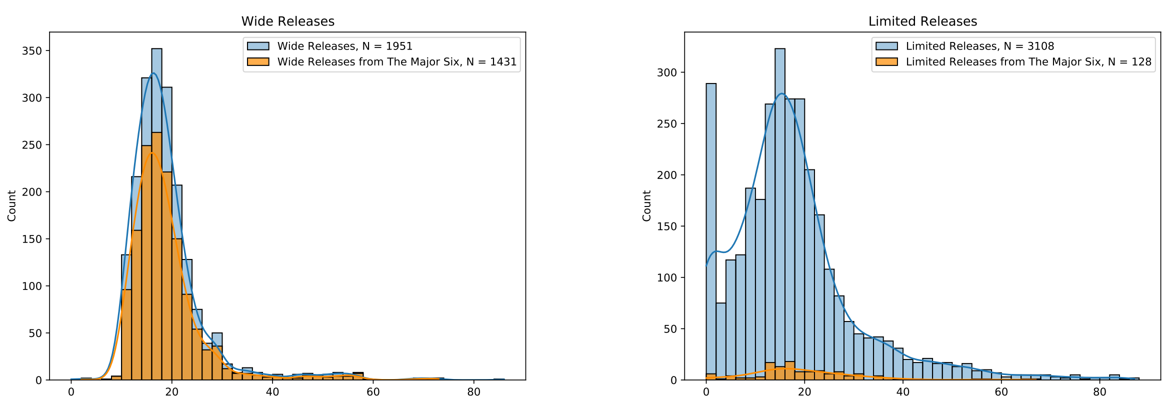


Figure 3.7: Sample Distribution of the Theatrical Window Length: $r_2 - r_1$

The theatrical timing data cover 13103 titles domestically released in theaters between 1/1/1997 and 12/31/2019, 6674 of which have the video release information. Excluding the ones that were handled by different distributors in the two markets, 5059 titles are left, 39% of which are wide releases and the rest 61% are limited releases. Figure 3.7 and Table 3.5 show the sample distribution and summary statistics of the theatrical window length defined as the number of weeks between the theatrical and video release dates ($r_2 - r_1$). Among the wide releases, 73% are distributed by the Major Six with 19 weeks as the mean and 17 weeks as the median. Meanwhile, only 4% of the limited releases are distributed by the Major Six. Non-major studios often choose limited releases due to the low budget.

Table 3.5: Summary Statistics of the Theatrical Window Length: $r_2 - r_1$

	All Sample Wide Releases	Major Six Wide Releases	All Sample Limited Releases	Major Six Limited Releases
count	1951	1431	3108	128
mean	18.8	18.7	17.4	19.7
std	7.9	7.7	13.0	11.5
min	0.0	2.0	0.0	0.0
25%	14.0	14.0	9.0	13.0
50%	17.0	17.0	15.0	17.0
75%	21.0	21.0	22.0	25.0
max	86.0	71.0	87.0	67.0

Additionally, Figure 3.8 is to show distributors' observed strategic timing choices. I create a two-dimensional space (r_1, r_2) to represent distributors' choice set. Given the initial (theatrical) release year of a certain title, the player (distributor) chooses two week numbers (r_1, r_2) to release the movie sequentially. I model the choices discretely, so both r_1 and r_2

are natural numbers. The theatrical release week r_1 is within 1 to 53 since we have at most 53 weeks in any given year. The video release week r_2 is greater or equal to r_1 because of the nature of sequential distribution. For example, Christopher Nolan's The Dark Knight (2008) distributed by Warner Bros. was released on July 18th, the 29th week of 2008, thus $r_1 = 29$. Its initial DVD release was on December 9th, the 50th week of the same year, therefore $r_2 = 50$, and eventually, $(29, 50)$ would be its observed equilibrium. Another example, Monsters, Inc. (2001) from Walt Disney was released to theaters on November 2nd, the 44th week of 2001, so $r_1 = 44$. Its DVD version was out on September 17th of 2002, the 38th week of 2002 or 91st $(53 + 38)$ week if we count from the beginning of 2001, and thus $r_2 = 91$. Finally, $(44, 91)$ is the observed equilibrium for Monsters, Inc. Like I mentioned in the previous section, in practice, the distributor often aims at a particular theatrical release season instead of a year for a given title, hence the choice space can be further reduced, which is helpful in the estimation stage.

3.5 Model

I model distributors' release time decisions in two stages. In the first stage, each player predicts the market shares of her movie based on the given characteristics, such as storylines, genres, star power, and so on. In the second stage, all the players engage in a strategic release time game with one another to maximize the potential aggregated market shares of their own portfolios of films. I further assume that the release time decision of all the sequential markets are made simultaneously, that is, the time to enter the theatrical market r_1 and the home video market r_2 are decided at the same time by the distributors.

The first stage of the movie demand estimation is presented in Chapter 2. I now take the demand estimates as given and use them as inputs to the game estimation. I follow the method of *pseudo-backward induction* proposed by Einav (2009 [30]) to empirically solve the release time game.

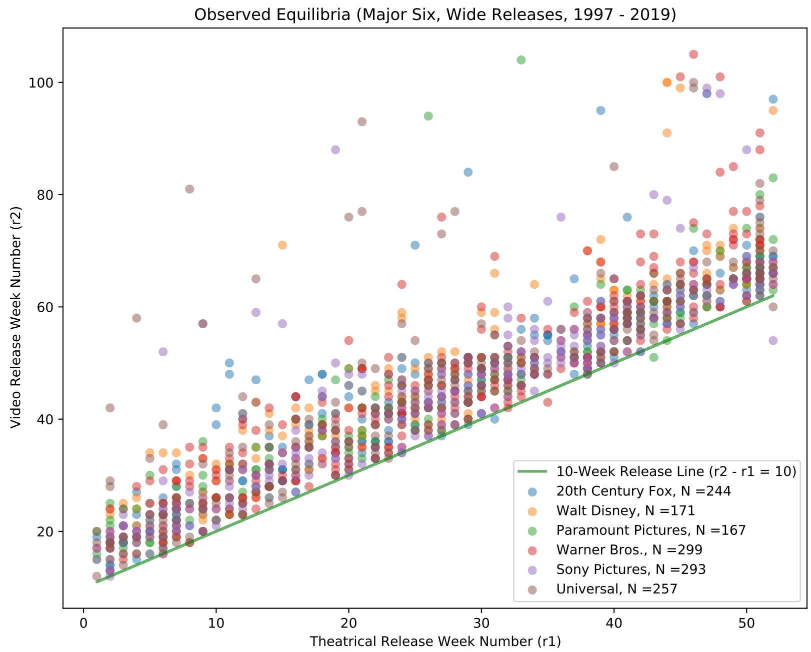


Figure 3.8: Observed Equilibria of (r_1, r_2): The Major Six (Wide Releases 1997 - 2019)

3.5.1 Strategic Timing Game

Consider N players (distributors) interacting with each other by making choices on movie release dates in the market m indexed as $i = 1, \ldots, N_m$. Each of the player i in market m has a portfolio of films to release indexed as $j = 1, \ldots, J_i^m$. The discrete and finite action space for player i distributing movie j in market m is denoted as A_{ij}^m. Thus, the action space for player i in market m can be written as (for simplicity of notation, the m subscripts are suppressed) $\prod_{j=1,2,\ldots,J_i} A_{ij} = A_{i1} \times A_{i2} \times \cdots A_{iJ_i}$, and the action space for all players in market m is $\prod_{i=1,2,\ldots,N,\, j=1,2,\ldots,J_i} A_{ij}$.

Given the actions of all players $a \in \prod_{i=1,2,\ldots,N,\, j=1,2,\ldots,J_i} A_{ij}$, the payoff for player i in market

m is given by

$$\pi_i(a;\theta) = \hat{\pi}_i(a;\theta) + \epsilon^i_{a_i} \tag{3.5}$$

where $\hat{\pi}_i(a;\theta)$ represents player i's estimated (expected) payoff, $\epsilon^i_{a_i}$ is the payoff shock which is assumed to be private information of player i and θ is a vector of parameters that we are interested in. Notice here that the payoff shock $\epsilon^i_{a_i}$ is a "non-strategic error term", that is, your rivals' payoff shocks $\epsilon^{-i}_{a_{-i}}$ do not enter your payoff function and thus, do not affect your release time decision.

Player i chooses a_i in order to maximize her objective function given beliefs about her opponents' actions a_{-i}. The probability that strategy a_i is chosen by player i conditional on other players' action a_{-i} is given by

$$\Pr(a_i \big| a_{-i}) = \Pr\left\{\hat{\pi}_i(a_i, a_{-i};\theta) + \epsilon^i_{a_i} \geq \hat{\pi}_i(a'_i, a_{-i};\theta) + \epsilon^i_{a'_i}\right\} \tag{3.6}$$

where $a'_i \in \prod_{j=1,2,\ldots,J_i} A_{ij}$. I further assume that $\epsilon^i_{a_i}$ is an i.i.d (across actions and players) draw from a type I extreme value distribution, thus the conditional probability can be written as a closed-form solution.

$$\Pr(a_i \big| a_{-i}) = \frac{\exp(\hat{\pi}_i(a_i, a_{-i};\theta))}{\sum_{a'_i \in \prod_{j=1,2,\ldots,J_i} A_{ij}} \exp(\hat{\pi}_i(a'_i, a_{-i};\theta))} \tag{3.7}$$

The strategic timing game in this paper is a one-time sequential game. Players move one by one according to a pre-specified order, which is known to the players but may be unknown to the econometrician. Let o be the pre-specified order and I use $o(k) = i$ to imply that player i is the k-th player to move, thus the last player to move is denoted as $o(N)$. I follow the method of *pseudo-backward induction* proposed by Einav (2009 [30]), which let us solve the game backwards. I briefly introduce the methodology in this section. The probability of the last player to move, $o(N)$, choosing the strategy, $a_{o(N)}$, given the rivals' decisions $a_{prev(o(N))}$ (all the rivals move before the last player $o(N)$), is

$$\Pr(a_{o(N)}|a_{prev(o(N))}) = \frac{\exp(\hat{\pi}_{o(N)}(a_{o(N)}, a_{prev(o(N))};\theta))}{\sum_{a'_{o(N)} \in \prod_{j=1,2,\ldots,J_{o(N)}} A_{o(N)j}} \exp(\hat{\pi}_{o(N)}(a'_{o(N)}, a_{prev(o(N))};\theta))} \tag{3.8}$$

where $prev(o(N)) = \{h : o^{-1}(h) < o^{-1}(o(N))\}$ denotes the set of players who play before player $o(N)$. Therefore, for each one of the possible configurations of the rivals' decisions $a_{prev(o(N))}$, we have a probability distribution indicating how likely the last player to move, $o(N)$, would choose the particular strategy $a_{o(N)}$. This allows us to update the payoff values for the second last player to move, $o(N-1)$, by integrating over player $o(N)$'s decision probabilities, that is

$$\hat{\pi}_{o(N-1)}(a_{prev(o(N))};\theta) = \sum_{a_{o(N)} \in \prod_{j=1,2,\ldots,A_{o(N)j}}} \Pr(a_{o(N)}|a_{prev(o(N))})\hat{\pi}_{o(N-1)}(a_{o(N)}, a_{prev(o(N))};\theta) \tag{3.9}$$

And thus,

$$\pi_{o(N-1)}(a_{prev(o(N))};\theta) = \hat{\pi}_{o(N-1)}(a_{prev(o(N))};\theta) + \epsilon^{o(N-1)}_{a_{o(N-1)}} \tag{3.10}$$

The modified payoffs above would directly enter the decision of player $o(N-1)$, which will again be used to update the payoffs of $o(N-2)$ and so on. Eventually, we can iteratively update the payoffs all the way up to the player who moves first. Hence, given an order o, the likelihood of observing a particular release schedule a is given by

$$\Pr(a|o) = \prod_{i=1}^{N} \Pr(a_i|a_{prev(i)}, o) \tag{3.11}$$

Following the common practice in the literature[14], I exploit the assumption that players act optimally, and thus the data I observe is generated by equilibrium behavior. I estimated the model parameters using maximum likelihood.

3.5.2 Payoff Function

Now I specify the payoff functions $\pi_i(a;\theta)$ for our industrial setting. First, I assume the (estimated) payoff of a portfolio of films to be the sum of the (estimated) payoffs of each film in the portfolio. Second, I define the (estimated) payoff of each movie in the portfolio to be

[14] Also mentioned in Einav and Nevo (2006)[31].

the cumulative undiscounted payoff through all of its initial and sequential markets. That is,

$$\pi_i(a_i, a_{-i}; \theta) = \sum_{j=1,m=1}^{j=J_i,m=M} \hat{\pi}_{ijm}(a_i, a_{-i}; \theta) + \epsilon^i_{a_i} \tag{3.12}$$

Third, I use (estimated) market shares to represent the payoffs. Using market shares rather than revenue can eliminate the effects of population growth and price inflation. From the demand estimation stage with the logit specification, we denote the (estimated) market share of movie j at week t to be

$$\hat{s}^m_{jt} = \frac{\exp(\delta^m_{jt})}{\sum_{k=0}^{J_t} \exp(\delta^m_{kt})} \tag{3.13}$$

where $m = 1, 2$ in our current setting, indicating theatrical and home video window respectively.

Thus, the final total payoff of movie j distributed through two window markets can be expressed as

$$\hat{\pi}_{ij} = \sum_{t=r^1_j}^{t=r^2_j} \hat{s}^1_{jt} + \sum_{t=r^2_j}^{t=\mathcal{T}} \hat{s}^2_{jt} \tag{3.14}$$

where $r^1_j < r^2_j < \mathcal{T}$, $\sum_{t=r^1_j}^{t=r^2_j} \hat{s}^1_{jt}$ represents the total (estimated) demand of movie j from the theatrical window aggregated from r^1_j to r^2_j since the movie will exit the first market before it enter the second, and $\sum_{t=r^2_j}^{t=\mathcal{T}} \hat{s}^2_{jt}$ is the total (estimated) demand from the sequential home video window aggregated from r^2_j to $\mathcal{T}$.

To reflect the costs of distribution and the weights distributors put on the windows, I add two more coefficients to estimate at the game stage. Equation 3.14 now is further modified to be the following:

$$\hat{\pi}_{ij} = w \sum_{t=r^1_j}^{t=r^2_j} (\hat{s}^1_{jt} - F) + (1 - w) \sum_{t=r^2_j}^{t=\mathcal{T}} \hat{s}^2_{jt} \tag{3.15}$$

where F represents the weekly distributional cost of the theatrical market in the unit of market share; the weekly distributional cost of the video market is normalized to be 0. The weight coefficient w shows the distributor's weights assigned to each window, reflecting their marketing preferences.

Eventually, if we expand the number of sequential markets from two to any finite number M, we can generalize player i's objective function to maximize as the following:

$$\begin{aligned}
\pi_i(a_i, a_{-i}; \theta) &= \sum_{j=1,m=1}^{j=J_i,m=M} \hat{\pi}_{ijm}(a_i, a_{-i}; \theta) + \epsilon^i_{a_i} \\
&= \sum_{j=1,m=1}^{j=J_i,m=M} w^m \sum_{t=r_j^m}^{t=r_j^{m+1}} \left(\hat{s}^m_{jt} - F^m \right) + \epsilon^i_{a_i} \\
&= \sum_{j=1,m=1}^{j=J_i,m=M} w^m \sum_{t=r_j^m}^{t=r_j^{m+1}} \left(\frac{\exp(\delta^m_{jt})}{\sum_{k=0}^{J_t} \exp(\delta^m_{kt})} - F^m \right) + \epsilon^i_{a_i} \\
&= \sum_{j=1,m=1}^{j=J_i,m=M} w^m \sum_{t=r_j^m}^{t=r_j^{m+1}} \left(\frac{\exp(X_j\beta_m + \tau^m_t - \lambda_m(t - r^m_j))}{\sum_{k=0}^{J_t} \exp(X_k\beta_m + \tau^m_t - \lambda_m(t - r^m_k))} - F^m \right) + \epsilon^i_{a_i}
\end{aligned} \tag{3.16}$$

where $\sum_{m=1}^{M} w^m = 1$. We sum over the time, the initial and sequential windows, and all the movies in the portfolio. β, τ_t, and λ are derived from the demand estimation stage via the logit or the nested logit specification. F and w are estimated in the stage of the strategic timing game.

3.6 Results: Optimal Video Release Time

The strategic timing game depicted in Section 4.2 allows each player's simultaneously decided choices to be multidimensional; for instance, for theatrical-video market combination, we have a two-dimensional choice set (r_1, r_2). When the number of sequential markets goes up, the choice set gets larger, increasing the dimensionality and causing computational burden. In this section, I move r_1 outside of the choice set and treat it as public information. I limit the attention to exploring how distributors choose the video release date (r_2) but take the choice of the theatrical release date (r_1) as given. I simplify the problem in this way since 1) our research interest is regarding the window length[15] $(r_2 - r_1)$ instead of the window

[15]The sample window length is plotted in Figure 3.10.

location in the timeline and 2) keeping r_1 as fixed and known to all players can reduce the action space by half and make the counterfactual analysis feasible.

3.6.1 The Estimation Setup

I lay out the setup and assumptions of the simplified game, most of which are cast to ease the estimation complexity and computational burden.

The Players: I assume the distribution of each film is independent of each other, in other words, each film represents a unique player, which also means that the number of movies in each player's portfolio is limited to 1. This assumption can later be relaxed so that the relationship between players and films is not necessarily one-to-one. Furthermore, I use a year-season combination to group movies. I assume the competition exists mainly among movies from the same season. This assumption aligns with the common industrial practice (described in Section 2.2) where distributors first aim for a specific season and then choose a week within the chosen season. For the estimation purpose in this paper, I set each season to be 5 weeks long and each season represents an independent timing game. I vary the number of movies (players) in a season to represent different levels of competitiveness and the movies (players) are chosen based on the rank of movie budgets or movie qualities; the estimation results are based on the top movies only.

Payoff Period: The estimated demand of each film is aggregated through the one-year period after the initial release; the payoff earned after the one year is ignored. As we have mentioned previously, the revenue decays as time goes, thus such assumption is consistent with the industrial focus. I modify the general payoff function (Equation 3.16) to fit our two-window setting.

$$\pi_i(a_i, a_{-i}; \theta) = \boldsymbol{w} \sum_{t=r_j^1}^{t=r_j^2} \left(\frac{\exp(X_j\beta_1 + \tau_t^1 - \lambda_1(t - r_j^1))}{\sum_{k=0}^{J_t} \exp(X_k\beta_1 + \tau_t^1 - \lambda_1(t - r_k^1))} - \boldsymbol{F} \right) + \boldsymbol{(1 - w)} \sum_{t=r_j^2}^{t=\mathcal{T}} \frac{\exp(X_j\beta_2 + \tau_t^2 - \lambda_2(t - r_j^2))}{\sum_{k=0}^{J_t} \exp(X_k\beta_2 + \tau_t^2 - \lambda_2(t - r_k^2))} + \epsilon_{a_i}^i \tag{3.17}$$

where $w \in [0, 1]$ is the weight that distributors put on the two windows and F reflects the cost the exhibitors may charge to agree to show the movies on the theatrical screens. The longer the theatrical window is, the higher the cost can be. For convenience, F is measured in the unit of the market share and the cost of making video copies is normalized to be 0.

The Timeline: The game starts after r_1 has been decided and made known to all players[16]. Each player decides on r_2 in the order of movie budgets from the highest to the lowest. For example, if three players compete with each other in the same season whose budgets are ranked as $B_1 > B_2 > B_3$, the timeline of this three-player game shows in Figure 3.9.

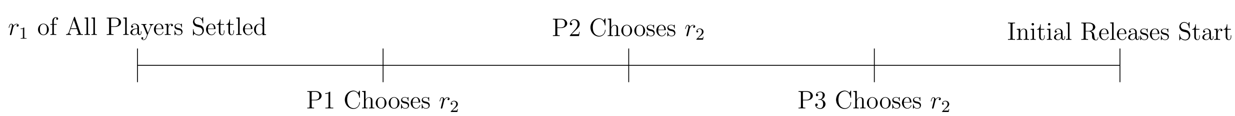


Figure 3.9: Timeline of The Three-Player Game

3.6.2 Parameter Estimates and Model Fit

I present the parameter estimates based on Equation 3.17. I conduct the estimation for two peak seasons across the sample years (2006 - 2019) and the results are shown in Table 3.6. To rationalize the observed equilibria, the estimated weight distributors put on the theatrical window is 0.86 for the Summer season and 0.81 for the Winter season. The estimated cost of

[16] I model the choices, both r_1 and r_2, to be made before the actual initial releases, because distributors sign contracts ex-ante in practice and release dates of windows are in general pre-scheduled.

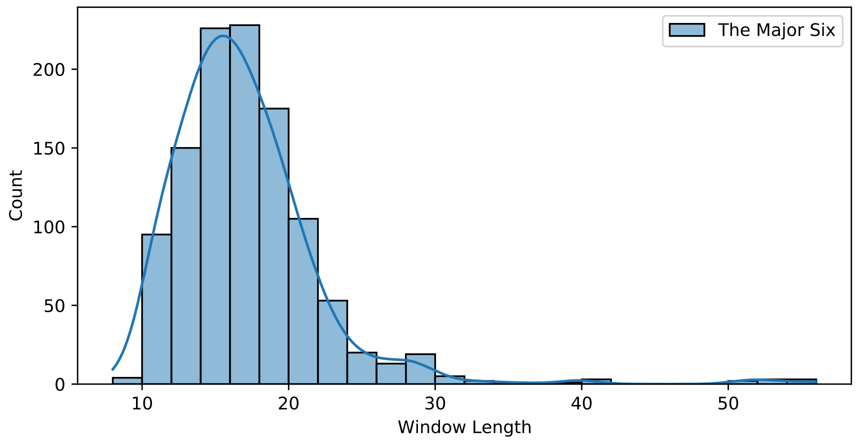


Figure 3.10: Window Length ($r_2 - r_1$) Distribution, Sample Year: 2006 - 2019

the theatrical window is 0.14 per week for Summer and 0.18 per week for Winter measured in the unit of market shares.

Under the parameter estimates displayed in Table 3.6, the estimated average window length ($r_2 - r_1$) for a 2-player game is 16 weeks while the actual observed window length is 18 weeks for the Summer Season. As for the Winter season, the estimated is 13 and the observed is 15.

3.7 Counterfactual Analyses

In this section, I conduct a series of counterfactual analyses for both Summer (week 26 - 30) and Winter (week 48 - 52) seasons. I change the value of one of the parameters keeping the rest constant. The goal is to investigate how the theatrical window length ($r_2 - r_1$) may change when the variable cost (F), distributors' weight (w), movie quality ($X\beta$), or consumers' utility decay rate (λ) changes. To reduce the computational burden, I set the number of players to be two ($N = 2$).

Table 3.6: Video Release Timing Game: Parameter Estimates (2006 - 2018)

Seasons	Summer (Week 26 - 30)	Winter (Week 48 - 52)
Number of Players	2	2
w	0.8620***	0.8069**
	(3.6997)	(3.1779)
F	0.1404	0.1778
	(0.9568)	(1.3606)

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The estimation results are based on the N movies with the highest budgets.

Table 3.7: Video Release Timing Game: Model Fit (2006 - 2018)

Seasons	Summer (Week 26 - 30)	Winter (Week 48 - 52)
Number of Players	2	2
Predicted $(r_2 - r_1)$	17.5385	15.0769
Observed $(r_2 - r_1)$	15.5000	13.0769

Variable Cost, *F* I first investigate the impact of the theatrical variable cost F on the window length $(r_2 - r_1)$ holding all the rest coefficients constant at their estimated values. Intuitively, if it gets more costly to distribute movies into theaters, distributors would shrink the theatrical window and let the products enter the sequential market sooner. The left panel of Figure 3.11 is consistent with the expectation. As F increases from 0 to 1[17], the theatrical window length is estimated to shrink monotonically to almost 0. It indicates that if the cost is too high compared to the revenue gain, distributors may strategically skip the first window and jump to the sequential one directly.

Distributors' Weight, w The other parameter of interest at the game stage is distributors' weight w on the two windows. It reflects distributors' preferences and measures the responsiveness of total effective demand to a change in levels of either theatrical or video revenue performance. I again keep the value of F at its estimated level and change the value of w to investigate how the window length changes as the weight changes. The right panel of Figure 3.11 validates that the higher the weight is on the theatrical window, the longer the theatrical window would be. I have also tested the case where the variable cost F is 0 and the weight on the theatrical window is 1, the estimated equilibrium of r_2 turns out to be 53 weeks (one year) from the initial release, which indicates that the theatrical window length would be set to be as long as possible if there is no variable cost and distributors only care about the first window.

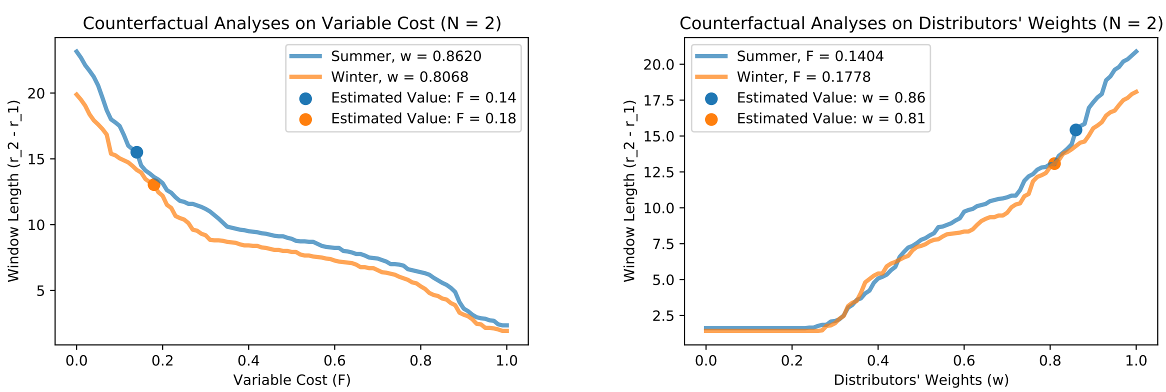


Figure 3.11: Counterfactual Analyses on F and w

[17] F is measured in the unit of market shares, thus the value of F should be within the range of $[0, 1]$.

Movie Quality, $\boldsymbol{X\beta}$ Under the Logit specification, consumers' mean utility is made up of three parts, the movie quality, $X\beta$, the utility decay, $\lambda(t - r_j)$, and the seasonality, τ. I now investigate how the movie quality may affect the window length decision. As mentioned in Figure 3.3, the movie demand commonly starts high and decreases afterward. The movie quality is the factor that decides the starting point of the revenue, or in other words, how strong the opening would be. For the counterfactual analyses, I add a multiplier coefficient in front of the movie quality to hypothetically adjust its value, thus the $X\beta$ part is now modified to be $b \cdot X\beta$. I change the value of the multiplier b to represent the changes in the movie quality. For instance, if $b = 0.5$, then the movie quality is discounted to be only half of the original value; if $b = 2$, then the movie quality doubles. Figure 3.12 is plotted to show the impact of the multiplier b on the window length in a 2-player market. We can observe that as the movie quality increases, the optimal theatrical window length also rises. Movies with higher quality can have a stronger opening in both windows compared to the original state, and such expanded effect is more significant in the initial window, and thus it would be optimal to enlarge the first window to make the best of the quality rise.

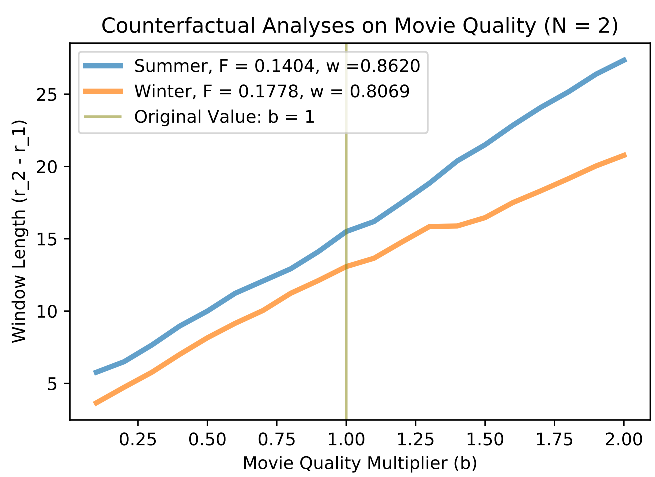


Figure 3.12: Counterfactual Analyses on Movie Quality

Utility Decay, $\boldsymbol{\lambda}$ Lastly, I explore how consumers' utility decay rates of the two windows impact distributors' window length decisions. I again adopt a multiplier coefficient γ in front of the utility decay rate λ and hypothetically manipulate the utility decay rate through the value of the multiplier. The left panel of Figure 3.13 depicts the change in the theatrical utility decay rate holding the video utility decay rate constant. It shows that the slower the utility decays, the longer the theatrical window should be. Especially, if the theatrical decay rate doubles, the equilibrium window length would shrink towards 0. Similarly, I change the video utility decay rate holding the theatrical one constant; the results are shown in the right panel of Figure 3.13. As the video utility decays faster, distributors would extend the first window. However, the magnitude of the change rates (slope of the curves) differs between the two windows, since the video utility decay rate originally is much smaller than the theatrical one.

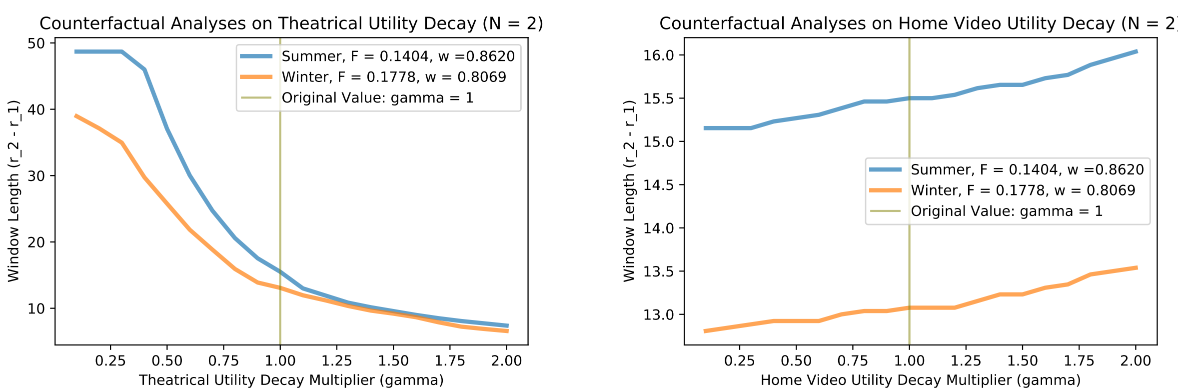


Figure 3.13: Counterfactual Analyses on Utility Decay

3.8 Conclusion

This paper uses market-level data from theatrical and home video markets to study movie distributors' release time decisions under sequential distribution. I combine demand estimation and strategic game modeling to investigate what factors may contribute to the current trend towards the shrinking theatrical window.

I model distributors' release time decisions in two stages. First, each player predicts the market shares of her movie based on the given movie characteristics and seasonality. After deriving the demand estimates and taking the parameters as given, I next model distributors' windowing strategies as a one-shot sequential-move game with incomplete information. All the players engage in a strategic release time game with one another to maximize the potential aggregated market shares of their own portfolios of films. I follow the method of pseudo-backward induction proposed by Einav (2009 [30]) to empirically solve the release time game.

At the game estimation stage, the two coefficients in the payoff function that we are interested in estimating are the weekly variable cost of theatrical distribution and the weight imposed by distributors on the two windows. The estimated variable cost is 0.14 per week for summer and 0.18 per week for winter measured in the unit of market shares. The estimated weight that distributors put on the theatrical window is 0.86 for summer and 0.81 for winter. After solving the simplified two-player game, I conduct a series of counterfactual analyses where I hypothetically change the value of one of the key coefficients holding other factors constant to investigate how the release time decisions may change. The results show that higher variable cost on theatrical distribution, lower weight on the theatrical window, poorer movie quality with weak opening performance, or faster theatrical utility decay can all potentially lead the distributors to strategically shrink the theatrical window and release video versions sooner.

I conclude this chapter by pointing out some opportunities for future research. First, although the idea of managing portfolios of movies is embedded in our game model, its effect is not estimated due to the computational burden. Modeling each players' portfolios of movies rather than individually can help to understand how distributors rank and control the competition internally. However, it will tremendously increase the dimension of each player's choice set, hence, future studies can be devoted to searching for more efficient ways of estimation. Second, with the rise of streaming services, the traditional release pattern starting from theaters is no longer strictly committed. It is now more frequently observed that movies are released onto streaming platforms directly skipping the theatrical window and streaming begins to dominate the post-theatrical market replacing home videos. Therefore, besides release time decisions, distributors can more freely choose what platforms to put their products on and in what sequence it may lead to profit maximization. The competition happens not only in the time dimension but also within venues. Future research can be conducted to investigate the value of such transition in the industry. Lastly, uncertainty plays an important role in the motion picture industry, which is not modeled in this paper. It is worthwhile in the future to explore how it may affect distributors' strategic decisions.

BIBLIOGRAPHY

[1] Ainslie, A., Drèze, X., and Zufryden, F. Modeling movie life cycles and market share. *Marketing science 24*, 3 (2005), 508–517.

[2] Anderson, S. P., De Palma, A., and Thisse, J.-F. *Discrete choice theory of product differentiation*. MIT press, 1992.

[3] August, T., Dao, D., and Shin, H. Optimal timing of sequential distribution: The impact of congestion externalities and day-and-date strategies. *Marketing Science 34*, 5 (2015), 755–774.

[4] Bass, F. M. A new product growth for model consumer durables. *Management science 15*, 5 (1969), 215–227.

[5] Bass, F. M., Krishnan, T. V., and Jain, D. C. Why the bass model fits without decision variables. *Marketing science 13*, 3 (1994), 203–223.

[6] Basuroy, S., and Chatterjee, S. Fast and frequent: Investigating box office revenues of motion picture sequels. *Journal of Business Research 61*, 7 (2008), 798–803.

[7] Basuroy, S., Chatterjee, S., and Ravid, S. A. How critical are critical reviews? the box office effects of film critics, star power, and budgets. *Journal of marketing 67*, 4 (2003), 103–117.

[8] Ben-Akiva, M., and Lerman, S. R. *Discrete choice analysis: theory and application to travel demand*. Transportation Studies, 1985.

[9] Berbeglia, F., Derdenger, T., and Tayur, S. Optimal timing of home video releases: A dynamic model of movie distribution.

[10] Berry, S., Levinsohn, J., and Pakes, A. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* (1995), 841–890.

[11] Berry, S. T. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics* (1994), 242–262.

[12] Bird, S., Klein, E., and Loper, E. *Natural language processing with Python: analyzing text with the natural language toolkit*. " O'Reilly Media, Inc.", 2009.

[13] Bresnahan, T. F., and Reiss, P. C. Empirical models of discrete games. *Journal of Econometrics 48*, 1-2 (1991), 57–81.

[14] Chatterjee, P. Online reviews: do consumers use them?

[15] Chen, P.-Y., Wu, S.-y., and Yoon, J. The impact of online recommendations and consumer feedback on sales.

[16] Chen, Y., and Shugan, S. Using new product sales to evaluate individuals in teams: Market-based metrics and empirical evidence from the entertainment industry. In *University of Arizona Eller College of Management*. 2004.

[17] Chevalier, J. A., and Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *Journal of marketing research 43*, 3 (2006), 345–354.

[18] Chiou, L. The timing of movie releases: Evidence from the home video industry. *International Journal of Industrial Organization 26*, 5 (2008), 1059–1073.

[19] De Vany, A. *Hollywood economics: How extreme uncertainty shapes the film industry*. Routledge, 2003.

[20] De Vany, A., and Lee, C. Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues. *Journal of Economic Dynamics and Control 25*, 3-4 (2001), 593–614.

[21] De Vany, A., and Walls, W. D. Bose-einstein dynamics and adaptive contracting in the motion picture industry. *The Economic Journal 106*, 439 (1996), 1493–1514.

[22] De Vany, A., and Walls, W. D. The market for motion pictures: rank, revenue, and survival. *Economic inquiry 35*, 4 (1997), 783–797.

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

[24] De Vany, A. S., and Walls, W. D. Motion picture profit, the stable paretian hypothesis, and the curse of the superstar. *Journal of Economic Dynamics and Control 28*, 6 (2004), 1035–1057.

[25] Dellarocas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management science 49*, 10 (2003), 1407–1424.

[26] Dellarocas, C., Zhang, X., and Awad, N. F. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive marketing 21*, 4 (2007), 23–45.

[27] Dhar, T., Sun, G., and Weinberg, C. B. The long-term box office performance of sequel movies. *Marketing Letters 23*, 1 (2012), 13–29.

[28] Duan, W., Gu, B., and Whinston, A. B. Do online reviews matter?—an empirical investigation of panel data. *Decision support systems 45*, 4 (2008), 1007–1016.

[29] Einav, L. Seasonality in the us motion picture industry. *The Rand journal of economics 38*, 1 (2007), 127–145.

[30] Einav, L. Not all rivals look alike: Estimating an equilibrium model of the release date timing game. *Economic Inquiry 48*, 2 (2010), 369–390.

[31] Einav, L., and Nevo, A. Empirical models of imperfect competition: A discussion. Tech. rep., CSIO Working Paper, 2006.

[32] Elberse, A. The power of stars: Do star actors drive the success of movies? *Journal of marketing 71*, 4 (2007), 102–120.

[33] Elberse, A., and Eliashberg, J. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing science 22*, 3 (2003), 329–354.

[34] Eliashberg, J., Elberse, A., and Leenders, M. A. The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing science 25*, 6 (2006), 638–661.

[35] Eliashberg, J., Jonker, J.-J., Sawhney, M. S., and Wierenga, B. Moviemod: An implementable decision-support system for prerelease market evaluation of motion pictures. *Marketing Science 19*, 3 (2000), 226–243.

[36] Eliashberg, J., Weinberg, C. B., and Hui, S. K. Decision models for the movie industry. In *Handbook of marketing decision models*. Springer, 2008, pp. 437–468.

[37] Ellickson, P. B., and Misra, S. Structural workshop paper—estimating discrete games. *Marketing Science 30*, 6 (2011), 997–1010.

[38] Feichtinger, G. Optimal pricing in a diffusion model with concave price-dependent market potential. *Operations Research Letters 1*, 6 (1982), 236–240.

[39] Frank, B. Optimal timing of movie releases in ancillary markets: The case of video releases. *Journal of Cultural Economics 18*, 2 (1994), 125–133.

[40] Godes, D., and Mayzlin, D. Using online conversations to study word-of-mouth communication. *Marketing science 23*, 4 (2004), 545–560.

[41] Hennig-Thurau, T., Houston, M. B., and Heitjans, T. Conceptualizing and measuring the monetary value of brand extensions: The case of motion pictures. *Journal of Marketing 73*, 6 (2009), 167–183.

[42] Hennig-Thurau, T., Houston, M. B., and Sridhar, S. Can good marketing carry a bad product? evidence from the motion picture industry. *Marketing Letters 17*, 3 (2006), 205–219.

[43] Hennig-Thurau, T., Houston, M. B., and Walsh, G. The differing roles of success drivers across sequential channels: An application to the motion picture industry. *Journal of the Academy of Marketing Science 34*, 4 (2006), 559–575.

[44] Hennig-Thurau, T., Marchand, A., and Hiller, B. The relationship between reviewer judgments and motion picture success: re-analysis and extension. *Journal of Cultural Economics 36*, 3 (2012), 249–283.

[45] Hennig-Thurau, T., Walsh, G., and Wruck, O. An investigation into the factors determining the success of service innovations: The case of motion pictures. *Academy of Marketing Science Review 6*, 1 (2001), 1–23.

[46] Hofmann, K. H. *Co-financing Hollywood film productions with outside investors: An economic analysis of principal agent relationships in the US motion picture industry.* Springer Science & Business Media, 2012.

[47] Holt, J., and Sanson, K. *Connected viewing: Selling, streaming, & sharing media in the digital age.* Routledge, 2013.

[48] Horsky, D., and Simon, L. S. Advertising and the diffusion of new products. *Marketing Science 2*, 1 (1983), 1–17.

[49] Hutto, C., and Gilbert, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (2014), vol. 8, pp. 216–225.

[50] Jones, J. M., and Ritz, C. J. Incorporating distribution into new product diffusion models. *International Journal of Research in Marketing 8*, 2 (1991), 91–112.

[51] Krider, R. E., and Weinberg, C. B. Competitive dynamics and the introduction of new products: The motion picture timing game. *Journal of Marketing Research 35*, 1 (1998), 1–15.

[52] Lehmann, D. R., and Weinberg, C. B. Sales through sequential distribution channels: An application to movies and videos. *Journal of Marketing 64*, 3 (2000), 18–33.

[53] Lilien, G. L., Rangaswamy, A., and Bruyn, A. The bass model: marketing engineering technical note. *Principles of Marketing Engineering 1515528067* (2007), 248.

[54] Lilien, G. L., Rangaswamy, A., and Van den Bulte, C. Diffusion models: Managerial applications and software. *New product diffusion models* (2000), 295–311.

[55] Litman, B. R. Predicting financial success of motion pictures. *The motion picture mega-industry* (1998).

[56] Liu, A., Liu, Y., and Mazumdar, T. Star power in the eye of the beholder: A study of the influence of stars in the movie industry. *Marketing Letters 25*, 4 (2014), 385–396.

[57] Liu, Y. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of marketing 70*, 3 (2006), 74–89.

[58] Loria, S. textblob documentation. *Release 0.15 2* (2018).

[59] Luan, J. Y., and Sudhir, K. Optimal inter-release time between sequentially released products. Tech. rep., Working Paper. Yale University, New Haven, CT, 2006.

[60] Mahajan, V., Muller, E., and Bass, F. M. New product diffusion models in marketing: A review and directions for research. *Journal of marketing 54*, 1 (1990), 1–26.

[61] Mahajan, V., and Peterson, R. A. Innovation diffusion in a dynamic potential adopter population. *Management Science 24*, 15 (1978), 1589–1597.

[62] Mazzeo, M. J. Product choice and oligopoly market structure. *RAND Journal of Economics* (2002), 221–242.

[63] Moldovan, S., and Goldenberg, J. Cellular automata modeling of resistance to innovations: Effects and solutions. *Technological Forecasting and Social Change 71*, 5 (2004), 425–442.

[64] Moul, C. C. *A concise handbook of movie industry economics*. Cambridge University Press, 2005.

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

[66] Nelson, R. A., and Glotfelty, R. Movie stars and box office revenues: an empirical analysis. *Journal of Cultural Economics 36*, 2 (2012), 141–166.

[67] Orbach, Y. Parametric analysis of the bass model. *Innovative Marketing 12*, 1 (2016), 29–40.

[68] Pesendorfer, W. Design innovation and fashion cycles. *The american economic review* (1995), 771–792.

[69] Prag, J., and Casavant, J. An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. *Journal of cultural economics 18*, 3 (1994), 217–235.

[70] Prasad, A., Bronnenberg, B., and Mahajan, V. Product entry timing in dual distribution channels: The case of the movie industry. *Review of Marketing Science 2*, 1 (2004).

[71] Purnawirawan, N., De Pelsmacker, P., and Dens, N. Balance and sequence in online reviews: How perceived usefulness affects attitudes and intentions. *Journal of interactive marketing 26*, 4 (2012), 244–255.

[72] Ravid, S. A. Information, blockbusters, and stars: A study of the film industry. *The Journal of Business 72*, 4 (1999), 463–492.

[73] Robinson, B., and Lakhani, C. Dynamic price models for new-product planning. *Management science 21*, 10 (1975), 1113–1122.

[74] Rogers, E. M., Singhal, A., and Quinlan, M. M. *Diffusion of innovations*. Routledge, 2014.

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

[76] Schmittlein, D. C., and Mahajan, V. Maximum likelihood estimation for an innovation diffusion model of new product acceptance. *Marketing science 1*, 1 (1982), 57–78.

[77] Seim, K. An empirical model of firm entry with endogenous product-type choices. *The RAND Journal of Economics 37*, 3 (2006), 619–640.

[78] Shon, J.-H., Kim, Y.-G., and Yim, S.-J. Dissecting movie genres from an audience perspective: Mti movie classification method.

[79] Smith, S. P., and Smith, V. K. Successful movies: A preliminary empirical analysis. *Applied Economics 18*, 5 (1986), 501–507.

[80] Sood, S., and Drèze, X. Brand extensions of experiential goods: Movie sequel evaluations. *Journal of Consumer Research 33*, 3 (2006), 352–360.

[81] Srinivasan, V., and Mason, C. H. Nonlinear least squares estimation of new product diffusion models. *Marketing science 5*, 2 (1986), 169–178.

[82] Swami, S., Eliashberg, J., and Weinberg, C. B. Silverscreener: A modeling approach to movie screens management. *Marketing Science 18*, 3 (1999), 352–372.

[83] Trumpbour, J. *Selling Hollywood to the world: US and European struggles for mastery of the global film industry, 1920-1950.* Cambridge University Press, 2007.

[84] Ulin, J. *The business of media distribution: Monetizing film, TV and video content in an online world.* CRC Press, 2013.

[85] Van den Bulte, C. Want to know how diffusion speed varies across countries and products? try using a bass model. *PDMA visions 26*, 4 (2002), 12–15.

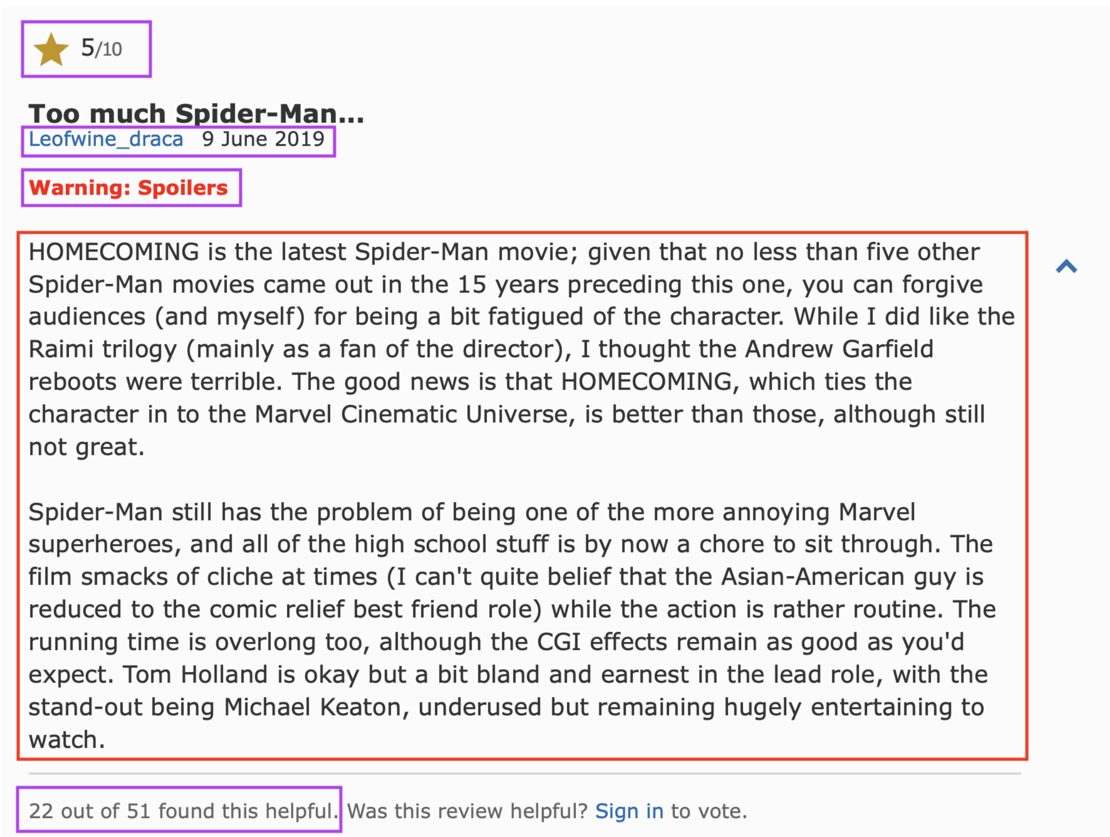
[86] Vogel, H. L. *Entertainment industry economics: A guide for financial analysis.* Cambridge University Press, 2020.

[87] Wilson, L. O., and Norton, J. A. Optimal entry timing for a product line extension. *Marketing Science 8*, 1 (1989), 1–17.

[88] Zufryden, F. S. Linking advertising to box office performance of new film releases-a marketing planning model. *Journal of Advertising Research 36*, 4 (1996), 29–42.

Appendix A

CHAPTER 1: WORD OF MOUTH AND MOTION PICTURES


5/10

Too much Spider-Man...

Leofwine_draca 9 June 2019

Warning: Spoilers

HOMECOMING is the latest Spider-Man movie; given that no less than five other Spider-Man movies came out in the 15 years preceding this one, you can forgive audiences (and myself) for being a bit fatigued of the character. While I did like the Raimi trilogy (mainly as a fan of the director), I thought the Andrew Garfield reboots were terrible. The good news is that HOMECOMING, which ties the character in to the Marvel Cinematic Universe, is better than those, although still not great.

Spider-Man still has the problem of being one of the more annoying Marvel superheroes, and all of the high school stuff is by now a chore to sit through. The film smacks of cliche at times (I can't quite belief that the Asian-American guy is reduced to the comic relief best friend role) while the action is rather routine. The running time is overlong too, although the CGI effects remain as good as you'd expect. Tom Holland is okay but a bit bland and earnest in the lead role, with the stand-out being Michael Keaton, underused but remaining hugely entertaining to watch.

22 out of 51 found this helpful. Was this review helpful? Sign in to vote.

Figure A.1: WOM Data: An Scraped Example from IMDB

Figure A.1 shows a scraped example from IMDB. Each observation in the WOM data set includes the full-text review, the reviewer's user identity number, the rating value out of a 10 scale, the date when the review is posted, if the review is a spoiler, the number of thumbs-up indicating the review is helpful, and the number of the thumbs-down indicating the review is unhelpful.

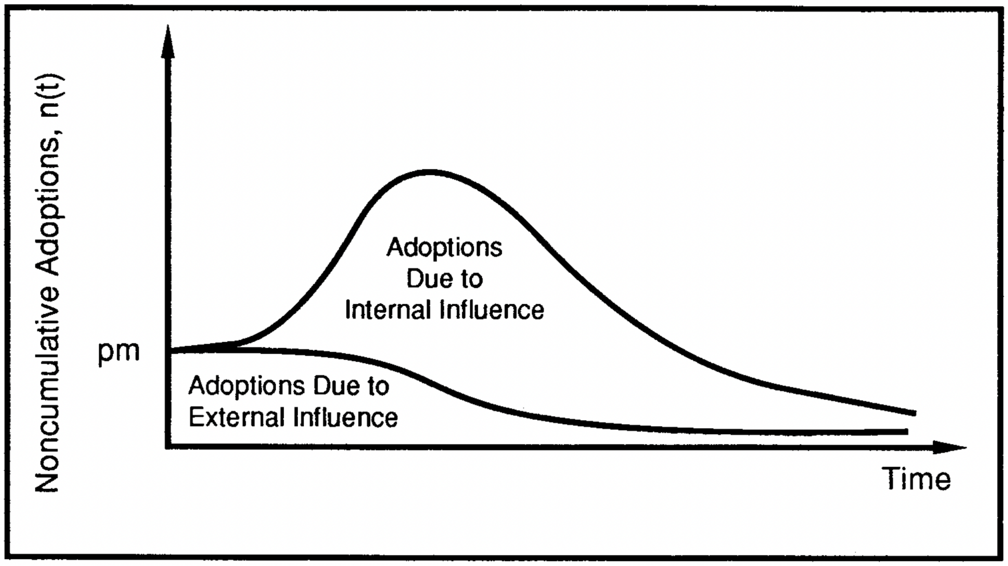


Figure A.2: Adoption Due to External and Internal Influences in the Bass Model from Mahajan, Muller, and Bass (1990 [60])

Figure A.2 displayed here is the Figure 1A from Mahajan, Muller, and Bass (1990 [60]). It plots the conceptual structure underlying the Bass model.

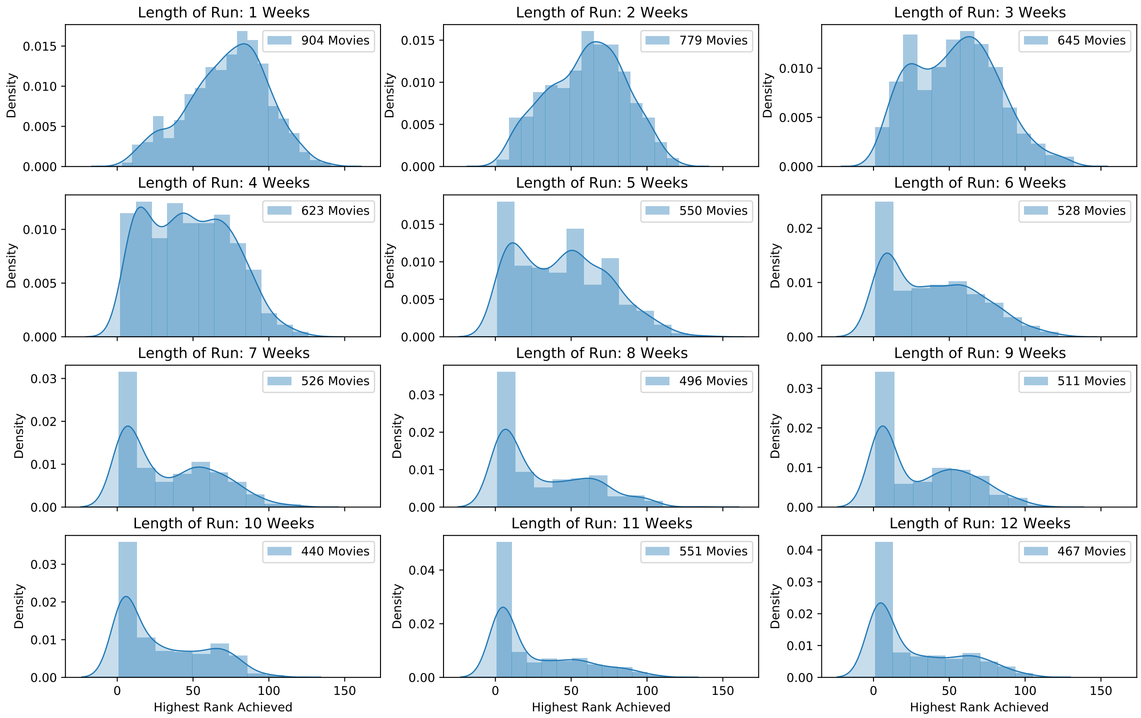


Figure A.3: Distribution of Highest Rank Achieved by Length of Theatrical Run

Figure A.3 shows the distribution of the highest rank a movie has achieved, plotted by movies' length of theatrical run. The pattern suggests that longer-lived movies tend to reach and sustain the high ranks.

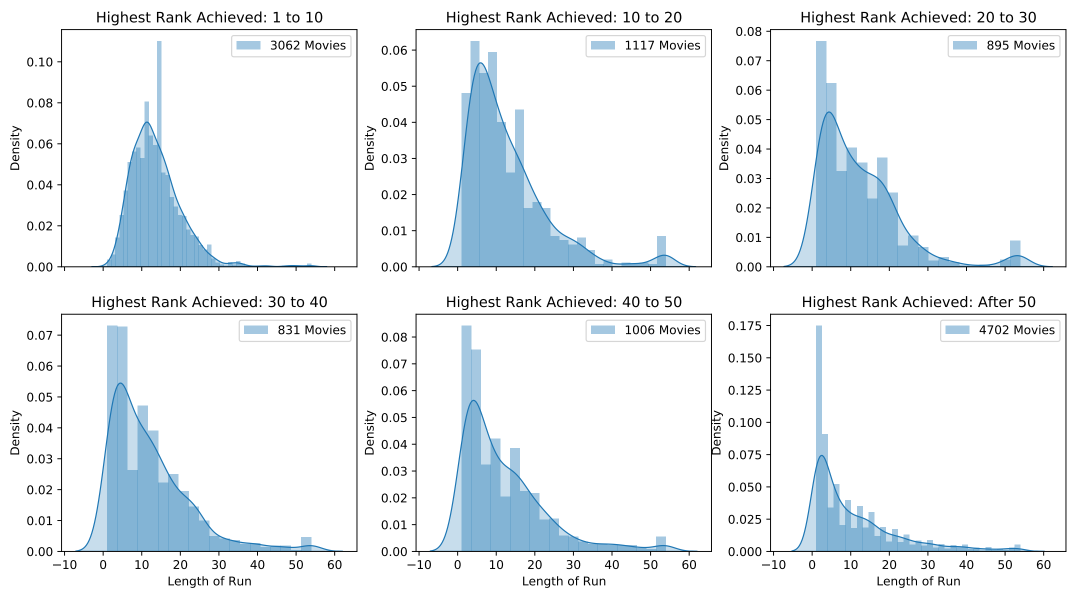


Figure A.4: Distribution of Length of Theatrical Run by Highest Rank Achieved

Figure A.4 shows the distribution of movies' length of theatrical run, plotted by their highest rank achieved. The pattern suggests that higher-ranked movies appear to also have longer runs.

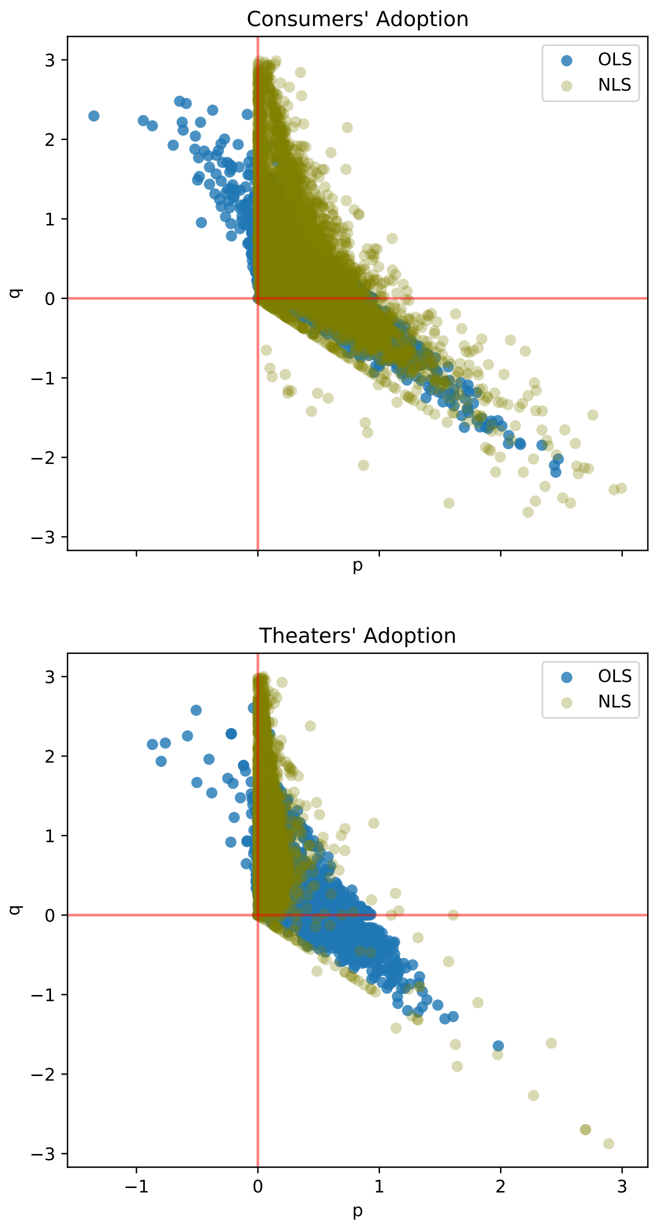


Figure A.5: Regression Results of the Bass Model: An Overview

Appendix B

CHAPTER 2: MOVIE DEMAND ESTIMATION: EVIDENCE FROM THEATRICAL AND HOME VIDEO MARKETS

Table B.1: Star Rank

Rank	Star Name	# of Sample Wide Releases Lead in
1	Samuel L. Jackson	47
2	Julianne Moore	42
3	Matt Damon	39
4	Robert De Niro	39
5	Johnny Depp	38
6	Ewan McGregor	37
7	Akshay Kumar	37
8	Mark Wahlberg	36
9	Nicolas Cage	36
10	Liam Neeson	35

I rank the sample stars by their productivity which is measured by the number of films each star has participated in as a leading actor or actress. There are in total 14746 stars and Table B.1 lists the top 10. Similarly, I rank the sample directors/writers by their productivity which is measured by the number of films each director/writer has participated in. Table B.2 lists the top 10 directors and Table B.3 lists the top 10 writers.

Table B.2: Director Rank

Rank	Director Name	# of Sample Wide Releases Directed
1	Woody Allen	21
2	Steven Soderbergh	20
3	Clint Eastwood	19
4	Tyler Perry	18
5	François Ozon	17
5	Steven Spielberg	17
7	Michael Winterbottom	16
7	Ridley Scott	16
9	Richard Linklater	15
10	Ron Howard	14

Table B.3: Writer Rank

Rank	Writer Name	# of Sample Wide Releases Written
1	Woody Allen	21
2	Luc Besson	20
3	Tyler Perry	19
4	François Ozon	12
5	Paul Laverty	11
5	Ehren Kruger	11
5	Steven Knight	11
8	Leigh Whannell	10
8	Joel Coen	10
8	M. Night Shyamalan	10

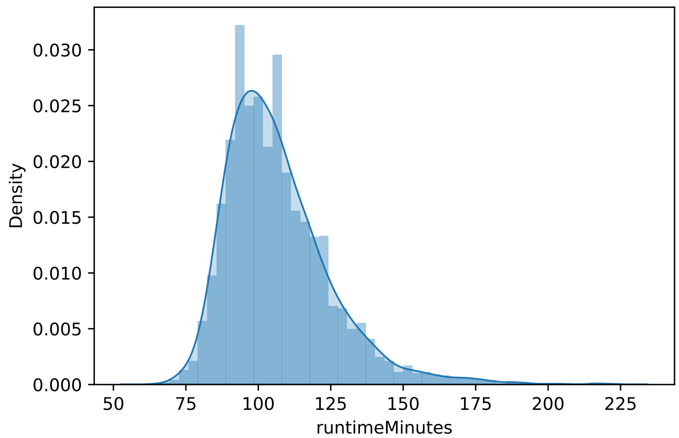


Figure B.1: Distribution of Runtime

Runtime Figure B.1 shows the sample distribution of runtime in minutes. The average runtime of our wide-release sample is 107 minutes and the median is 104 minutes.

Table B.4: Demand Estimation: Theatrical (2006 - 2019)

	Logit	**Logit**	**NL**	**NL 1st Stage**	**NL**	**NL 1st Stage**
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	IV-2SLS	OLS	IV-2SLS	OLS
weeks in release	-0.317***	-0.407***	-0.147***	-0.313***	-0.204***	-0.402***
	(-39.54)	(-52.01)	(-4.04)	(-38.14)	(-7.23)	(-50.71)
runtimeMinutes	0.0183***		0.00841**	0.0180***		
	(5.72)		(3.15)	(5.53)		
sequel	0.297**		0.139**	0.293**		
	(4.06)		(3.00)	(4.06)		
country	0.178*		0.0881*	0.171*		
	(2.74)		(2.32)	(2.67)		
language	0.281		0.110	0.313		
	(0.62)		(0.49)	(0.71)		
color	0.410		0.181	0.408		
	(1.77)		(1.53)	(1.76)		
top stars	0.145**		0.0708**	0.134**		
	(4.05)		(2.96)	(3.57)		
top directors	0.363**		0.168***	0.364**		
	(4.14)		(3.50)	(4.14)		
top writers	0.277***		0.126**	0.283***		
	(5.57)		(2.80)	(5.46)		
Adventure	0.282*		0.123*	0.288*		
	(2.51)		(2.04)	(2.50)		
Animation	1.237***		0.573***	1.225***		
	(12.63)		(4.26)	(12.40)		
Biography	0.0148		0.0129	0.00618		
	(0.08)		(0.15)	(0.04)		
Comedy	-0.152		-0.0802	-0.141		
	(-1.78)		(-1.80)	(-1.76)		
Crime	-0.418*		-0.197*	-0.432**		
	(-3.01)		(-2.39)	(-3.30)		

Documentary	-0.787***		-0.376**	-0.754**		
	(-4.50)		(-2.68)	(-4.12)		
Drama	-0.322*		-0.144*	-0.330**		
	(-2.89)		(-2.42)	(-3.01)		
Horror	-0.123		-0.0584	-0.111		
	(-1.34)		(-1.31)	(-1.21)		
Other Genres	0.149		0.0669	0.181		
	(0.57)		(0.52)	(0.69)		
20th Century Fox	0.664***		0.310**	0.662***		
	(7.47)		(3.08)	(7.60)		
Paramount Pictures	0.975***		0.449***	0.977***		
	(9.80)		(3.43)	(9.64)		
Sony Pictures	1.054***		0.482***	1.057***		
	(15.65)		(3.88)	(15.68)		
Universal	1.073***		0.496***	1.065***		
	(9.40)		(3.66)	(9.22)		
Walt Disney	1.355***		0.623***	1.355***		
	(9.35)		(3.40)	(9.18)		
Warner Bros.	0.931***		0.433***	0.930***		
	(8.31)		(3.49)	(8.27)		
$\log(s_{j\|g})$			0.538***		0.503***	
			(4.61)		(7.49)	
IV ratings				-0.714***		-0.256
				(-6.16)		(-1.56)
IV top stars				-0.0887		-0.112
				(-0.44)		(-0.70)
IV weeks in release				0.181***		0.165***
				(9.59)		(7.38)
group count				-0.0317***		-0.0167**
				(-4.23)		(-2.81)
Constant	-10.01***	-4.903***	-6.078***	-2.819	-3.835***	-0.991
	(-15.12)	(-42.36)	(-6.45)	(-1.83)	(-26.41)	(-0.86)
Observations	22320	22320	22320	22320	22320	22320

Adjusted R^2	0.580	0.792		0.578		0.783
Movie Fixed Effect	No	Yes	No	No	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.4 shows the full results of the theatrical demand estimation.

Table B.5: Demand Estimation: Home Video (2006 - 2019)

	Logit	**Logit**	**NL**	**NL 1st Stage**	**NL**	**NL 1st Stage**
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	IV-2SLS	OLS	IV-2SLS	OLS
weeks in release	-0.0750***	-0.0727***	-0.0237	-0.0745***	-0.0183***	-0.0752***
	(-28.70)	(-55.80)	(-1.60)	(-29.90)	(-4.04)	(-49.74)
runtimeMinutes	0.0192***		0.00605	0.0190***		
	(7.26)		(1.39)	(7.04)		
sequel	0.282***		0.0906	0.281***		
	(6.58)		(1.51)	(6.67)		
country	0.185***		0.0610	0.182***		
	(4.53)		(1.59)	(4.70)		
language	0.0947		0.0272	0.100		
	(0.46)		(0.37)	(0.47)		
color	-0.0644		-0.0200	-0.0654		
	(-0.31)		(-0.27)	(-0.31)		
top stars	0.0977***		0.0313	0.0945**		
	(4.97)		(1.57)	(4.14)		
top directors	0.224***		0.0738	0.218***		
	(5.33)		(1.81)	(5.45)		
top writers	0.143**		0.0448	0.141**		
	(3.96)		(1.65)	(3.78)		
Adventure	0.113		0.0323	0.121		
	(1.12)		(0.77)	(1.15)		
Animation	0.861***		0.275	0.850***		
	(7.95)		(1.47)	(7.87)		
Biography	-0.271*		-0.0818	-0.278**		
	(-2.96)		(-1.23)	(-3.05)		
Comedy	-0.205***		-0.0669	-0.201***		
	(-4.53)		(-1.34)	(-4.52)		
Crime	-0.566***		-0.176	-0.565***		
	(-6.41)		(-1.59)	(-6.32)		

Documentary	-0.875**		-0.285	-0.864**		
	(-3.11)		(-1.53)	(-3.02)		
Drama	-0.322***		-0.103	-0.321***		
	(-4.34)		(-1.36)	(-4.32)		
Horror	-0.306**		-0.0960	-0.305**		
	(-3.38)		(-1.30)	(-3.49)		
Other Genres	-0.298		-0.0918	-0.298		
	(-0.86)		(-0.74)	(-0.86)		
20th Century Fox	0.207**		0.0653	0.203**		
	(3.55)		(1.21)	(3.46)		
Paramount Pictures	-0.0965		-0.0323	-0.0938		
	(-1.11)		(-0.98)	(-1.09)		
Sony Pictures	0.223**		0.0680	0.224**		
	(3.30)		(1.14)	(3.25)		
Universal	0.120		0.0387	0.116		
	(1.46)		(0.90)	(1.40)		
Walt Disney	0.474**		0.150	0.468**		
	(4.09)		(1.27)	(3.99)		
Warner Bros.	0.154		0.0499	0.149		
	(1.95)		(1.10)	(1.88)		
log($s_{j\|g}$)			0.686***		0.751***	
			(3.29)		(12.22)	
IV ratings				-0.577		-0.310
				(-0.86)		(-1.65)
IV top stars				-0.216		-0.214
				(-0.33)		(-1.08)
IV weeks in release				0.00186		-0.00573
				(0.08)		(-0.85)
group count				-0.00664*		-0.00344***
				(-2.19)		(-3.49)
Constant	-11.73***	-8.767***	-6.887***	-2.288	-5.407***	-1.832
	(-23.95)	(-52.81)	(-4.45)	(-0.53)	(-19.14)	(-1.47)
Observations	69790	69790	69790	69790	69790	69790

Adjusted R^2	0.494	0.627		0.473		0.576
Movie Fixed Effect	No	Yes	No	No	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.5 shows the full results of the video demand estimation.

Table B.6: Demand Estimation (Alternative): Theatrical (2006 - 2019)

	Logit		**Nested Logit**	
	(1)	(2)	(3)	(4)
	OLS	OLS	IV-2SLS	IV-2SLS
weeks in release	-0.593***	-0.661***	-0.285***	-0.397***
	(-62.81)	(-36.57)	(-3.92)	(-8.61)
weeks in release^2	0.0126***	0.0122***	0.0061***	0.0073***
	(25.73)	(12.58)	(3.90)	(7.18)
log($s_{j\|g}$)			0.522***	0.402***
			(4.25)	(5.97)
Observations	22320	22320	22320	22320
Adjusted R^2	0.663	0.863		
Movie Fixed Effect	No	Yes	No	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.6 shows the results of the theatrical demand estimation with the alternative model.

Table B.7: Demand Estimation (Alternative): Home Video (2006 - 2019)

	Logit		**Nested Logit**	
	(1)	(2)	(3)	(4)
	OLS	OLS	IV-2SLS	IV-2SLS
weeks in release	-0.221***	-0.219***	-0.126	-0.241***
	(-166.54)	(-109.81)	(-14.33)	(-13.21)
weeks in release^2	0.0026***	0.0026***	0.0015***	0.0029***
	(113.69)	(89.45)	(14.32)	(13.23)
log($s_{j\|g}$)			0.431***	-0.102
			(10.77)	(-1.22)
Observations	69790	69790	69790	69790
Adjusted R^2	0.620	0.746		
Movie Fixed Effect	No	Yes	No	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.7 shows the results of the home video demand estimation with the alternative model.

B.1 Nested Logit Probability Choice Formulas

The utility of consumer i from watching movie j in week t is given by

$$\begin{aligned} u_{ijt} &= \delta_{jt} + \zeta_{it} + (1-\sigma)\epsilon_{ijt} \\ &= X_j\beta + \xi_j - \lambda(t - r_j) + \tau_t + \zeta_{it} + (1-\sigma)\epsilon_{ijt} \\ &\quad i = 1, \ldots, I_t, \quad j = 1, \ldots, J, \quad t = 1, \ldots, T, \quad 0 \le \sigma \le 1 \end{aligned} \tag{B.1}$$

where ϵ_{ijt} follows Type 1 extreme value distribution and $\zeta_{it} + (1-\sigma)\epsilon_{ijt}$ is also an extreme value random variable. Additionally, ζ_{it} is idiosyncratic group preference and σ characterizes the correlation of utilities that a consumer experiences among the products in the same group.

The selection probability of product j conditional on group g being selected is:

$$s_{jt|t} = \frac{\exp(\frac{\delta_{jt}}{1-\sigma})}{D_t}, \text{ where } D_t = \sum_{k \in \mathcal{J}_t} \exp(\frac{\delta_{kt}}{1-\sigma})$$

The selection probability of product 0 conditional on group 0 being selected is $s_{0|0} = 1$ since only product 0 is in group 0. Therefore, $D_0 = \exp(\frac{\delta_0}{1-\sigma}) = 1$.

The unconditional selection probability of the group g being selected is:

$$\begin{aligned} s_t &= \frac{D_t^{(1-\sigma)}}{\sum_t D_t^{(1-\sigma)}} \\ s_0 &= \frac{1}{\sum_{gt} D_t^{(1-\sigma)}} \end{aligned}$$

Thus, the unconditional selection probability of product j, in group g, is:

$$\begin{aligned} s_{jt} &= s_{jt|t} s_t = \frac{\exp(\frac{\delta_{jt}}{1-\sigma})}{D_t^{\sigma}[\sum_t D_t^{(1-\sigma)}]} \\ s_{00} &= s_{00|0} s_0 = \frac{1}{\sum_t D_t^{(1-\sigma)}} \end{aligned}$$

We further derive the expressions for $\frac{s_{jt}}{s_{0t}}$:

$$\begin{aligned}\frac{s_{jt}}{s_{0t}} &= \frac{\exp(\frac{\delta_{jt}}{1-\sigma})}{D_t^{\sigma}} \\ &= \frac{[\exp(\frac{\delta_{jt}}{1-\sigma})]^{\sigma}}{D_t^{\sigma}} \times [\exp(\frac{\delta_{jt}}{1-\sigma})]^{1-\sigma} \\ &= (s_{jt|t})^{\sigma} \times [\exp(\frac{\delta_{jt}}{1-\sigma})]^{1-\sigma}\end{aligned}$$

Taking the log on both sides, we can have the Nested Logit closed-form expression:

$$\begin{aligned}\log s_{jt} - \log s_{0t} &= \delta_{jt} + \sigma \log(s_{jt|g}) \\ &= X_j\beta - \lambda(t - r_j) + \tau_t + \sigma \log(s_{jt|g}) + \xi_j\end{aligned}$$

Appendix C

CHAPTER 3: OPTIMAL RELEASE TIME IN SEQUENTIAL DISTRIBUTION: THE WINDOWING STRATEGY OF THE U.S. MOTION PICTURE INDUSTRY

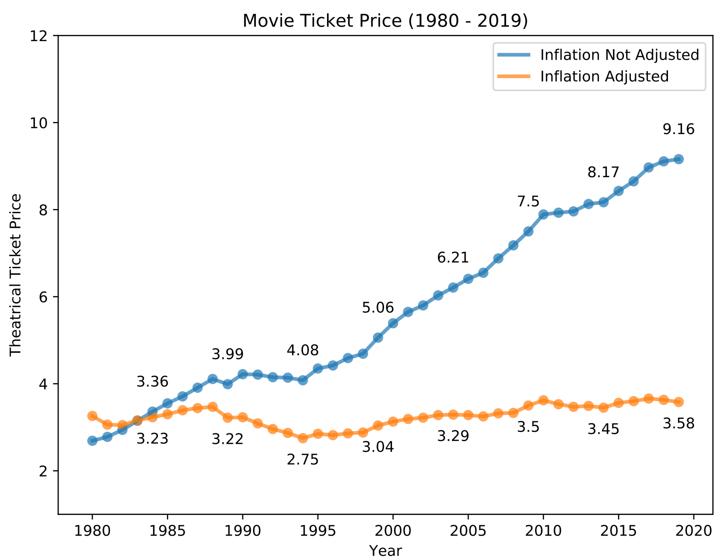


Figure C.1: Average U.S. Ticket Prices

Figure C.1 shows the annual average U.S. movie ticket prices from 1980 to 2019. Although the nominal prices keep increasing over the years, the inflation-adjusted ones remain stable. The price data are from NATO (The National Association of Theatre Owners). CPI data are used for inflation adjustment.

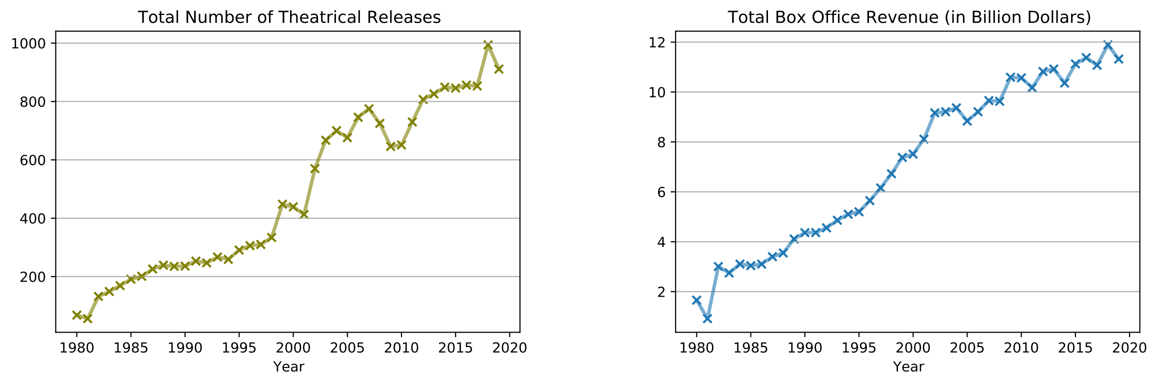


Figure C.2: Growth in The Industry

Figure C.2 shows that the U.S.motion pictures industry keeps expanding over the years. More products are produced (left panel) and higher revenue is generated (right panel).

Table C.1: Number of U.S. Movie Screens and Cinema Sites

Year	Indoor Screens	Drive-in Screens	Total Screens	Indoor Sites	Drive-in Sites	Total Sites	Ticket Price
2019	40613	559	41172	5548	321	5869	9.16
2018	40313	524	40837	5482	321	5803	9.11
2017	39651	595	40246	5398	349	5747	8.97
2016	40009	595	40604	5472	349	5821	8.65
2015	39411	595	40006	5484	349	5833	8.43
2014	39356	656	39956	5463	393	5856	8.17
2013	39368	656	40024	5326	393	5719	8.13
2012	39056	606	39662	5317	366	5683	7.96
2011	38974	606	39580	5331	366	5697	7.93
2010	38902	618	39520	5399	374	5773	7.89
2009	38605	628	39233	5561	381	5942	7.50
2008	38201	633	38834	5403	383	5786	7.18
2007	38159	635	38794	5545	383	5928	6.88
2006	37765	650	38415	5543	396	5939	6.55
2005	37040	648	37688	5713	401	6114	6.41
2004	35795	640	36435	5629	402	6031	6.21
2003	35016	634	35650	5700	400	6100	6.03
2002	35022	666	35688	5712	432	6144	5.80
2001	34823	683	35506	5813	440	6253	5.65
2000	35696	683	36379	6550	442	6992	5.39
1999	36448	683	37131	7031	446	7477	5.06
1998	33418	750	34168	6894	524	7418	4.69
1997	31050	815	31865	6903	577	7480	4.59
1996	28905	826	29731	7215	583	7798	4.42
1995	26995	848	27843	7151	593	7744	4.35

Table C.1 shows the number of U.S. movie screens and cinema sites.

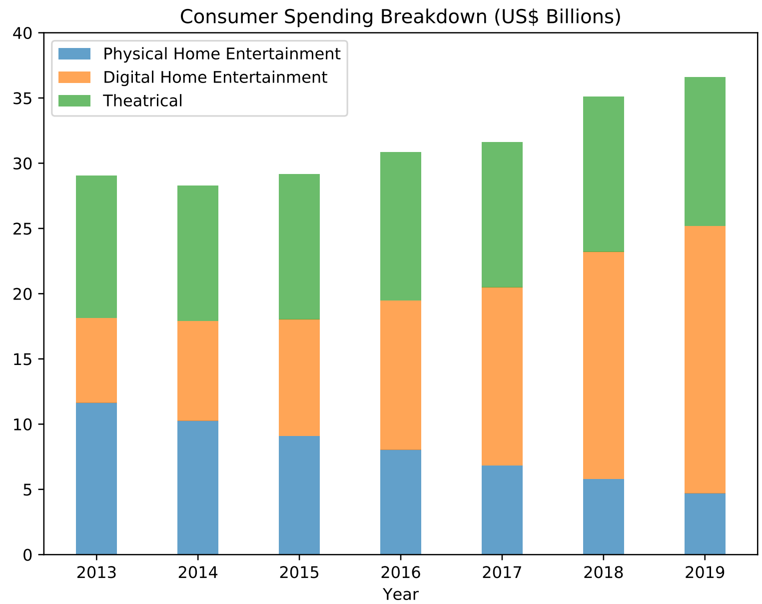


Figure C.3: Consumer Spending Breakdown

Figure C.3 shows consumers' spending breakdown on theatrical, physical home video, and digital home video markets. Digital home entertainment includes Electronic Sell-Through (EST), Video-On-Demand (VOD), and paid subscription streaming. Physical home entertainment includes rental and retail consumer spending. Over the years, consumers' spending on digital home videos is significantly expanding and the total spending on home entertainment exceeds the theatrical revenue. The data source is MPAA 2018 Theatrical Home Entertainment Market Environment (THEME) Report.