

UCLA

UCLA Electronic Theses and Dissertations

Title

Essays on Information and Firm Behavior

Permalink

<https://escholarship.org/uc/item/7bz261t6>

Author

Kwok, Siwei

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Essays on Information and Firm Behavior

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Economics

by

Siwei Kwok

2016

© Copyright by
Siwei Kwok
2016

ABSTRACT OF THE DISSERTATION

Essays on Information and Firm Behavior

by

Siwei Kwok

Doctor of Philosophy in Economics

University of California, Los Angeles, 2016

Professor Connan Andrew Snider, Chair

This dissertation consists of two chapters. In the first chapter, which is co-authored with Marcus Studart, we study the interaction of information and competition in incentivizing firms to produce high quality. We estimate a discrete quality choice game, using restaurant hygiene inspection data in Los Angeles County, from 1995 through 1998. Our results show that information is sufficient for competition to have an effect on quality provision. We also find that after the mandatory disclosure of information to consumers, a restaurant's equilibrium quality increases in the number of competitors up to a certain threshold. Beyond this threshold, an additional firm has a negative effect on quality choice. This result may indicate that too much competition reduces the returns of quality provision, which we interpret as a result of firms having more difficulty retaining customers.

In the second chapter, I investigate the notion of “hype” in the U.S. motion picture industry, which occurs whenever an upcoming film is heavily advertised irrespective of its underlying quality. I address the questions of 1) whether hype is a measurable phenomenon in data from the movie industry, 2) why the practice of hype is feasible, and 3) how the feasibility of hype impacts the movie industry. I first search for evidence of hype using a data set of weekly advertising expenditures, box office revenues, critical review outcomes, and movie characteristics of wide release theatrical films from 2003 to

2012. Using critical review outcomes as a measure of underlying quality, I observe that advertising levels are similar across film quality levels. I find that advertising levels fall dramatically between the pre-release and post-release phases for all movies, but they decay much more sharply for low quality films.

Building on this finding, I exploit weekly variation in the data by using regression analysis and a matching difference-in-differences estimator to quantify the causal effect of release and high critical reviews on advertising behavior. I find that advertising prior to a film's release is not statistically related to underlying film quality, but distributors with movies revealed to be high quality spend 62% more on advertising after release than other films. I then build a theoretical model of advertising on the movie industry based on Butters (1977) which incorporates consumer verification of film quality as a key feature to support this result. I provide conditions that ensure the existence and uniqueness of an advertising equilibrium in the model. I compute numerical examples of advertising equilibria and show that under certain parameter values, the distributors of bad films have no incentive to advertise if consumers are sufficiently informed of quality. These findings suggest that consumer access to critical reviews could explain the feasibility of hype in the movie industry.

The dissertation of Siwei Kwok is approved.

Shi Zhang

Hugo Andres Hopenhayn

Jinyong Hahn

Connan Andrew Snider, Committee Chair

University of California, Los Angeles

2016

Dedicated to my father

TABLE OF CONTENTS

1	Quality Disclosure and Competition: Evidence from the Los Angeles Restaurant Market	1
1.1	Introduction	1
1.2	Model	4
1.2.1	Environment	4
1.3	Data	6
1.4	Estimation	10
1.4.1	Identification	12
1.4.2	Estimation Model	13
1.5	Results	16
1.6	Conclusion	19
2	Advertising and ‘Hype’ in the U.S. Movie Industry	25
2.1	Introduction	25
2.2	Industry	29
2.3	Data	31
2.4	Evidence of hype	33
2.4.1	Regression Estimation	35
2.4.2	DID Estimation	38
2.5	Theoretical Model	42
2.5.1	Timing	42
2.5.2	Film Quality	43
2.5.3	The Firm’s Decision	44

2.5.4	The Consumer's Decision	45
2.5.5	Film Studio Revenue	48
2.5.6	Profit Maximizing Advertising Levels	50
2.5.7	Equilibrium	51
2.5.8	Numerical Solution Example	55
2.5.9	Continuous Types	56
2.5.10	Two Periods	57
2.6	Conclusion	59
Appendix	82
A.1	Asymptotic Variance of Two Step Estimator	82
B.2	Compilation of Movie Industry Data	85
References	88

LIST OF FIGURES

1.1	Average hygiene score plotted against average per capita income in each zip code before the law (June 1995-Dec 1997)	7
1.2	Average hygiene score plotted against the number of restaurants in each zip code before the law (June 1995-Dec 1997)	9
1.3	Average hygiene score versus average per capita income in each zip code after the law (Jan 1998-Dec 1998	10
1.4	Average hygiene score versus the number of restaurants in each zip code after the law (Jan 1998-Dec 1998)	11
2.1	Distribution of Rotten Tomatoes Scores (Wide Releases)	63
2.2	Advertising Expenditures vs. Rotten Tomatoes Score (Wide Releases) . .	64
2.3	Distribution of Opening Week Advertising Decay for $q \leq 30$ (Wide Releases)	65
2.4	Distribution of Opening Week Advertising Decay for $q \geq 50$ (Wide Releases)	66
2.5	Pre-release Advertising Expenditures vs. Score (Limited then Wide Releases)	67
2.6	Number of Pre-release Critical Reviews vs. Score (Wide Releases)	69
2.7	Proportion of Pre-release Critical Reviews vs. Score (Wide Releases) . .	70
2.8	Total Number of Critical Reviews vs. Score (Wide Releases)	71
2.9	Total Number of Critical Reviews vs. Budget (Wide Releases)	72
2.10	Advertising Strategies vs. Verification Probability (Numerical Solution 1)	80
2.11	Advertising Strategies vs. Verification Probability (Numerical Solution 2)	81

LIST OF TABLES

1.1	Thresholds Used in Estimation: L1, L2 and L3 denotes the inspection score beyond which the restaurant is considered as providing high quality. . . .	12
1.2	1st Stage Estimates (Specification (1))	21
1.3	2nd Stage Estimates (Specification (1))	22
1.4	1st Stage Estimates (Specification (2))	23
1.5	2nd Stage Estimates (Specification (2))	24
2.1	Summary Statistics - Wide Releases	62
2.2	Summary Statistics - Limited Releases (Followed by Wide)	62
2.3	Summary Statistics - Top Critical Review Publications	68
2.4	OLS Regression Results	73
2.5	IV Regression Analysis Results - Instrumenting for Score	74
2.6	IV Regression Analysis Results - Instrumenting for Budget	75
2.7	IV regression results, with number of published reviews used as instrument for film's production budget. Advertising expenditure is measured as pre-release or post-release total. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.	75
2.8	Balance test of covariates of propensity score method	76
2.9	Propensity Score of Treatment (Receiving $\theta > 50$)	77
2.10	Propensity Score of Treatment (Receiving $\theta > 50$) with Opening Week Box Office Revenue	78
2.11	DID Estimation Results	79

ACKNOWLEDGMENTS

I am indebted to my advisor Professor Connan Snider for his guidance and encouragement throughout my Ph.D. career at UCLA. I am thankful for his wisdom from both his teaching as my professor and advising as my mentor, and I am especially grateful for his moral support whenever I faced adversity. I am also thankful to Professors Jinyong Hahn, Hugo Hopenhayn, and Shi Zhang for their guidance over the years as my professors and members of my committee. I am also deeply grateful to Professors Simon Board and Moritz Meyer-ter-Vehn for their guidance and encouragement, especially during my final years at UCLA.

I am thankful to Professors Ginger Jin and Phillip Leslie for their helpful comments on the first chapter and to Mike Rossi for his input on the second chapter. I am especially grateful to Ginger Jin for her mentorship while I was a student at the University of Maryland at College Park, which propelled me to pursue a Ph.D. in economics.

I am thankful to Tiago Caruso, Sangyup Choi, Sean Fahle, Andraz Kavalari, Joseph Kuehn, Dennis Kuo, Matt Miller, Ravaris Moore, Amanda Nguyen, Ana Luisa Pessoa Araujo, Mario Silva, Jesper Sørensen, Marcus Studart, Allen Tran, Kyle Woodward, Richard Xie, and Menghan Xu, from whom I have received great support, strong friendship, and fond memories during my time at UCLA.

Lastly, I am thankful to my mother for all of her love and support.

VITA

- 2007–2008 Student Researcher, NCBI/NLM, National Institutes of Health. Worked as an intern under the auspices of Dr. Yi-Kuo Yu during summers. Coded data manipulation programs for a peptide database search research project. Generated and polished figures in manuscript for publication. Coded numerical programs for biostatistics research project involving hypothesis testing in the fitting of random distributions.
- 2009 B.S. (Mathematics) and B.A. (Economics), University of Maryland at College Park.
- 2009 Summer Intern, Bureau of Economics, Federal Trade Commission. Worked closely with Dr. Laura Hosken. Coded data analysis programs for study on wind turbines and energy production. Designed database form and collected data for study on Virtual Worlds.
- 2010 M.A. (Economics), UCLA.
- 2010–2014 Teaching Assistant, Economics Department, UCLA. Taught sections of Economics 1 (Principles of Microeconomics, 101 (Microeconomic Theory), 106T (Economics of E-commerce and Technology), 106G (Introduction to Game Theory), and 170 (Industrial Organization: Theory and Tactics) under Professors Anne Bresnock, William Zame, Connan Snider, Simon Board, Michael Sproul, Moritz Meyer-ter-Vehn, and Edward Kung.
- 2012–2013 Instructor, Economics Department, UCLA. Taught courses on microeconomic theory during Summer Session A in 2012 and Summer Session C in 2013.

2013 Instructor, Economics Department, UCLA. Taught an introductory course to macroeconomics during Fall Quarter.

PUBLICATIONS

Alves G, Ogurtsov AY, Kwok S, Wu WW, Wang G, Shen RF, Yu YK (2008) Detection of co-eluted peptides using database search methods. *Biol. Direct*, 3:27

CHAPTER 1

Quality Disclosure and Competition: Evidence from the Los Angeles Restaurant Market

1.1 Introduction

Quality disclosure schemes such as mandatory quality ratings have become an increasingly popular approach to ensuring informed consumer choices and incentivizing firm quality investment. The context for these systems are varied, ranging from the performance of public schools to the mortality rate in dialysis centers. There is also a burgeoning academic literature evaluating the effectiveness of these programs in influencing consumer choice (Ippolito and Mathios (1990), Jin and Sorensen (2006)), incentivizing firm quality investment (Powers et al. (2008), Benneer and Olmstead (2008)), and as well as attempts to manipulate or game the system (Dranove et al. (2003)).

Standard models of oligopolistic competition suggest that the optimal design and the resulting efficacy of information disclosure schemes may depend crucially on the structure of competition present in the markets in which the systems are implemented. In this paper we present the first study, to our knowledge, of the interaction between market structure and mandatory quality disclosure schemes. To do this, we revisit the results of Jin and Leslie (2003) who demonstrate that a new mandatory quality disclosure law, which required restaurant to prominently display hygiene inspection grade cards to consumers, caused an increase in restaurants' hygiene quality levels. They also find that the quality disclosure also caused consumer demand to become more sensitive to changes in hygiene score, and caused foodborne illness hospitalizations to decrease. Shih uses the same data

in conjunction with locational data to demonstrate that restaurants under a voluntary hygiene disclosure policy improved hygiene more when located near a restaurant with a mandatory disclosure policy .

In this paper, we determine what factors induce firms to provide higher hygiene quality. We estimate a game of incomplete information, where the payoffs of restaurants are functions of competitors' quality choice, as well as covariates including the demographics of the market and restaurant characteristics. We model two static games: one before and one after the hygiene quality disclosure scheme is implemented. We proceed in this way because we do not have access to revenue or prices, therefore we choose not to directly model consumer choice and, instead, consider how the new scheme affects the payoff structure of firms. We use a two step estimation method as in Bajari, Hong, Krainer and Nekipelov (2006). The method breaks the estimation into a reduced form first step estimation and uses the fitted value probabilities to estimate a standard discrete choice model in the second step. At their core, our results decompose the average treatment effects found in Jin and Leslie (2003) into heterogeneous effects that depend on the characteristics of local competition and the market.

Our results show that initially the quality choice of competitors had a negligible effect on a restaurant's own quality choice, although on average having more competitors increased the probability of providing higher hygiene quality. The effect changes after the law, when the results of previous hygiene inspections are mandatorily disclosed to the consumer. Now competitors' choice have a positive effect with magnitude ten times greater than before, which means that if firm i believes that its competitors will provide high quality, it will be optimal for i to provide high quality too. The intuition is that before the law, consumers did not have easy access to hygiene quality information, except possibly a noisy signal by word-of-mouth from other consumers. Since the probability of realizing the true hygiene quality is small for the consumer, firms do not have incentives to invest in it. After the law, the disclosure of information boosts competition, since the consumer can punish low hygiene practices by going to a competitor that offers higher

hygiene standards.

The first step estimation reveals that, although the average payoff does not include the competitors high quality choice, the total number of firms in the zip code area had a positive effect in quality provision in both periods. The effect is nonlinear and becomes more intense when restaurants start displaying the grade cards. In markets with a small number of firms, an additional restaurant increases the probability that all other firms will provide high quality. The effect is positive until we reach some threshold, after which additional firms have negative effects on quality provision. Although the number of firms in the market is not a perfect proxy for competition intensity, the result is interesting since we do not know any theory paper, other than Bar-Isaac (2005), which explores the subject ¹. We provide a possible source of estimation bias, but the intuition relies on the fact that more competition allows consumers to more easily switch from low quality firms to higher quality options, but as the number of firms becomes large, competition becomes too fierce and it is harder to keep customers with investments in quality. The investment is as costly as before, but the expected return on reputation investment is lower. In all, these results not only confirm the findings of Jin and Leslie (2003), but also suggest that the pathway through which information disclosure incentivizes restaurants to invest more in the quality is competition. With this consideration, policymakers can design better quality disclosure schemes that seize upon potential social welfare gains from information to consumers.

Our paper is structured in the following way. In Section 2 we describe the model structure and define an equilibrium of the incomplete information game used. Next we describe the two step approach to estimate discrete games and describe identification issues. In Section 3, we describe the restaurant data used and specify functional forms in the first and second steps of estimation. In Section 4, we describe the estimation process

¹The author provides a specific example of a demand function from Sutton (1991) where the number of firms has a nonlinear effect on reputation incentives. However his example generates a different result: incentives are high when the number of firms are either small or big. At intermediate levels the incentives to build reputation for quality is small.

and discuss results in Section 5. In Section 6, we conclude our paper. We also include in the appendix the derivation of the estimator’s asymptotic distribution.

1.2 Model

Although the data set has a panel structure from 1995 to 1998, we choose to model the game as a static one, because neither revenue nor price information were available. As a result, our best option was to treat consumer behavior as a primitive of the game. We therefore chose to focus our model on the firms’ choices, specifying a restaurant’s profit to be a function of its quality choice and several covariates from the data, which serves as a proxy for what determines consumer behavior. In addition, we note that there is a structural break in the data at the end of 1997, which was when the new hygiene information disclosure law was adopted in Los Angeles County. If consumer behavior changed as a result of this event, we should observe changes in the primitives of our model. To look for these changes, we decided to estimate two different static games—the first played before the law was passed and the second played afterwards—and compare parameter estimates.

1.2.1 Environment

In each market $m = 1, \dots, M$, there is a finite number of restaurants denoted $i = 1, 2, 3 \dots n_m$ that each have a unique manager. Throughout this paper, the words manager and restaurant will be used interchangeably. The managers simultaneously choose an action $a_i \in \{L, H\}$ to maximize the payoff to their respective restaurants, where L and H represent low and high hygiene quality levels respectively. We define $A = \{L, H\}^{n_m}$ to be the set of all possible combination of actions involved in the game and $a = (a_1, a_2, a_3, \dots, a_{n_m})$ to be an element of this set. The variable $s_i \in S_i$ denotes the state for restaurant i and $s = (s_1, s_2, s_3, \dots, s_{n_m}) \in S$ denotes the vector of all restaurants states, where $S = \times S_i$.

The state s is common knowledge to the players and observable to the econometrician. However, associated with each restaurant action is another state variable $\varepsilon_i(a_i)$, which is private information only to the manager of restaurant i . We assume the private states to be identically and independently distributed across restaurants with density $f(\varepsilon_i(a_i))$ with support E . The payoff to the manager i has the following additive form:

$$U_i(a, s, \varepsilon_i) = \pi_i(a_i, a_{-i}, s) + \varepsilon_i(a_i) \quad (1.1)$$

The manager's payoff consists of an observable part which is common knowledge to all players and a privately observed state. The value of the observable part depends on the actions of all players in the market and the realized state $s \in S$ of the game. The optimal policy of the manager for restaurant i is a mapping $g : S \times E \mapsto \{H, L\}$, so that his decision does not depend on ε_{-i} since it consists of privately known information of the other players. Given this function and knowledge about the structure of game, the other players are able to infer the probability that manager i chooses H as the hygiene quality conditional on the observable state s in the following way:

$$\sigma_i(a_i = H|s) = \int_{\{\varepsilon_i \in E: g(s, \varepsilon_i) = H\}} f(d\varepsilon_i) \quad (1.2)$$

Using Equation 1.2 we can calculate the conditional probability of observing action $a \in A$ given the state s . Let $\sigma(a|s) = \times_{i=1,2,3..n_m} \sigma_i(a_i|s)$. Using the conditional distribution of the actions of the $-i$ players, we can compute the expected utility of manager i 's choice $a_i \in \{H, L\}$ conditional on the state s .

$$E_{\sigma_{-i}} [U_i(a, s, \varepsilon_i)|s] = \sum_{-i} \sigma_{-i}(a_{-i}|s) \pi_i(a_i, a_{-i}, s) + \varepsilon_i(a_i) \quad (1.3)$$

Then, the optimal decision for agent i given s is to choose an action in $\{H, L\}$ such that the conditional expected utility given by Equation 1.3 is maximized. The maximization problem is the following:

$$g_i(s, \epsilon_i) = \operatorname{argmax}_{a_i \in \{H, L\}} \{E_{\sigma_{-i}} [U_i(a, s, \epsilon_i) | s]\} \quad (1.4)$$

The probability that manager i chooses high quality is calculated in the following way:

$$\begin{aligned} \sigma_i(a_i = H | s) = \\ P[\{\epsilon_i \in E : \sum_{-i} \sigma_{-i}(a_{-i} | s) \pi_i(H, a_{-i}, s) + \epsilon_i(H) \geq \sum_{-i} \sigma_{-i}(a_{-i} | s) \pi_i(L, a_{-i}, s) + \epsilon_i(L)\}] \end{aligned} \quad (1.5)$$

We define an equilibrium of the hygiene quality choice game in the following way:

Definition An equilibrium is a set of policy functions $\{g_i\}_i$ and choice probabilities $\{\sigma_i\}_i$ such that for any player i , given the policy function of the other players g_{-i} , the inferred choice probabilities σ_{-i} are given by Equation 1.2, g_i is the solution to the maximization problem in Equation 1.4 and σ_i is given by Equation 1.5.

The existence of such an equilibrium has been proven in previous papers including McKelvey and Palfrey (1995), but the equilibrium may not be unique. We address this important question in the next subsection.

1.3 Data

We use panel data on restaurants in Los Angeles County from Jin and Leslie (2003) to estimate our model. Each observation in the data set is the result of an official hygiene inspection of a restaurant. Along with the date of and the resulting score from the inspection, the data include census tract demographic information and a significant number of restaurant details. The census data include information about average income, ethnic composition, and socioeconomic makeup in each census tract. The restaurant informa-

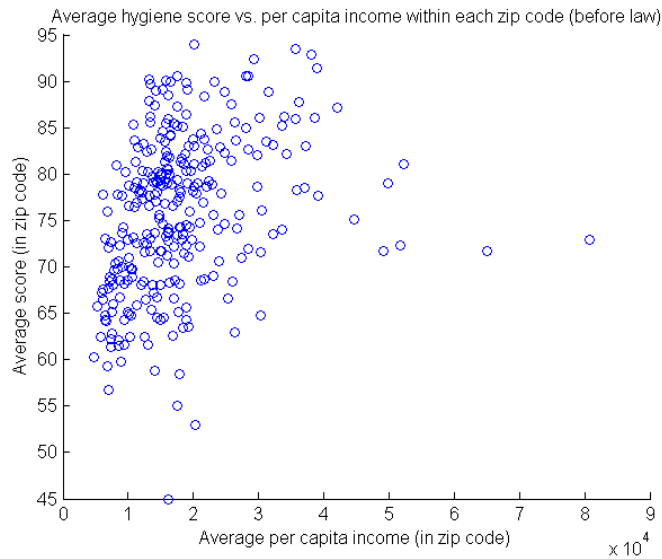


Figure 1.1: Average hygiene score plotted against average per capita income in each zip code before the law (June 1995-Dec 1997)

tion includes whether or not it has a license to sell alcoholic beverages and dummies to indicate whether the restaurant size is small, medium, or large. Other characteristics are whether or not the restaurant is part of a chain and the age of the restaurant. The data set covers hygiene inspections conducted from June 1995 to December 1998, and includes the type of inspection (e.g. if it was a routine inspection or initiated by a case of food poisoning). The inspections done from 1995 to 1997 were mostly initiated by routine inspections and the average number of inspections was 1.9 from 1995 to 1997 and 2.1 in 1998. The hygiene scores are on a scale from 0 to 100, but scoring criteria were changed during the sample period. Before July 1997, the inspections consisted of objective and subjective assessments. Under this design, the DHS inspector first deducts pre-specified points for each violation. For instance, a food temperature violation results in a deduction of 5 points, an improperly sanitized eating utensils violation is worth 5 points, and an indication of the presence of cockroaches is worth 3 points. Before July 1997, inspectors were also supposed to give a subjective assessment of the status of the establishment, which could be excellent (no deduction), average (20 points), fair (30 points) or poor (40

points). In the whole sample period, restaurants could be closed if they receive a score below 60 in two consecutive inspections or when there is a “severe” hygiene problem. In the period after the inspection procedure was redesigned to be more objective, we observe an immediate increase of scores by an average of 8.9 points. In view of this inspection standards discrepancy, we normalized the hygiene inspection scores after the redesign occurred in our estimation.

In December 1997 the Los Angeles County Board of Supervisors voted in favor of the grade card ordinance, which would come into effect in the next month to make the display of hygiene inspection results mandatory. Despite this legislation, it was the cities themselves that had the authority to determine when the law would take effect. It took several months until all cities in the county had implemented the law. Before then, restaurants were free to display the grades voluntarily if their city had not implemented the law yet. The grades are displayed on a letter scale, where “A” represents a hygiene score between 90 and 100, “B” represents one between 80 and 89, “C” represents one between 70 and 79 and if the score is less than 70 the restaurant is issued a card that reports the actual number score. The data initially contain hygiene inspection results for 24304 restaurants, but after dropping the ones with missing census data and dropping restaurants that were listed as being a bar or a cafe (e.g. Starbucks), we were left with a total of 16673 restaurants. The rationale behind dropping the bars and cafes is that those types of establishments are not in direct competition with traditional restaurants. All of the demographic data come from the U.S. Census, where the units of observation are census tracts, regions in general smaller than zip code areas. The total number of census tracts and zip codes are 1373 and 304 respectively. We have raised demographic data from the census tracts to the zip code level, because our study analyzes the strategic interaction between restaurants in providing hygiene quality. The census tract cannot be used as a unit level of analysis because due to its small size, restaurants in some census tract are likely to be competing with restaurants from another census tract.

A good proxy for the level of competition between two restaurants would be the pair’s



Figure 1.2: Average hygiene score plotted against the number of restaurants in each zip code before the law (June 1995-Dec 1997)

distance or a variable to designate food style similarities. Unfortunately, the locations of restaurants are not available to us and most of the data set is missing the restaurant's style or food type. We instead make the assumption that a zip code is a market and that every restaurant is a direct competitor of the others in the same zip code. A consequence of this assumption is that the number of players in one instance of the game can exceed 100. This fact reinforces our need for a specific parameter choice to identify the model as discussed in the last section.

We merged the information of all census tracts contained under a zip code to be the demographics of that particular zip code. The zip code per capita income is computed as the average of per capita income of the census tracts in the area. We do so because neither the population of each census tract nor an identifier of the census tract is available, so we assume that the census tracts' coverage is divided evenly within a zip code.

We also calculated correlations between the scores, number of restaurants and average income in the zip codes. Before the implementation of the law, the correlation between average hygiene score and per capita income in a market was 0.322. A scatterplot of the

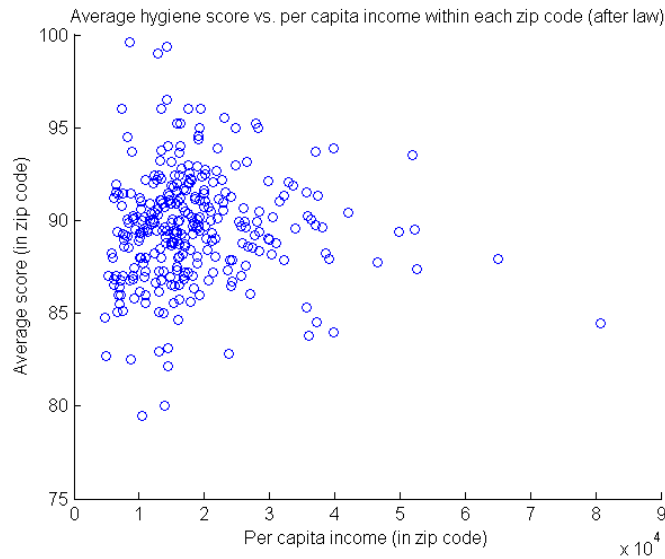


Figure 1.3: Average hygiene score versus average per capita income in each zip code after the law (Jan 1998-Dec 1998)

average scores as a function of per capita income is given in Figure 1.1.

The correlation between average hygiene scores and the number of restaurants in a market in the sample before the law is -0.145 . A scatterplot of the average scores as function of the number of restaurants is given in Figure 1.2. After the law was implemented, correlation between average income and average hygiene score in a market is reduced to 0.03 , while the correlation between average scores and the number of restaurants is more negative than before, -0.236 . We repeat the graphs for the sample after the law in Figures 1.3 and 1.4.

1.4 Estimation

We estimate two sets of primitives for the two different games observed, where one is played before and the other after the law is implemented. In the first game, the consumers do not observe the scores, and have to rely on learning from past experiences to choose which restaurant to draw services from. We consider hygiene inspections from

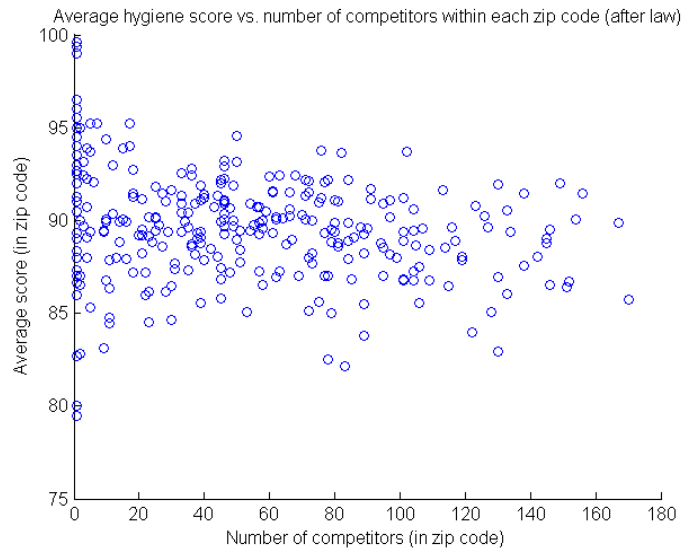


Figure 1.4: Average hygiene score versus the number of restaurants in each zip code after the law (Jan 1998-Dec 1998)

the beginning of the data set to June 30, 1997. We exclude inspections from July 1997 to when the law was passed, because the inspection standards changed in this period, and because we believe that the news investigations (which inspired the law) may have temporarily changed the game dynamics, creating noise in our estimation.

In the second game, the local government provides direct information to consumers about recent hygiene practices, so consumers can condition their choices on the information available. If consumers start making their decisions on that information, the restaurant payoff function should change. We consider all hygiene inspections from January 1998 to the end of our data set, including restaurants that did not post their hygiene score because their cities had not started enforcing the law. These restaurants were included, because they account for only 1,002 out of 32,273 hygiene inspections in this period of time. Our main questions are what variables determine quality, what is the importance of competition to quality provision, and how the law affected the game equilibrium.

Throughout this paper, we assumed that hygiene quality is a discrete choice variable with only two values, high and low. This assumption simplifies enormously the estimation

	L1	L2	L3
Before Law	70	75	80
After Law	78.9	83.9	88.9

Table 1.1: Thresholds Used in Estimation: L1, L2 and L3 denotes the inspection score beyond which the restaurant is considered as providing high quality.

process. Although we could have a greater number of choices, we kept the choice to be binary for simplicity, even though we believe that the results would not be quite different.

Our estimation depends on actions $a_i \in \{H, L\}$ which we do not observe. Our solution was to use the hygiene score as a proxy for quality. We choose a threshold \bar{L} and establish a rule such that if the restaurant's recent average score is below \bar{L} , the restaurant was producing a low quality product. Since our threshold choice is ad hoc, we use three different thresholds for each game, resulting in a total of 6 estimations. Because after July 1997 the grades were inflated by an average of 8.9 points, we adjusted the threshold values for the estimation after the law. The smallest threshold in the game before the law is 70, the second 75 and the third 80 which lead to respective thresholds after the law of 78.9, 83.9, and 88.9, as displayed in the 1.1 below.

1.4.1 Identification

Before describing the estimation method, we discuss identification issues. Our main objective is to estimate $\{\pi_i\}_{i=1,2,3..n_m}$ which consists of the primitives of the game. We will say the model is identified if different values for the primitives generate different choice probabilities. This allows us to uniquely recover the functions $\{\pi_i\}_{i=1,2,3..n_m}$ from the choice probabilities, as long as the distribution of the private state variable is known. We normalize $\pi_i(L, a_{-i}, s) = 0$ and assume the difference between the two error terms in Equation 1.5 is generated from a type I extreme value distribution. The normalization is necessary for identification and the extreme value distribution assumption is convenient for approaching this problem using maximum likelihood estimation. Inverting the

distribution, we form a system of n_m equations for each s as below.

$$F^{-1}(\sigma_i(a_i = H|s)) = \sum_{-i} \sigma_{-i}(a_{-i}|s)\pi_i(H, a_{-i}, s) \quad (1.6)$$

If we assume the choice probabilities to be known, we can solve for the sequence $\{\pi_i\}_{i=1,2,3..n_m}$, but not uniquely since there are $n_m 2^{n_m-1}$ unknowns for each state. In order to make the model identified, we need exclusion restrictions so that we have at least the same number of equations as unknowns. A common way to achieve this is to restrict the payoff function π_i to depend only on s_i and not s_{-i} . If the data have enough variation in the states, such that s_{-i} conditional on s_i has at least 2^{n_m-1} points, the model is non-parametrically identified. Unfortunately our model has many players and the limited extent of the data does not allow us to non-parametrically identify the model. Our approach is instead to specify a parametric form and make exclusion restrictions, the details of which we describe in the next subsection.

1.4.2 Estimation Model

In our model, restaurant i enjoys a utility level which depends on whether it chooses the $a_i = H$ or $a_i = L$ and is linear in the observable state and the aggregate decisions of competing restaurants. This is captured in the following specification:

$$\pi_i(a_i, a_{-i}, s) = \begin{cases} s'_i \alpha^1 + \delta^1 \sum_{j \neq i} I(a_j = H) & \text{when } a_i = H \\ s'_i \alpha^2 + \delta^2 \sum_{j \neq i} I(a_j = H) & \text{when } a_i = L \end{cases} \quad (1.7)$$

With the utility specification as shown in Equation 1.7, we cannot identify the parameters α^1 , α^2 , δ^1 , or δ^2 separately. The best we can do is to estimate a model with a linear form commonly used in firm entry games as in Berry (1992) given by:

$$\pi_i(a_i, a_{-i}, s) = \begin{cases} s'_i \alpha + \delta \sum_{j \neq i} I(a_j = H) & \text{when } a_i = H \\ 0 & \text{when } a_i = L \end{cases} \quad (1.8)$$

where $\delta = \delta^1 - \delta^2$ and $\alpha = \alpha^1 - \alpha^2$.

Plugging in Equation 1.8 into Equation 1.3 along with the extreme value distribution assumption on the unobservable terms yields the probability of agent i choosing high hygiene quality given the state s , which is given by:

$$\sigma_i(a_i = H|s) = \frac{\exp(s'_i \alpha + \delta \sum_{j \neq i} \sigma_j(a_j = H|s))}{1 + \exp(s'_i \alpha + \delta \sum_{j \neq i} \sigma_j(a_j = H|s))} \quad (1.9)$$

We estimate the model above using a two step method approach as in Bajari, Hong, and Nekipelov (2004) and Bajari et al. (2006). This method is as efficient as a one step method but is computationally faster. The two step method consists of forming a reduced form model in the first step, yielding consistent estimators for the choice probabilities $\{\sigma_i\}$. We decided to use probit estimation for this step. In the second step we use the fitted values $\hat{\sigma}_i(a_i = H|s)$ to estimate the primitives of the model, using a binary choice estimation. To do this, we find the parameters that maximize the following likelihood function:

$$L(\beta) = \prod_{m=1}^M \prod_{i=1}^{n_m} \left(\frac{\exp(x_{i,m})}{1 + \exp(x_{i,m})} \right)^{I(a_{i,m}=H)} \left(1 - \frac{\exp(x_{i,m})}{1 + \exp(x_{i,m})} \right)^{I(a_{i,m}=L)} \quad (1.10)$$

where $x_{i,m} = s'_i \alpha + \delta \sum_{j \neq i} \sigma_j(a_j = H|s)$.

The estimation is consistent and gives efficient estimators unless agents play different equilibria across the markets. Since multiple equilibria is a possibility, we make the strong assumption that players coordinate on the same equilibrium in the sample.

In this first step we use the following two specifications for the probit model:

Specification 1:

$$\begin{aligned} \sigma(a_i = H|s) = & \Phi(\gamma_{INC}INC_i + \gamma_{AL}AL_i + \gamma_{NIC}NIC_i + \gamma_NN_i \\ & + \gamma_{PAL}PAL_i + \gamma_{PNIC}PNIC_i) \end{aligned} \quad (1.11)$$

Specification 2:

$$\begin{aligned} \sigma(a_i = H|s) = & \Phi(\gamma_{INC}INC_i + \gamma_{AL}AL_i + \gamma_{NIC}NIC_i + \gamma_NN_i + \gamma_{N^2}N_i^2 \\ & + \gamma_{PAL}PAL_i + \gamma_6PAL_i^2 + \gamma_7PNIC_i) \end{aligned} \quad (1.12)$$

where the symbol Φ denotes the standard normal cumulative distribution function.

The variable INC_i denotes average income per capita in the zip code area of restaurant i , AL_i is a indicator variable which is equal to 1 if the restaurant has a license to sell alcoholic beverages, and NIC_i is a binary variable which is equal to 1 if the restaurant is not part of chain. The variable N_i is the number of restaurants in the zip code area and lastly PAL_i and $PNIC_i$ are the percentages of restaurants that have alcohol license and are not part of a chain in the zip code area respectively. The reduced form estimation above uses the each zip code as a repetition of the (same) game with differing numbers of players. We view each market as a repeated game because the cost structure of maintaining hygiene levels in a restaurant should be the same across markets. Since the number of players is not constant throughout the zip code areas, the number of restaurants is included as one of the explanatory variables.

After estimating the first step, we use the fitted values of the choice probabilities $\hat{\sigma}_i(a_i = H|s)$ as inputs in the likelihood function in Equation 1.10. The linear specification with the actual variables used in our regressions is given by:

$$\pi_i(a_i, a_{-i}, s) = \begin{cases} \alpha_1 INC_i + \alpha_2 AL_i + \alpha_3 NIC_i + \delta \sum_{j \neq i} I(a_j = H) & \text{if } a_i = H \\ 0 & \text{otherwise} \end{cases} \quad (1.13)$$

We note that only income per capita, whether the restaurant has a liquor license, and whether the restaurant is part of a chain are explanatory variables. The other regressors used in the first step, such as number of competitors in the neighborhood, percentage of restaurants with license, and percentage of chain restaurants are not included in the second step, which is a part of our identification strategy. If the first stage estimates $\hat{\sigma}_i(a_i = H|s)$ and the term $s'\alpha$ in Equation 1.8 depend on $s \in S$, our regression would suffer from severe collinearity, which would make identifying the separate effects of the vector $(\alpha, \delta)'$ on the choices probabilities difficult. The problem is similar to a standard simultaneous equation model, meaning that exclusion restrictions are necessary to identify parameters.

It seems intuitive that neither the number of competitors, the proportion of competitors with alcoholic license nor the proportion of independent restaurants should enter directly into restaurant i 's profit function. On the contrary, these variables indirectly affect the profits through competitive behavior, which is captured by the term δ . With these restrictions, our model is identified.

1.5 Results

The results from specification (1) are found in Tables 1.2 and 1.3. The parameter associated with competitors' hygiene quality choice δ is either negative and close to zero, or insignificant under any specification before the law. The intuition behind this result is that poor hygiene practices are mostly observable when a customer suffers from food poisoning, which should only happen with sufficiently small probability since the DHS shuts

down restaurants when the chance of contamination is high enough. If the probability of observing the true quality is small, the incentives to invest in high quality are smaller. Restaurant i should not care much about the actions of other restaurants, because choosing to produce high hygiene quality would have a small effect on its reputation. If more competitors are choosing high quality, the reputation gain to firm i choosing high quality is even lower, and may not be large enough to compensate for the cost of that choice. In this case, it is possible to have competing restaurants' high quality choices negatively affecting manager i 's probability of choosing high quality, as is observed in the estimates.

On the other hand, the new law dramatically changes the primitives of the payoff function. This can be explained by noting that the government now directly informs the consumer about the past quality choice of managers. If consumers believe that the grades are good proxies for the current hygiene practices, they will avoid restaurants with lower grades. Then the hygiene quality choices of the competitors directly affects restaurant i 's choice, because i can be punished with the loss of customers if it chooses quality lower than its competitors.

Our choice to have three different thresholds for high and low quality choices helps to clarify the last result. The estimate for δ after the law reduces from 0.0108 to 0.0075 and 0.005 as we increase the threshold from 78.9 to 83.9 and 88.9. The intuition is that having a grade below "C", when others have grades above, is much worse than having a grade lower than "B", when other restaurants have grades of "B" or "A".

Another variable of interest is per capita income in the market. In our estimation, the probability of high hygiene quality is increasing in income, although the effect of income diminishes after the law is implemented. As can be seen in Table 1.3, the estimates for $\hat{\alpha}_{INC}$ are always positive, but they decrease from 0.0836 to 0.0717 in the first, from 0.0593 to 0.0426 in the second, and from 0.0477 to 0.0285 in third, when the law implemented. Although we do not have data on other product dimensions, such as food quality, it seems reasonable to believe that wealthier zip code areas in the sample contain higher level restaurants. These tend to have higher prices and consequently a smaller customer

base. With a smaller clientele, the restaurant have more to lose when one client is not satisfied. This happens because first, the per client revenue is larger and second, negative information flows faster in wealthier neighborhoods since high level restaurants are frequently rated by professional surveys such as Zagat.

The main question we attempt to answer in this paper is how the degree of competition and incentives to produce quality are related. An ideal proxy for competition would be number of restaurants divided by some market size index, which would be some function of population and income. Unfortunately we do not have census data about population size in the zip codes areas, so our next best choice was to use the number of restaurants as our proxy for competition intensity. Our estimates for γ_N are negative in all specifications, both before and after the law. This means that an additional firm in the game reduces the probability that the current restaurants produce high hygiene. That does not seem to be compatible with the fact the δ estimate becomes strictly positive after the law. For this reason we also analyzed specification (2), which also includes in the payoff function the number of restaurants squared in the zip code.

The results of the specification (2) are found in Tables 1.4 and 1.5. The additional term in the first stage does not change the second stage estimates much, except for $\hat{\delta}$ which became slightly positive in the first two benchmarks before the law. Our interpretation does not change since the parameter is very close to zero in both of these benchmarks, and not significant in the second benchmark. This indicates that competition might have nonlinear effects on quality choice. Our results show that one additional firm has a positive effect on quality choice when the number of restaurants is below 92, in benchmark (1) and (3), and when below 69 in benchmark (2) before the law. After some threshold is achieved, an increase in the number of restaurants reduces the probability to produce high hygiene. We do not know of any theoretical papers that could help explain this empirical fact.

After the law was implemented, the negative effects of competition are reduced, since the effect of an additional firm in the zip code area is positive until the total number

reaches 152, 146 and 105 in benchmarks (1), (2) and (3) respectively. The regressions show that more competition provides incentives to invest in quality, but eventually the incentives are reduced after some threshold number of restaurants is reached. One possible explanation is that if a market is saturated by restaurants, that is, many firms serve a relatively small customer pool, it may become harder for restaurants to keep clients loyal to their businesses, what could reduce the returns of building a reputation for being clean.

The other parameters in the payoff function were indicators for if the restaurant has a license to sell alcohol, and if it is not part of a chain. On average, the alcohol license reduces the probability of restaurant choosing high quality (in all specifications), both before and after the law. However, before the law, chain restaurants were more prone to provide quality, which is reasonable since reputation of the whole brand is affected, when one of its affiliates starts shirking on quality. Surprisingly, after the law, the parameter changes sign in benchmark (1) and (2).

One possible source of estimation bias is that in some of our markets, there may be no reputational incentives for restaurants to invest in hygiene quality. For example, in areas where tourists are commonly found such as Hollywood or Universal City, we might expect consumers to be one-shot (i.e. tourists who eat at the restaurant once during their vacation and never return). In our estimation, we ignore this possibility because in every zip code, regardless of whether it's popular among tourists or not, there are residents who can be viewed as repeat customers of these restaurants.

1.6 Conclusion

In this paper, we have analyzed the effect of an oligopolistic market structure on the incentives to produce quality, in the context of quality disclosure. Our approach was to search for empirical evidence in favor of our view that a lack of competition reduces quality. We then estimated a discrete quality choice game using Los Angeles County restaurant hygiene inspection data, from 1995 to 1998. As a simplification, we analyzed

the restaurants' hygiene choice, despite the fact that the good they supply has many other dimensions, such as food quality, location, and service. This choice worked with our estimation strategy, because hygiene procedures can be treated as an homogeneous good, and the supply cost should be very similar for the restaurants in the sample. That permitted us to use each zip code as a realization of the game and to estimate the primitives of the restaurants payoff function.

Our results suggest that competition improves average hygiene quality and that the increase in the information flow to consumers strengthens this effect. It also shows that too much competition may reduce incentives for investment quality. We do not know any theory that explains this fact, but our intuition is that too much competition can make harder for firms keep clients, so that the incentives to invest in quality are reduced, or that consumers learn too slowly about individual firms in a saturated producer market. With these results, policymakers can better understand the mechanism through which information improves quality in markets and design optimal information disclosure schemes to maximize welfare gains.

Table 1.2: 1st Stage Estimates (Specification (1))

Period	Grade Threshold	<i>INC</i>	<i>AL</i>	<i>NIC</i>	<i>N</i>	<i>PAL</i>	<i>PNIC</i>	Log-likelihood
Before	70	0.0254***	-0.0894***	-0.6314***	-0.0028***	1.9559***	0.2715***	-8673.6
		(0.0011)	(0.0238)	(0.0388)	(0.0003)	(0.1239)	(0.0629)	
		0.0224***	-0.0919***	-0.5644***	-0.0025***	1.6280***	-0.1216**	-9726.2
Before	75	(0.0011)	(0.0228)	(0.0331)	(0.0003)	(0.1174)	(0.0572)	
		0.0262***	-0.0962***	-0.5693***	-0.0022***	1.2550***	-0.6267***	-9198.7
		(0.0012)	(0.0234)	(0.0310)	(0.0003)	(0.1229)	(0.0581)	
After	78.9	0.0006	-0.0686**	-0.5287***	-0.0021***	0.7573***	2.2057***	-3817.9
		(0.0017)	(0.0333)	(0.0777)	(0.0004)	(0.1813)	(0.1091)	
		-0.0002	-0.0402	-0.6399***	-0.0018***	0.8166***	1.5719***	-7258.8
After	83.9	(0.0014)	(0.0256)	(0.0485)	(0.0003)	(0.1399)	(0.0740)	
		0.0015	-0.0216	-0.7186***	-0.0016***	0.8910***	0.8422***	-9915.5
		(0.0013)	(0.0225)	(0.0350)	(0.0003)	(0.1229)	(0.0599)	

Notes: The period “Before” and “After” denotes the time period before and after the law, respectively. The grade threshold denotes the quality level beyond which the restaurant is considered as providing high quality. For each period, different grade thresholds are used. Thresholds are either 70, 75 and 80 before the law, and 78.9, 83.9 and 88.9 after law. The variable *INC* denotes average income per capita in the zip code area of restaurant, *AL* is a indicator variable which is equal to 1 if the restaurant has a license to sell alcoholic beverages, and *NIC* is a binary variable which is equal to 1 if the restaurant is not part of chain. The variable *N* is the number of restaurants in the zip code area and lastly *PAL* and *PNIC* are the percentages of restaurants that have alcohol license and are not part of a chain in the zip code area respectively. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 1.3: 2nd Stage Estimates (Specification (1))

Period	Grade Threshold	$\hat{\alpha}_{INC}$	$\hat{\alpha}_{AL}$	$\hat{\alpha}_{NIC}$	$\hat{\delta}$	log-likelihood
Before	70	0.0836*** (0.0019)	-0.0431 (0.0392)	-0.6824*** (0.0444)	-0.0001 (0.0007)	-8788.9
	75	0.0593*** (0.0017)	-0.0634* (0.0363)	-0.8968*** (0.0387)	-0.0014* (0.0008)	-9808.4
	80	0.0477*** (0.0019)	-0.1077*** (0.0374)	-1.3402*** (0.0375)	-0.0055*** (0.0012)	-9337.7
After	78.9	0.0717*** (0.0029)	-0.0696 (0.0660)	0.6145*** 0.0830	0.0108*** (0.0008)	-4192.4
	83.9	0.0426*** (0.0020)	-0.0069 (0.0526)	0.1429*** (0.0475)	0.0075*** (0.0006)	-7679.6
	88.9	0.0285*** (0.0016)	0.0269 (0.0354)	-0.4547*** (0.0414)	0.0050*** (0.0007)	-10151.3

Notes: The period “Before” and “After” denotes the time period before and after the law, respectively. The grade threshold denotes the quality level beyond which the restaurant is considered as providing high quality. For each period, different grade thresholds are used. Thresholds are either 70, 75 and 80 before the law, and 78.9, 83.9 and 88.9 after law. The variable *INC* denotes average income per capita in the zip code area of restaurant, *AL* is a indicator variable which is equal to 1 if the restaurant has a license to sell alcoholic beverages, and *NIC* is a binary variable which is equal to 1 if the restaurant is not part of chain. The variable *N* is the number of restaurants in the zip code area and lastly *PAL* and *PNIC* are the percentages of restaurants that have alcohol license and are not part of a chain in the zip code area respectively. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 1.4: 1st Stage Estimates (Specification (2))

Period	Grade Threshold	$\hat{\gamma}_{INC}$	$\hat{\gamma}_{AL}$	$\hat{\gamma}_{NIC}$	$\hat{\gamma}_N$	$\hat{\gamma}_{N^2}$	$\hat{\gamma}_{PAL}$	$\hat{\gamma}_{PAL^2}$	$\hat{\gamma}_{PNIC}$	log-likelihood
Before	70	0.0264*** (0.0011)	-0.0825*** (0.0239)	-0.6425*** (0.0388)	0.0092*** (0.0013)	-0.0001*** (0.0000)	1.3521*** (0.3555)	0.6616 (0.4602)	-0.0916 (0.0857)	-8636.6
	75	0.0221*** (0.0011)	-0.0913*** (0.0229)	-0.5985*** (0.0330)	0.0069*** (0.0013)	-0.0001*** (0.0000)	2.0098*** (0.3457)	-0.5145 (0.4527)	-0.5136 (0.0792)	-9671.4
	80	0.0286*** (0.0012)	-0.0901*** (0.0234)	-0.5725*** (0.0310)	0.0092*** (0.0013)	-0.0001*** (0.0000)	1.1180*** (0.3725)	-0.2114 (0.4837)	-1.0529*** (0.0841)	-9158.2
After	78.9	0.0191*** (0.0016)	-0.0892*** (0.0325)	-0.0380 (0.0788)	0.0303*** (0.0019)	-0.0002*** (0.0000)	0.0560 (0.5468)	0.2407 (0.6946)	-0.0102 (0.1318)	-4020.0
	83.9	0.0110*** (0.0014)	-0.1091*** (0.0256)	-0.6517*** (0.0481)	0.0146*** (0.0014)	-0.0001*** (0.0000)	-0.3303 (0.4143)	0.2114 (0.5415)	1.0161*** (0.0973)	-7305.4
	88.9	0.0053*** (0.0013)	-0.0187 (0.0225)	-0.7084*** (0.0350)	0.0105*** (0.0013)	-0.0001*** (0.0000)	0.7058* (0.3637)	-0.0135 (0.4778)	0.2900*** (0.0827)	-9903.8

Notes: The period “Before” and “After” denotes the time period before and after the law, respectively. The grade threshold denotes the quality level beyond which the restaurant is considered as providing high quality. For each period, different grade thresholds are used. Thresholds are either 70, 75 and 80 before the law, and 78.9, 83.9 and 88.9 after law. The variable INC denotes average income per capita in the zip code area of restaurant, AL is a indicator variable which is equal to 1 if the restaurant has a license to sell alcoholic beverages, and NIC is a binary variable which is equal to 1 if the restaurant is not part of chain. The variable N is the number of restaurants in the zip code area and lastly PAL and $PNIC$ are the percentages of restaurants that have alcohol license and are not part of a chain in the zip code area respectively. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 1.5: 2nd Stage Estimates (Specification (2))

Period	Grade Threshold	$\hat{\alpha}_{INC}$	$\hat{\alpha}_{AL}$	$\hat{\alpha}_{NIC}$	$\hat{\delta}$	Log-likelihood
Before	70	0.0808*** (0.0019)	-0.0474 (0.0393)	-0.7189*** (0.0453)	0.0014* (0.0007)	-8787.2
	75	0.0560*** (0.0017)	-0.0696* (0.0364)	-0.9374*** (0.0391)	0.0009 (0.0009)	-9809.3
	80	0.0439*** (0.0020)	-0.1152*** (0.0374)	-1.3712*** (0.0376)	-0.0019 (0.0013)	-9345.1
After	78.9	0.0711*** (0.0029)	-0.0685 (0.0660)	0.6301*** (0.0825)	0.0108*** (0.0009)	-4193.5
	83.9	0.0419*** (0.0021)	-0.0037 (0.0810)	0.1388*** (0.0457)	0.0077*** (0.0006)	-7676.7
	88.9	0.0276*** (0.0016)	0.0261 (0.0356)	-0.4715*** (0.0417)	0.0057*** (0.0007)	-10145.1

Notes: The period “Before” and “After” denotes the time period before and after the law, respectively. The grade threshold denotes the quality level beyond which the restaurant is considered as providing high quality. For each period, different grade thresholds are used. Thresholds are either 70, 75 and 80 before the law, and 78.9, 83.9 and 88.9 after law. The variable *INC* denotes average income per capita in the zip code area of restaurant, *AL* is a indicator variable which is equal to 1 if the restaurant has a license to sell alcoholic beverages, and *NIC* is a binary variable which is equal to 1 if the restaurant is not part of chain. The variable *N* is the number of restaurants in the zip code area and lastly *PAL* and *PNIC* are the percentages of restaurants that have alcohol license and are not part of a chain in the zip code area respectively. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

CHAPTER 2

Advertising and ‘Hype’ in the U.S. Movie Industry

2.1 Introduction

The week of theatrical release is considered to be the most important in a film’s life. Typically, a film’s highest box office revenue grossing week comes from its opening, an event which ultimately decides the life-span of the theatrical run. Successful openings give rise to information cascades among consumers, where the choices of initial movie-goers serve as a signal which begets more movie-goers and can be further perpetuated by word-of-mouth effects. The film’s success in the theatrical stage also often carries over to the latter stages of the film’s life, such as in the international theatrical, DVD/Blu-ray, and streaming video-on-demand markets. For these reasons, distributors typically commit the lion’s share of films’ marketing budgets to pre-release advertising campaigns. Empirical papers on the movie industry have long recognized advertising as a significant determinant of box office revenue, but until recently the literature has generally considered it simply as a shifter of demand without considering the possibility of strategic behavior in advertising. Examples of recent articles that do acknowledge strategic behavior in advertising include Liu (2015) and Joo (2009), who study how film distributors strategically use advertising as a catalyst for social learning about movie quality. These papers focus on advertising’s interaction with word-of-mouth transmission of information about movie quality. As a consequence, they pay little attention to its interaction with critical reviews.

This is a limitation because as a measure of underlying quality, critical reviews are an important determinant of a movie’s box office performance. Following release, a film’s

positive reviews attract casual viewers who did not see it in the opening week, and negative reviews deter potential consumers. Critical reviews are so influential on the opening success of films that distributors often strategically manage the availability of critical reviews relative to the release date¹. Advertising not only informs consumers of the film's existence, but it also causes some to read reviews in an attempt to ascertain the film's quality. By this reasoning, distributors with low quality films should have less incentive to advertise because the benefit of increased awareness of the film is offset as consumers learn about its poor quality by reading reviews. Contrary to this expectation, the industry often sees distributors spend heavily on national pre-release advertising campaigns only to have their films receive terrible critical reviews. Recent examples include *Green Lantern* (2011) and *Kingdom of Heaven* (2005), which had pre-release advertising expenditures of \$50 million and \$26 million and received low Rotten Tomatoes scores² of 26 and 39, respectively.

A word commonly used to describe this practice is “hype”, which is defined in this context as the heavy pre-release marketing effort of studios to generate excitement for an upcoming film. This phenomenon is especially puzzling because it also runs counter to the traditional view of advertising as a signal of quality pioneered by Nelson (1974) and Kihlstrom and Riordan (1984), where firms with higher quality products spend more on advertising because they are confident that the expenditure can be recovered from product sales. This paper seeks to explain the behavior of film distributors by examining the interaction between their advertising decisions, the informational structure of the domestic theatrical market, and a film's underlying quality. I provide evidence that for bad films, the success of the hype strategy relies on exploiting consumers' incomplete information with respect to underlying quality in the U.S. film industry.

¹See Brown, Camerer, and Lovallo (2012).

²Rotten Tomatoes is an online critical review aggregator which collects movie reviews published in journals and newspapers and summarizes each one as “fresh” or “rotten” (i.e. good or bad) depending on the review's verdict. The film's “tomatometer” is the percent of reviews that deemed the movie “fresh”. Available from <http://www.rottentomatoes.com/>.

In this paper, I first empirically document and characterize the hype phenomenon. I build a panel data set of movie characteristics, box office revenue, critical reviews and weekly advertising expenditures for wide theatrical releases in the U.S. from 2003 to 2012 and find trends that are consistent with the phenomenon of hype. I use regression analysis to estimate a reduced form specification of the advertising behavior of film studios. I find that after accounting for all the observable characteristics of a film, film studios decrease advertising significantly following the release of the film. This drop in advertising is lessened significantly with each one-point increase in a film’s Rotten Tomatoes score, but even films with perfect scores are predicted to experience drops. In the pre-release stage, the score (as a measure of a film’s pre-determined underlying quality) bears no significant effect on advertising expenditures, a finding suggestive of films pooling in their pre-release advertising strategies.

Exploiting the weekly nature of the box office and advertising expenditure data, I then employ matching difference-in-differences (DID) estimators to estimate the impact of a film’s theatrical release and critical reviews on advertising strategies of “bad” and “good” films. I find that release and being rated as high quality by critics have a significant impact on advertising strategies. Before a film’s release, there is no significant difference in advertising strategies between films with similar characteristics (summarized by a propensity score), but differences emerge after release. Viewing the change in the informational structure of the market (i.e. underlying film quality being revealed through critical reviews) as part of the treatment, the estimated effect is used to motivate the research question: how does the feasibility of hype as an advertising strategy impact the movie industry?

To answer this question, I build a model of advertising in the movie industry based on Butters (1977) in which consumers are informed of the product’s existence by the Poisson arrival of advertisement exposure and randomly check reviews. In the model, only film studios know their film’s underlying quality, so the informational structure is defined by how often consumers exogenously verify the quality of the film by checking critical

reviews. After proving existence and uniqueness of the equilibrium, I numerically solve for equilibrium advertising strategies and find that the advertising strategy of a high quality film is increasing (and that of a low quality film is decreasing) in the “completeness” of information. Also, high quality films always advertise more than low quality films, except for when information is incomplete (i.e. consumers never check reviews) in which case advertising levels are the same. In a world where critical reviews are available ahead of a movie’s release, distributors with high quality films would spend more on pre-release advertising (and less with low quality films) than in the real world. As a consequence, the phenomenon of hype that is observed in the data would not occur in this counterfactual world. Fewer consumers would end up seeing bad films and more consumers see good films in the opening week since being exposed to advertising would be a stronger signal of high quality.

The theatrical film industry is a unique landscape for studying advertising because films are short-lived experience goods whose quality is known either upon consumption or after its release via critical reviews. The films are already produced, so after the scheduling of the film’s release, the only decision left for film distributors to make is advertising level. The costs of advertising are the same regardless of underlying film quality, so the theatrical market is a suitable environment to study the interaction of advertising and information structure. The combination of incomplete information, short-livedness, and the homogenous cost of advertising gives rise to distorted reputational incentives as studios are not necessarily punished for delivering a bad film because there is no repurchase. Moreover, directors are more closely associated with the success or failure of films than the studio or the distributor supporting the film. Unless the director is independent, the distributor handles the marketing of the film. Consumer forgetfulness or lack of attention paid to distributors and their practices allows hype to remain a force in the movie industry, and this paper seeks to explain how the theatrical market is impacted by its presence.

The paper is structured as follows. In Section 2.2, I describe the U.S. film industry

and the elements of film studios' advertising decisions. In Section 2.3, I discuss my data set and present descriptive statistics. In Section 2.4, I discuss evidence of hype in the movie industry, which I observe in trends and estimation results. In Section 2.5, I build a theoretical modeling of advertising and provide conditions for existence and uniqueness of the equilibrium in support of the findings in Section 2.4 and in Section 2.6, I conclude the paper.

2.2 Industry

Today, the U.S. film industry collects nearly 10 billion dollars in box office revenue annually in the domestic theatrical market and additional revenue from subsequent markets, including international theatrical, DVD/Blu-ray, and streaming video-on-demand. This lifecycle begins with the film's planning, which can take several years before production begins. During this stage, the script is written, a director is chosen, roles are filled, filming locations are picked, and funding is secured from the planned distributor. After the cut of the movie is "in the can" and the film is in the post-production phase, the distributor of the movie begins making purchases of advertising slots scheduled for months ahead of time. During this stage, preview trailers are produced to provide consumers with a glimpse of the film and are released in several iterations on the Internet and as attachments to films currently in theaters. These begin with a "teaser" trailer followed by a series of different versions that eventually display the film's official release date when it is finalized. Throughout this post-production stage, release dates change often, as distributors aim to release their films during high demand periods, while also strategically timing the releases in response to competitors' release dates³.

By the post-production phase of a movie's life, movie distributors have a fairly good idea of how the movie will be received by critics. For this reason, distributors are deliberate in when to provide pre-release screenings to critics and whether to even allow

³See Einav (2010).

pre-release critical reviews to be published prior to the official release date. The reasoning is that the lack of reviews is less damaging to a movie's potential than a published negative review⁴. If a distributor has enough confidence in its film's mainstream potential, the film will be slotted for a wide release in which it will show in up to thousands of theaters across the country during its opening weekend. In the industry, an opening that involves 600 or more theaters is considered to be "wide"; otherwise, it is considered "limited". Depending on the strength of the distributor's prior, it may instead plan for the film to have a limited release, either by entering a film festival or by releasing the movie to the public, albeit in only a handful of theaters, to reach a better understanding of how the film would fare during a wide release. The results of the limited release help the distributor determine if the film will be prepared for a wide release and, if so, the number and location of theaters in which the film will open. This practice also has a secondary purpose of generating initial positive word-of-mouth effects that may culminate in a successful wide release.

Following a film's wide release, the movie distributor makes a second advertising purchase decision which governs how the post-release marketing campaign will unfold. This decision is dependent on the movie's performance during its opening weekend, not only in terms of box office performance but also critical and audience reviews, which have become increasingly important with the advent of the internet and the greater access to it afforded by mobile devices. In addition to Rotten Tomatoes, other websites including Fandango and IMDb also provide convenient access to critical reviews, with more of an emphasis on audience-submitted scores. With the rise of these websites, news of a movie's quality can spread rapidly among potential movie-goers. Additionally, social media websites such as Twitter and Facebook provide additional channels through which word-of-mouth effects can quickly travel. Typically, the domestic theatrical run of the film lasts for a few months, and then the film is pulled from theaters. After the film's run in the theater ends, it continues on to the foreign theatrical, DVD/Blu-ray, and streaming

⁴See Charles Moul's *A Concise Handbook of Movie Industry Economics*.

markets, sometimes with overlap between these stages.

Entry of distributors into the movie industry happens very infrequently due to the gargantuan size of long-term incumbents. Historical box office performance data from The Numbers indicate that Warner Bros., Walt Disney, Sony Pictures, 20th Century Fox, Paramount Pictures, and Universal each had box office revenues exceeding \$20 billion between 1995 and 2015 and market shares exceeding 10% from movies that they have distributed. This is consistent with my data set, where the same six distributors were responsible for most wide releases, accounting for more than 70% of observations. Since 2000, more than twenty film studios have been created in the U.S., but only a few have been able to compete with the top incumbents, which include the Weinstein Company, Lionsgate Films, and Relativity Media. However, none of these entrants have broken through to the top 5 of annual market shares, with the exception of Lionsgate in 2014, when it found enormous success with *The Hunger Games* and the last film of the *Twilight* series. While the number of successful entrants has been limited, there have been notable mergers and acquisitions among movie studios. Dreamworks, which had many successful films throughout the 90's, was acquired by Viacom, the parent company of Paramount Studios. Summit Entertainment was purchased by Lionsgate, along with the rights to the *Twilight* franchise.

2.3 Data

The data consist of weekly observations from the beginning of 2003 through 2012 and are compiled from four sources. First, weekend box office revenue data from the U.S. theatrical market were collected from The Numbers⁵. For a given movie on a given weekend, the data set also includes total gross box office revenue collected up to that date (which includes weekday sales), the movie's genre, the movie's distributor (in the U.S.), the number of theaters at which the movie was shown, and the number of days

⁵Available from <http://www.the-numbers.com/>.

since the movie’s first release. Similarly, movie characteristics were obtained from IMDb⁶. For each movie, the genre, MPAA rating (e.g. PG-13, R, etc.), official wide release date, and production budget were collected. Critical reviews (summarized as a “tomatometer” score) and their publication dates were collected from Rotten Tomatoes. I discuss how these data were collected in more detail in the appendix.

Lastly, weekly advertising expenditures were obtained from Kantar Media for all films in the U.S. from the same time period. For a given movie in a given week, the Kantar data set contains the dollar amount of advertising expenditures in national and local markets. The data are further broken down to specify the type of media in which the advertising occurred, such as outdoor advertising (e.g. billboards), radio, magazine, and TV. In this paper, I only consider spending on national advertising, which includes advertisements on network TV, network radio, national newspapers, and other mediums that are viewed by consumers across the U.S. For my sample, I only include movies that were advertised nationally and had budgets written in U.S. dollars, and I dropped films for which any of the components were missing. I display summary statistics for the resulting merged data set in Table 2.1.

A histogram of Rotten Tomato scores is shown in Figure 2.1 accompanied by a kernel density-based approximation of the scores’ distribution. The distribution has a peak at about a score of 20, after which it steadily declines until 80, where it begins to sharply decline. Out of the sample, two-thirds have scores lying between these two scores. 22% of films have scores lying below 20, and the remaining 11% have scores lying above 80. This indicates that lowly rated movies are more common than highly rated ones. Moreover, the majority of movies in the middle of the distribution are below average, which suggests that either critical reviews are very conservative in their assessments or that it is difficult to produce a good movie, even with film budgets allotting hundreds of millions of dollars for the filming process. Throughout this paper, I will use the variable q and the word “score” to refer to a film’s Rotten Tomatoes score.

⁶Available from <http://www.imdb.com/>.

The publication dates and number of reviews by “top critics” listed on Rotten Tomatoes for each film were also collected. To be designated a top critic, a reviewer must satisfy one of several sufficient conditions set by Rotten Tomatoes, such as being a contributor in a print publication in the top 10% of circulation. Summary statistics for the timing of publication of reviews are found in Table 2.3. These data are used as an instrument in my estimation strategy, which I discuss later.

2.4 Evidence of hype

With these data on hand, I search for trends that are consistent with the practice of hype in the U.S. movie industry. A scatterplot of advertising expenditures before and after a film’s release as a function of score is shown in Figure 2.2, along with superimposed fractional-polynomial prediction curves. Except for at the very beginning of the spectrum, the prediction curve for pre-release advertising more closely resembles a horizontal line than its post-release counterpart. This is suggestive of a pooling equilibrium in advertising strategies across film qualities during the pre-release phase, with a separating equilibrium occurring in the post-release phase. Between the two phases, there is a natural decay in advertising for films of all qualities for two reasons. First, the potential market size in the opening week is larger than that in subsequent weeks, because repeat viewers are not common. Second, the movie studio lets word-of-mouth transmission act as a substitute for advertising.

Even though advertising decay occurs for all films, distinct patterns can be observed when films are grouped by quality. Films that end up with scores of less than 30 tend to experience a sharper decay in advertising from the opening week to the second week of release than good films, as seen in Figure 2.3. The highest peak of the distribution is in the first bin, which indicates nearly complete decay in advertising expenditure by the second week for the greatest number of movies in this subset. There are also several local peaks, which suggests that conditional on having poor critical reviews, there is variation in the

continuation of advertising. Some lowly rated films do well in the opening week at the box office and their distributors decide to continue advertising for it. In contrast, Figure 2.4 shows the histogram and kernel density of advertising decays for high quality films, with its unique distinct peak at around 0.25. Additionally, there are many more high quality films that maintain higher post-release advertising campaigns than low quality films. As seen in Figure 2.3, the right tail of the distribution of advertising decay ends around 0.5 for low quality films, but in Figure 2.4, the tail extends well into the 1.0 range for high quality films.

One possible explanation for why the distributors of bad films advertise highly in the pre-release period is that distributors do not know their films' quality. To counter this argument, I examined the set of films that initially had a limited opening and later opened to a wide audience. This is characterized by a film initially having limited run in fewer than 600 theaters, the film being pulled from theaters, and the film lastly having a second opening that is considered wide. The summary statistics for this "limited-then-wide release" sample is found in Table 2.2. In this case, I consider pre-release advertising to include all advertising that occurred prior to the initial limited release through the week of the wide release. Figure 2.5 shows a scatter plot of the pre-release advertising expenditures versus eventual score for the sample. The figure shows that films that are rated badly still choose to advertise highly in the pre-release period. It is noteworthy that it can also be seen in the figure that there appears to be selection bias in the sample. The distribution of scores in this sample is skewed left, with movies in this sample tending to have higher scores. This is intuitive because films that initially have a limited release are more likely to go wide if they see favorable results from the initial limited release, whether they be critical reviews or box office revenue. Films that receive poor feedback from a limited release are less likely to seek a wide theatrical release and will instead consider skipping to the non-theatrical market.

Most movies have reviews published on the days leading up to the film's release, with the majority of reviews being published on the actual release date. However, some movies

have reviews with delayed publications that become available on the day of release at the earliest. From the critical review data, these movies always end up receiving low scores. Conversely, looking at the number of reviews that are published before a movie's release, there is a clear indication that movie studios have a good idea of how their films will be perceived by critics. This is seen in Figures 2.6 and 2.7, where there is a noticeable trend between the number and proportion of pre-release reviews allowed and score. Together with the findings from the limited release sample, this is strong evidence against the argument that ignorance on the part of movie distributors is the explanation for the pre-release advertising patterns seen in the data. Under the assumption that film distributors know the quality of their films, I characterize the effect of movie characteristics and other determinants on advertising behavior for the rest of this section.

2.4.1 Regression Estimation

As notation for the regression estimation model, let $a_{j\tau}$ denote the advertising spending for film j during period τ , which can either be 0 for the pre-release phase or 1 for the post-release phase. Let $a_{j0} = \sum_{t=-\infty}^0 \tilde{a}_{jt}$ denote the entire pre-release advertising campaign, where \tilde{a}_{jt} represents advertising spending in week t of film j 's life, and $t = 0$ represents the opening week. Similarly, let $a_{j1} = \sum_{t=1}^{\infty} \tilde{a}_{jt}$ denote the entire post-release advertising campaign. $q_j \in [0, 100]$ is the quality score of the film from Rotten Tomatoes and w_j contains movie characteristics including the film's budget and dummy variables for the film's MPAA rating (G, PG, PG-13, R) and for the film's genre (Action, Adventure, Drama, Comedy, Thriller, Horror).

I consider the following general model specification for advertising:

$$\log(a_{j\tau}) = \beta_0 + \beta_1 q_j + \beta_2 \mathbb{I}(\text{after release}_{j\tau}) + \beta_3 \mathbb{I}(\text{after release}_{j\tau}) \cdot q_j + w_j + \delta_{j\tau} + \varepsilon_{j\tau} \quad (2.1)$$

On Equation 2.1's right side, $\delta_{j\tau}$ contains time fixed effects related to the film's week of release ($\tau = 0$) and the week after release ($\tau = 1$), w_j is the film fixed effect (which might

include the film's budget), q_j is the film's Rotten Tomatoes score, and $\varepsilon_{j\tau}$ is an unobserved error term. This can be interpreted as something that affects the advertising decision, which is observable only to the film distributor. I discuss the potential endogeneity issue with this term in the identification subsection.

The purpose of this specification is to study how a change in the underlying quality and its interaction with release status affects the film distributor's advertising decision. This is a reduced form of a possible advertising policy function where in expectation, it is dependent on the observable characteristics of the film, the film studio's signal of its underlying quality, and the month and year of its release. The coefficients of interest are β_1, β_2 , and β_3 , which measure the effect of underlying quality and change in information structure on the release of the film on advertising behavior of the distributor.

2.4.1.1 Identification

One possible source of endogeneity is that the unobserved heterogeneity in advertising choices embodied in $\varepsilon_{j\tau}$ is due to misunderstandings of consumer demand or of the potential marketability of the film. In this interpretation, the film distributor is misinformed about consumer demand and the chances of its film's success at the box office, and consequently makes randomly erroneous advertising decisions. This could be correlated with variation in the film's underlying quality that the studio realizes. For example, film studios might be overly optimistic for films that have higher quality levels.

A second explanation is that the advertising shock is intentional, in which case ε_{jt} can be viewed as the film distributor's deliberate deviation from the expected advertising strategy. An example of this would be if the film studio has spent an enormous budget in producing the film and then invests more in advertising, hoping to make up for the cost in the opening week.

To address possible correlation between ε and the covariates, I use the total number of top critical reviews as an instrument. For both of these explanations for endogeneity

issues arising in the model, the ex ante shock (in both the pre-release and post-release periods) is independent of the level of interest in the film among movie critics. This is because the film studio's decision-making or beliefs about the market are independent from the resulting attention received from critics. For the underlying quality explanation, a word-of-mouth process could unfold in the community of top critics that operates differently from that of the general audience, the latter being the desired target for the film studio's advertising campaign. The word-of-mouth effects among critics should be independent from the film studio's ex ante beliefs about consumer demand for the film. The relationship between the number of reviews and the film's score is seen in Figure 2.8.

For the second explanation for endogeneity, film critics could become interested in reviewing a film based on the characteristics, the most conspicuous of which being the film's production budget. This attention could be the result of correlation between the presence of star actors, actresses, and directors and the production budget. It could also be because films that have huge budgets are notable, irrespective of the attached cast and crew. Regardless, there is a strong correlation between the budget and the number of resulting critical reviews, as seen in Figure 2.9. The interest in the movie due to the film's budget should be independent of intentional deviations from expected advertising committed by the film studios.

2.4.1.2 Results

The results of the ordinary least squares (OLS) regression are found in Table 2.4. The results of the instrumental variables (IV) regressions, where the number of critical review publications for a film is used to instrument for the film's score and the logarithm of the film's budget, are found in Tables 2.5 and 2.6 respectively. In general, these indicate the stronger dependence of advertising behavior on the release status of the film. For the average movie in the sample which has a score of 43, if the movie has been released, the logarithm of advertising is predicted to decrease by about 3, an indication that post-release advertising for this type of movie will drop to about 5% of the pre-release level. For

better movies, the drop is less, but there is still a distinct difference between advertising levels in the two release statuses. Note that even a film with a perfect score of 100 is predicted to experience a drop in advertising in the post-release phase.

Even though films of all types see drops in their advertising following release, what is also notable are the differences in advertising level before release. When instrumenting for score, the effect of score on advertising before release (i.e. $\mathbb{I}(\text{After Release}) = 0$) is statistically zero in Specifications 4 and 6 and significant at only the 10% level in Specification 5, as seen in Table 2.5. When instrumenting for budget, the effect of score in the pre-release stage is insignificant or weakly significant in all specifications as can be seen in Table 2.6. When all observable characteristics of the film are accounted for, there is a significant effect of score on post-release advertising, but there is at best a weakly significant effect in the pre-release period. This is an indication that, holding film characteristics constant, a film having a lower score will not deter the film’s distributor from advertising.

In the next subsection, I instead consider the identification issue that movie characteristics might have different impacts on advertising expenditures depending on whether it is the pre-release or post-release period. To address this, I use a DID approach, where treatment for a film is considered to be its release and subsequent revelation as having high quality.

2.4.2 DID Estimation

I consider the following model specification for studio advertising behavior for film j in period $\tau \in \{0, 1\}$:

$$\log(a_{j\tau}) = \beta_0 + \beta_1\theta_j + \beta_2\mathbb{I}(\text{after release}_{j\tau}) + \phi\mathbb{I}(\text{after release}_{j\tau}) \cdot \theta_j + w_j + \delta_{j\tau} + \varepsilon_{j\tau} \quad (2.2)$$

where $\tau = 0$ represents the pre-release period, $\tau = 1$ represents the post-release period, and $a_{j\tau}$ is the total advertising expenditure in period τ . θ_j is a indicator variable of

whether film j is considered to be “good” or “bad”, where $\theta_j = 1$ if $q \geq 50$ and $\theta_j = 0$ otherwise. w_j includes film’s budget, a dummy for the film’s MPAA rating, a dummy for the film’s genre, the number of theaters at which the film is opening. $\delta_{j\tau}$ represents time fixed effects related to the film’s week of release. The term $\varepsilon_{j\tau}$ is unobservable to the econometrician, but is known to the film distributor. The coefficient of interest is ϕ , which represents the effect that a movie being revealed to be high quality has on the advertising strategy of the film studio. In other words, this measures the effect that the publication of reviews has on the advertising strategy of the film.

Due to the likelihood that movie characteristics influence quality and the high dimensionality of the data set, I employ a kernel-based matching estimator which matches films on propensity score. In this case, the propensity score $P(w) = Pr(\theta = 1|w)$ measures the probability of a movie with characteristics w being rated as having high quality by critics. Two assumptions are necessary to allow for the use of this technique: 1) $E(a_{j0}(0) - a_{j1}(0)|P(w), \theta = 1) = E(a_{j0}(0) - a_{j1}(0)|P(w), \theta = 0)$ and 2) $0 < Pr(\theta = 1|w) < 1$ where $a_{j\tau}(0)$ represents how the film with quality θ would advertise in period τ if it were not in the treatment group⁷. The second condition is easy to satisfy, since there is no set of movie characteristics that will guarantee good critical reviews. If there was, then all film studios would employ these characteristics as a strictly dominant strategy.

The first condition is more difficult to satisfy. Advertising expenditures are required to have the same decay between between bad films and good films had they been rated bad. There might be some unobserved appeal that drives audiences to see the film regardless of whether the movie is rated highly or not. A high box office return despite earning low reviews would cause a film distributor to continue to advertise in the subsequent weeks. Examples tend to be large summer blockbusters such as *Transformers* or *Twilight* that, despite having low reviews, succeed at the box office during their opening week. In the data, such movies are uncommon and are usually associated with enormous budgets,

⁷See Heckman, Ichimura, and Todd (1997).

famous directors and actors, and franchises. I assume that beyond these exceptional titles, advertising is used to draw consumers' attention to critical reviews as long as they are available. A distributor with a film that actually has high underlying quality, if excluded from the treatment group, would have the same incentive to direct consumers to critical reviews through advertising as the distributors of bad films, so the first condition should also be satisfied.

Let S denote the entire sample of films and let I denote the set of movies that have $\theta = 1$. Following the notation in Equation 2.2, the treatment effect ϕ can be estimated with:

$$\hat{\phi} = \frac{1}{n_G} \sum_{j \in I} \{a_{j1} - \hat{E}[a_{j1} | P(X_j), \theta_j = B]\} - \frac{1}{n_G} \sum_{j \in I} \{a_{j0} - \hat{E}[a_{j0} | P(X_j), \theta_j = B]\} \quad (2.3)$$

where n_G is the number of good films and X_j includes characteristics w_j and δ_τ . To construct the counterfactual advertising strategy of a distributor with a film of type $\theta = 1$, the propensity score $P(X_j)$ is estimated with a logit model and used to calculate $\hat{E}[a_{j\tau} | P(X_j), \theta_j = B] = \sum_{k \in S \setminus I} W_k(P(X_j)) a_{k\tau}$, where weights are given by $W_j(P(X_i)) = \frac{K(\frac{P(X_i) - P(X_j)}{h_n})}{\sum_{k \in S \setminus I} K(\frac{P(X_i) - P(X_k)}{h_n})}$, the Epanechnikov kernel is used for $K(\cdot)$, and the bandwidth size h_n is chosen by default to be 0.06.

2.4.2.1 Results

The results of the DID analysis under various specifications are found in Table 2.11. In the first two specifications, I include the number of opening theaters, the budget, the MPAA rating, the genre, and month and year dummies, and vary what value to use for advertising expenditures. In Specification 1, a_{jt} represents only opening week or second week advertising expenditures while in Specification 2, $a_{j\tau}$ represents either pre-release or post-release advertising. Table 2.9 contains the logit propensity score estimation results.

While the budget coefficient is significant, the regression's relatively low pseudo R^2 of 0.069 suggest that a high budget may not guarantee a high resulting quality. The horror genre dummy is dropped due to collinearity, and the results indicate that relative to horror, the other genres have more success with critics, which is unsurprising since horror movies appeal to a very niche part of the market. The month and year of release have an insignificant effect on this propensity score. Table 2.8 provides the results of a two-sided t-test on the difference of the means of the treatment and control groups of films in the pre-release period, which affirms these findings by showing a weakly significant difference between average outcomes of the two groups, with insignificant differences in the estimated impact of the rest of the covariates.

The results in Table 2.11 show that there is no significant difference in advertising expenditures between the distributors of high and low quality films that have similar propensity scores. This implies that before release, films that have the same characteristics are advertised in similar ways. In the post-release period, I find a significant difference in advertising levels. The final DID estimate accounts for the decay in advertising for both film types in the post-release period, and predicts a significant treatment effect of positive critical reviews. In Specification 1, where advertising is specified to be either opening week or second week, if a movie is revealed to be high quality, its advertising expenditure would increase by 117% relative to low quality films. In Specification 2, the change in entire pre-release advertising (which includes the opening week) and post-release advertising would be 155%. Specifications 3 and 4 also include opening box office revenues as a matching covariate and vary the accounting of a_{jt} in the same way as in the first two specifications. In Specification 3 the predicted increase in opening week advertising would be 61.6% and in Specification 4, pre-release advertising would increase by 78%.

The significance of the change in advertising due to the publication of positive critical reviews appears to be robust to the changes in model specification and when it is combined with the facts gathered in the rest of this section, it suggests that film distributors behave

in a manner that is consistent with the practice of hype. In particular, the distributors of bad films advertise heavily in the pre-release period only to abandon their product once review publications reveal their true quality. In the next section, I build a theoretical model of advertising which focuses on changes in the availability of information on film quality as a key feature in the movie industry which allows the feasibility of a hype strategy by film distributors.

2.5 Theoretical Model

In this section, I build a model of the movie industry which focuses on the advertising decisions of film studios. I do not include release date scheduling or casting as decisions in the model, which themselves are important choices for the film distributor. Instead, I take the film's release date and production outcome as given and focus entirely on the resulting advertising decisions of the film distributor. The purpose of this model is to demonstrate that under reasonable assumptions, the distributors of both bad and good films will advertise in similar ways in the pre-release period but will diverge in the post-release period, which supports the findings of the DID analysis. The result is generated from considering that the probability with which consumers verify the quality of films varies across a film's lifetime, and is generally less in the pre-release period than it is in the post-release period. At the end of this section, I show existence and uniqueness of equilibrium conditions, demonstrate examples of equilibria through the computation of numerical solutions, and characterize extensions to the model.

2.5.1 Timing

I consider a single film distributor which has produced a film and faces a mass of consumers of size M who are initially uninformed of the existence of the film. The film has characteristics given by w that are determined exogenously. At the beginning of period t , the quality $\theta \in \{B, G\}$ is drawn from distribution $F(\theta|w)$, and corresponds to

the film having either low or high quality. The value of θ is known only to the movie studio who makes an advertising purchase $a \geq 0$, letting consumers know of the existence of the film.

A consumer reached by the advertising observes all the characteristics of the film except for θ and decides whether or not they are interested in the film. If so, she learns θ with probability r . At the end of period t , she makes a decision to see the movie or not. As a result of the verification process, the consumer may or may not know the true quality of the film. Consumers who are not reached by advertising do not face this decision and automatically choose to not watch the movie. Allowing for dynamics would enrichen the model, but for now, I will discuss the model only in the context of a single period. I consider a two-period extension in Section 2.5.10.

2.5.2 Film Quality

The underlying exogenous quality of the film θ is influenced by factors such as the identities of the director and actors of the film, the genre of the film, the MPAA rating of the film, and the budget. Let observable characteristics of the film be summarized by w . During film production, the eventual quality of the film is influenced by unobservable events that are summarized by ε , which is private information of the film studio and is unobservable to the econometrician. An example of this would be a setback in film production due to a mistake by an editor. The distribution of these stochastic events $g(\varepsilon)$ is known to film studios and consumers alike. In the case of two film types, the quality is determined by:

$$\theta = \begin{cases} G & \text{w.p. } \mu(w) \\ B & \text{w.p. } 1 - \mu(w) \end{cases} \quad (2.4)$$

where $\mu(w)$ denotes the distribution of underlying movie quality as a function of the movie's characteristics. Movies with $\theta = G$ give consumers greater utility upon viewing than those with $\theta = B$. Letting $U(\theta)$ be the consumer's utility of watching a movie with

quality θ and characteristics w , I assume that $U(G, w) > U(B, w)$ for all w .

2.5.3 The Firm's Decision

After the underlying quality of a film is realized, the film distributor makes an advertising decision. For notational brevity, I will drop the w argument in all equations in the rest of this section. Additionally, when the term “film type” is used, it is intended to refer to a film of $\theta = G$ versus $\theta = B$, for fixed w . The distributor chooses to spend a on advertising. For a film with quality θ and advertising expenditure a , let $R(a|\theta)$ be the film distributor's revenue.

2.5.3.1 Advertisements

As in Butters (1977), I model advertising as the costly sending of messages that inform consumers randomly of the existence of a product. With flexible prices, the cost of advertising on a types of ads that would get viewership from a larger number of consumers would be more expensive. I assume that these prices adjust so that on average, the cost of exposing one consumer to an advertisement is equalized regardless of the medium of advertising. Let b be the cost of sending messages to a random number of consumers such that on average one consumer is reached. The film distributor can choose an advertising expenditure $a \in [0, \bar{a}]$ where $a = 0$ represents no advertisements at all. The finite upper bound \bar{a} can in principle be very high, but not infinite, since there is limited advertising space in the world.

At an advertising level a , the number of messages that are sent, each with an average number of consumers reached of 1, is $\frac{a}{b}$. Whether or not a consumer receives a message is independent across messages, so it is possible for the same consumer to receive redundant advertisements. When a consumer receives an advertisement, she becomes aware of the existence of the film and becomes a *reached* consumer. What she does next depends on whether she is interested in the film and becomes a *potential* movie-goer and further,

whether or not she reads the reviews of the movie and becomes an *informed* potential movie-goer.

The number of complete messages that an individual consumer receives is distributed binomial with parameters $n = \frac{a}{b}$ and $p = \frac{1}{M}$. As the number of consumers and the number of messages sent is taken to the limit with $\frac{a}{b} \cdot \frac{1}{M} = \lambda$ as constant, the number of messages that a consumer receives is distributed Poisson with parameter λ . It follows that for a single consumer, the probability that she does not receive any message is $\exp\{-\lambda\} = \exp\{-\frac{a}{b} \cdot \frac{1}{M}\}$. Such a consumer cannot become a potential consumer because she does not know about the existence of the movie. The proportion of reached consumers who receive at least one message and become aware of the movie is $1 - \exp\{-\frac{a}{b} \cdot \frac{1}{M}\}$. Because the number of movie-goers and the amount of advertising expenditures in the movie industry are each in the many millions, values in the model can be considered to approximate the asymptotic values.

2.5.4 The Consumer's Decision

The utility maximization problem is limited to consumers who are reached by advertising. Those who do not receive advertising messages are left unaware of the movie's existence, and have no viewing decision to make. For reached consumers, the problem consists of two stages. In each, reached consumers and potential consumers are weighing the expected utility of seeing a film to the outside option. I represent the outside option of the consumer as the consumer not seeing the movie (and staying home) and receiving U_0 .

2.5.4.1 Consumer Utility Maximization Problem

If the risk-neutral consumer i is reached by the advertisements of the film distributor and becomes aware of film's existence, she will form a conditional expectation about her mean utility from watching the film. Whether or not the reached consumer becomes a

potential consumer depends on this conditional expectation and also on an idiosyncratic preference shock ξ_i , which is private information to consumers. I assume ξ_i is distributed according to the probability density function $h(\xi)$, which is known to both consumers and the film studios. Consumer i 's total expected utility from seeing the film is given by $u(\xi) = E_c [U(\theta)|\text{reached}] + \xi_i$ where the conditional expectation depends on the fact that the consumer was reached by the advertising strategy of the film studio. The consumer will become potential if her expected utility from seeing the film net of the fixed ticket price T is greater than from the value of the outside option, or if:

$$E_c [U(\theta)|\text{reached}] + \xi_i - T > U_0 \quad (2.5)$$

For notational brevity, I let $V_1 = E_c [U(\theta)|\text{reached}] - T$ be the mean initial impression that a consumer would get from being reached by advertising and seeing the characteristics w . A reached consumer therefore solves the following problem:

$$\max\{V_1 + \xi_i, U_0\} \quad (2.6)$$

2.5.4.2 Conditional Expected Quality

A consumer who is reached by advertising learns the observables of the movie, including its production budget, MPAA rating and genre. The reached consumer also knows that they've been reached by advertising, but does not know the actual advertising decision a of the movie studio, only the equilibrium advertising strategies a_θ^* . This assumption is reasonable because consumers would not see the current advertising strategies of film studios, but may learn about them after the film leaves theaters. Consumers have an idea of how film studios have behaved in the past (and thus how they are expected to behave in the present) but cannot see any deviations in advertising as they occur. In a fully dynamic model, consumers would update their beliefs ex post, but I consider only a static model where consumers do not update their beliefs about film studio behavior

because entrants in film distribution are rare, as discussed in Section 2.2.

Conditional on these characteristics and the advertising choice, the consumer forms beliefs about the quality of the film that can be summarized by $Pr_c(\theta = G|\text{reached})$. This element can be thought of as the consumer's initial impression of the film, which is summarized by:

$$Pr_c(\theta = G|\text{reached}) = \frac{\mu Pr(\text{reached}|\theta = G)}{\mu Pr(\text{reached}|\theta = G) + (1 - \mu) Pr(\text{reached}|\theta = B)} \quad (2.7)$$

The consumer thus forms expectations of the quality of the new film, which is conditional on its observable characteristics:

$$E_c[U(\theta)|\text{reached}] = Pr_c(\theta = G|\text{reached})U(G) + Pr_c(\theta = B|\text{reached})U(B) \quad (2.8)$$

The indifferent consumer is the one whose idiosyncratic shock $\bar{\xi}$ causes Equation 2.5 to hold with equality. The value of this shock is given by:

$$\begin{aligned} \bar{\xi} = T - & \left(\frac{\mu(1 - \exp(-\frac{a_G}{bM}))}{\mu(1 - \exp(-\frac{a_G}{bM})) + (1 - \mu)(1 - \exp(-\frac{a_B}{bM}))} U(G) \right. \\ & \left. + \frac{(1 - \mu)(1 - \exp(-\frac{a_B}{bM}))}{\mu(1 - \exp(-\frac{a_G}{bM})) + (1 - \mu)(1 - \exp(-\frac{a_B}{bM}))} U(B) \right) \quad (2.9) \end{aligned}$$

where U_0 is normalized to 0. This term is equal to $-V_1$, and it separates the consumers who would become interested in the film from those who would dismiss it, and thus is the crux of the film distributor's advertising decision. The share of potential movie-goers is given by the share of consumers for whom $\xi > \bar{\xi}$. For each potential movie-goer, they will check the reviews of the movie with probability $r \in [0, 1]$. The intuition of why r should be strictly less than 1 in a film's pre-release phase is that word of mouth effects are minimal and consumers may not have adequate access to the early reviews of critics. In the post-release period, it is reasonable to assume that r will increase, since

critical reviews and word of mouth effects are more widely available. I assume that this probability is exogenous. One way in which this assumption can be rationalized is by considering that since there is no cost to consumers for checking the reviews for a film, all consumers choose to search, but they may not be successful depending on the availability of reviews and the consumer's access to web technology. Consumers who do check reviews become informed because they know θ and then solve the problem:

$$\max\{U(\theta) - T + \xi_i, 0\} \quad (2.10)$$

For notational brevity, I make the abbreviation $V_2(\theta) = U(\theta) - T$ so that interested consumers who verify the film's quality end up watching the film if $V_2(\theta) + \xi_i > 0$. Potential consumers who remain uninformed go to see the movie because of their initial interest in the first stage of the problem, and they pay the film distributor T .

2.5.5 Film Studio Revenue

The ticket price T and verification probability r are taken to be parameters of the environment so they are omitted as arguments for notational brevity. Later, I will allow r to vary in order to represent changes in the information structure of various stages in a film's life. The expected box office revenue earned from a single reached consumer is given by:

$$E[R_i|\theta] = \int_{-\infty}^{+\infty} \mathbb{I}(V_1 + \xi > 0)((1 - r)T + r[\mathbb{I}(V_2(\theta) + \xi > 0) \cdot T])h(\xi)d\xi \quad (2.11)$$

As discussed in section 2.5.3.1, each consumer has a chance of $1 - \exp\{-\frac{a}{bM}\}$ to receive

a message, the expected total revenue of the film studio choosing advertising level a is:

$$E[R(a)|\theta] = M(1 - \exp\{-\frac{a}{bM}\}) \cdot \int_{-\infty}^{+\infty} \mathbb{I}(V_1 + \xi > 0)((1-r)T + r[\mathbb{I}(V_2(\theta) + \xi > 0) \cdot T])h(\xi)d\xi \quad (2.12)$$

Consider the indifferent consumer with preference shock $\bar{\xi}$. Consumers with preference shocks below this value will not be interested in the film, so the film studio has no chance of capturing these consumers, whether they're informed of the film's existence or not. Notationally, I will now use $\bar{\xi}$ instead of $-V_1$. The expected revenue is then:

$$E[R(a)|\theta] = M(1 - \exp\{-\frac{a}{bM}\}) \int_{\bar{\xi}}^{+\infty} ((1-r)T + r[\mathbb{I}(V_2(\theta) + \xi > 0) \cdot T])h(\xi)d\xi \quad (2.13)$$

From 2.9, the total revenues for bad and good films are given by:

$$E[R(a)|B] = M(1 - \exp\{-\frac{a}{bM}\})T((1-r)(1 - H(\bar{\xi})) + r(1 - H(-V_2(B)))) \quad (2.14)$$

$$E[R(a)|G] = M(1 - \exp\{-\frac{a}{bM}\})T(1 - H(\bar{\xi})) \quad (2.15)$$

The incentives for films to advertise can be seen from the revenue function's derivatives. Beginning with the revenue for a good film (Equation 2.15), taking the first derivative yields:

$$\frac{\partial E[R(a_G)|G]}{\partial a_G} = T \left[\frac{1}{b} \exp(-\frac{a_G}{bM})(1 - H(\bar{\xi})) - M(1 - \exp(-\frac{a_G}{bM}))h(\bar{\xi}) \frac{\partial \bar{\xi}}{\partial a_G} \right] \quad (2.16)$$

Assume $a_B = 0$. Then, $\bar{\xi} = T - U(G)$, so $\frac{\partial \bar{\xi}}{\partial a_G} = 0$. Then, $\frac{\partial E[R(a)|G]}{\partial a_G} = \frac{1}{b} \exp(-\frac{a_G}{bM})(T(1 - H(\bar{\xi}))) > 0$ for all a_G . If $a_B > 0$, then $\frac{\partial \bar{\xi}}{\partial a_G} > 0$ so that $\frac{\partial E[R(a_G)|G]}{\partial a_G} > 0$. The good film

distributor will find advertising profitable if Equation 2.16 is greater than 1, which is possible with low enough b and high enough T . Otherwise, the distributor will not advertise and in the model, no consumer will ever find out about the movie's existence.

The derivative of the revenue for a bad film is given by:

$$\frac{\partial E[R(a_B)|B]}{\partial a_B} = T \left[\frac{1}{b} \exp\left(-\frac{a_B}{bM}\right) \left((1-r)(1-H(\bar{\xi})) + r(1-H(-V_2(B))) \right) - (1-r)M(1-\exp\left(-\frac{a_B}{bM}\right))h(\bar{\xi}) \frac{\partial \bar{\xi}}{\partial a_B} \right] \quad (2.17)$$

The first term in this derivative is the marginal gain in revenue from reaching more consumers. The second term reflects the change in consumer beliefs with an increase in a_B : consumers would more likely believe advertising to be associated with bad films, so the indifferent consumer would require an even higher shock $\bar{\xi}$ to remain indifferent. The two terms work in opposite directions with an increase in a_B . In order for $\frac{\partial E[R(a_B)|B]}{\partial a_B}$ to be strictly positive, the following condition is required:

$$\frac{\partial \bar{\xi}}{\partial a_B} < \frac{1}{bM} \frac{\exp\left(-\frac{a_B}{bM}\right) \left((1-r)(1-H(\bar{\xi})) + r(1-H(-V_2(B))) \right)}{(1-r)(1-\exp\left(-\frac{a_B}{bM}\right))h(\bar{\xi})} \quad (2.18)$$

Similar to good films, bad films can be incentivized to advertise with appropriate values of b and T . Accounting for the fact that some values of b and T can cause at least some distributors to be disincentivized from advertising, I define the advertising policy functions in the next subsection.

2.5.6 Profit Maximizing Advertising Levels

Since the cost of the advertising decision a is a itself, film studio profits can be written as a function of advertising as follows:

$$E[\Pi(a)|\theta] = E[R(a)|\theta] - a$$

Optimal film studio advertising levels are determined from first order conditions on the profit functions of film studios and are given by:

$$a_B^* = \max \left(-bM \log \left(\frac{b}{T[(1-r)(1-H(\bar{\xi})) + r(1-H(-V_2(B)))]} \right), 0 \right) \quad (2.19)$$

$$a_G^* = \max \left(-bM \log \left(\frac{b}{T(1-H(\bar{\xi}))} \right), 0 \right) \quad (2.20)$$

where $H(\cdot)$ denotes the cumulative distribution function associated with $h(\cdot)$, $\bar{\xi}$ is determined by Equation 2.9, and the max operator is to disallow negative advertising levels. Note that these conditions are unlike typical best response functions, because the position of the indifferent consumer $\bar{\xi}$ is unmovable by the current period's advertising decisions by film studios. As defined in the next subsection, any equilibrium in this model will be self-fulfilling with respect to consumer expectations.

2.5.7 Equilibrium

With the setup of the model, I define an equilibrium in this environment, with discussions of existence and uniqueness of this equilibrium in the following subsections.

Definition 1 *An advertising equilibrium consists of advertising strategies a_B^* , a_G^* of the distributors of bad films and good films such that given consumer verification probabilities r , the cost of advertising b , the market population size M , and the ticket price T :*

1. *Film studios maximize their two-period profits with their advertising choices taking consumers' conditional expectations $\bar{\xi}$ as given.*
2. *Consumers' movie-going decisions maximize their expected utility.*
3. *a_B^* , a_G^* match consumer expectations $\bar{\xi}$.*

2.5.7.1 Existence

An interior equilibrium exists with certain assumptions on b and T . Note that both advertising policy functions a_B^*, a_G^* are continuous functions of $\bar{\xi}$. The function $\bar{\xi}$ itself is also continuous in a_B^*, a_G^* except for when both advertising levels are equal to 0. This would be the case only if b and T were such that:

$$b > T(1 - H(\bar{\xi})) \geq T[(1 - r)(1 - H(\bar{\xi})) + r(1 - H(-V_2(B)))] \quad (2.21)$$

for any value of a_G, a_B . In this case, the equilibrium would be $a_G^* = 0, a_B^* = 0$. To ensure an equilibrium in which at least one film type advertises, I make the following assumption:

Assumption 1 $b \leq T(1 - H(-V_2(G)))$

Assumption 1 is relaxed with a decrease in b . The effect of changing T is ambiguous because $1 - H(-V_2(G))$ is decreasing in T . While higher ticket prices will generate more revenue per ticket sold, the number of tickets sold will decrease due to a higher required ξ for a consumer to see the film.

If $r = 0$, then the advertising strategies of good and bad film are the same. If $r > 0$, then there are values of $b, T, U(B), U(G)$ such that even if Assumption 1 is satisfied, $b > T[(1 - r)(1 - H(\bar{\xi})) + r(1 - H(-V_2(B)))]$. In particular, lower values of $U(B)$ (or higher values of r) will make the sum $T[(1 - r)(1 - H(\bar{\xi})) + r(1 - H(-V_2(B)))]$ decrease, while $T(1 - H(-V_2(G)))$ remains the same, so that it is possible for $a_G^* > 0, a_B^* = 0$ in equilibrium.

I have already established that even if $a_B = 0$, good films are still incentivized to advertise with appropriate b and T . Thus, when the inequality given by Equation 2.18 does not hold so that $a_B^* = 0$, good films will still choose $a_G^* > 0$ in equilibrium. Even with bad films not advertising, equilibrium advertising by good films is bounded due to

the concavity of the revenue function, established by its second derivative:

$$\frac{\partial^2 E [R_0(a_0)|G]}{\partial a_G^2} = -\frac{1}{b^2 M} \exp\left(-\frac{a_G}{bM}\right) (T(1 - H(\bar{\xi}))) = -\frac{1}{b^2 M} \exp\left(-\frac{a_G}{bM}\right) (T(1 - H(-V_2(G)))) < 0.$$

Therefore, with $a_B = 0$, the set of possible a_G is bounded above.

The set of possible equilibrium a_G^* is bounded above even if $a_B > 0$. The maximum value of $1 - H(\bar{\xi})$ is 1, which yields as the good film's maximum best response $\bar{a}_G = -bM \log(\frac{b}{T})$. Thus, a_G^* is bounded to the interval $[0, -bM \log(\frac{b}{T})]$, assuming that $b < T$.

Similarly, a bad film's advertising strategy is bounded to the same interval because the maximum value that $(1 - r)(1 - H(\bar{\xi})) + r(1 - H(-V_2(B)))$ could have is 1, if $a_G = 0$ and $\bar{\xi} = -V_2(B)$. The incentives to advertise for bad films will always be less than or equal to the same incentives for good films, since the size of the mass of consumers that good films sell to is limited exactly to $1 - H(\bar{\xi})$, whereas for bad films, they lose some of $1 - H(\bar{\xi})$ due to the verification probability r . Therefore in any equilibrium, $a_B^* \leq a_G^*$ and the set of possible a_B is bounded to a smaller interval than the set of possible a_G . Given these findings, I use Brouwer's fixed-point theorem where I interpret the equilibrium policy functions for good and bad films as mapping from the set $[0, -bM \log(\frac{b}{T})] \times [0, -bM \log(\frac{b}{T})]$ to itself to make the following statement:

Proposition 1 *Suppose Assumption 1 is satisfied. Then, there exists an equilibrium in advertising strategies a_B^*, a_G^* where $a_B^* \leq a_G^*$ with $a_G^* > 0$ and $a_B^* \geq 0$.*

2.5.7.2 Uniqueness

Uniqueness of the equilibrium is established by demonstrating the monotonicity of each best response function under certain assumptions. For fixed r , the first derivatives of each film type's advertising best response functions with respect to the other film type's advertising strategy are given by:

$$\frac{\partial a_B(a_G)}{\partial a_G} = - \underbrace{M}_{>0} \underbrace{((1-r)T(1-H(\bar{\xi})) + rT(1-H(-V_2(B))))}_{>0} \underbrace{b(1-r)Th(\bar{\xi})}_{>0} \underbrace{\frac{\partial \bar{\xi}}{\partial a_G}}_{<0} \quad (2.22)$$

$$\frac{\partial a_G(a_B)}{\partial a_B} = - \underbrace{MT(1-H(\bar{\xi}))(bTh(\bar{\xi}))}_{>0} \underbrace{\frac{\partial \bar{\xi}}{\partial a_B}}_{>0} \quad (2.23)$$

and thus, $\frac{\partial a_B(a_G)}{\partial a_G} > 0$ and $\frac{\partial a_G(a_B)}{\partial a_B} < 0$ if $a_B > 0$ and $a_G > 0$. This result is intuitive: the presence of advertising by the distributors of bad films decreases the return of advertising for the distributors of good films. Consumers who become reached by advertising are more likely to believe that the advertising has come from a bad film. On the other hand, the presence of the advertising of good films increases the gains to advertising for the distributors of bad films, since a reached consumer would be more likely to believe that the advertisement is from the distributor of a good film.

With this result, I make the following statement:

Proposition 2 *Suppose Assumption 1 is satisfied. Then for each r , there exists a unique equilibrium (a_B^*, a_G^*) where $a_B^* \leq a_G^*$*

2.5.7.3 Equilibrium's Dependence on r

Previously I have assumed r to be a constant parameter of the environment. Now I allow it to vary in order to generate comparative statics regarding the equilibrium. In this section, I show that the equilibrium advertising strategy of the distributors of good films is increasing and that of the distributors of bad films is decreasing as a function of r .

The derivative of the equilibrium advertising strategy of the bad film is given by:

$$\frac{\partial a_B^*}{\partial r} = \frac{bM(H(\bar{\xi}) - H(-V_2(B)) - (1-r)h(\bar{\xi})\frac{\partial \bar{\xi}}{\partial r})}{(1-r)(1-H(\bar{\xi})) + r(1-H(-V_2(B)))} \quad (2.24)$$

While $H(\bar{\xi}) < H(-V_2(B))$ by construction, the ultimate sign of Equation 2.24 depends on the size of the term $(1-r)h(\bar{\xi})\frac{\partial \bar{\xi}}{\partial r}$. If the difference in the utilities from watching good and bad films is sufficiently high, then the bad film's equilibrium strategy will be decreasing in r . Intuitively, an increase in r generates more informed consumers from the pool of interested, but uninformed, consumers, so that the expected revenue of the bad film will decrease. This is because the required shock for an informed consumer to see the bad film is higher than for interested, but uninformed, consumers.

The counterpart to Equation 2.24 for good films is given by:

$$\frac{\partial a_G^*}{\partial r} = -\frac{bMh(\bar{\xi})\frac{\partial \bar{\xi}}{\partial r}}{1-H(\bar{\xi})} \quad (2.25)$$

which also depends on the sign of $\frac{\partial \bar{\xi}}{\partial r}$. If an increase in r causes $a^*(B)$ to decrease, $\frac{\partial \bar{\xi}}{\partial r} < 0$ and a_G^* will increase.

2.5.8 Numerical Solution Example

With the existence and uniqueness of the equilibrium established, I find the numerical equilibrium solutions where advertising strategies are consistent with the consumer expectations that are dependent on them. Assuming fixed consumer utility of viewing movies with $\theta = G$ and $\theta = B$, the resulting equilibrium advertising strategies are a function of the quality verification probability r . To demonstrate features of this equilibrium, I assume $T = \$5.3$, that $M = 230.7$ million, $b = 0.05$, and that $\mu = 0.5$.⁸ I assume that $U(\theta = G) = 4.0$ and $U(\theta = B) = 3.74$, so that both average utilities are below the ticket

⁸ b is approximated from the price of a Super Bowl advertisement. In 2014, 112.2 million people watched the Superbowl and the cost of an 30-second TV advertisement was \$4.5 million in 2015.

price T . Lastly, I assume a Type 1 extreme value distribution for ξ .

The resulting equilibrium advertising strategies are found in Figure 2.10. In this specification, the advertising level of bad films falls completely to 0 as r approaches 0.94. This is because T is so far removed from U_B that it would be very unlikely for the distributor of a bad film to find a consumer whose preference shock is high enough to make them want to see it. The fact that T is higher than U_G also causes the distributor of a good film to not invest too heavily in advertising, even when distributors of bad films give up on advertising completely. Since T is relatively high, it still requires a relatively high ξ_i for a consumer to become interested in the good film upon being reached by advertising.

As a second example, I consider a second specification where all the parameters are the same except that I assume that $U(\theta = G) = 5.48$ and $U(\theta = B) = 5.144$, so that T falls in between the two average utility levels. The result is found in Figure 2.11. In this case, the difference between T and U_B is small enough that the distributors of bad films still advertise in equilibrium even when $r = 1$.

2.5.9 Continuous Types

A natural extension to this model is to generalize the distribution of movie types to allow for continuous variation in underlying quality. For a movie with characteristics w , its underlying quality $q \in [0, 1]$ is distributed according to the continuous distribution function $F(q|w)$ with probability density function $f(q|w)$. I assume that consumer utility $u(q, w)$ is such that $u(q', w) > u(q, w)$ for all $q' > q$ and w . As before, I will drop all instances of w in the rest of the equations in this subsection.

Upon being reached by advertising, with assumed advertising strategies $a(q)$, the indifferent consumer has the idiosyncratic shock $\bar{\xi}$ where:

$$\bar{\xi} = T - \int_0^1 Pr(q|\text{reached})u(q) dq = T - \int_0^1 \frac{(1 - \exp(-\frac{a(q)})}{bM})}{\int_0^1 (1 - \exp(-\frac{a(q)})}{bM}) f(q) dq} u(q) f(q) dq \quad (2.26)$$

Then, the equilibrium advertising strategy for a film with quality q is given by:

$$a^*(q) = -bM \log\left(\frac{b}{T((1-r)(1-H(\bar{\xi})) + r(1-H(\max(\bar{\xi}, T-u(q))))))}\right) \quad (2.27)$$

In the equilibrium of this generalized version of the model, all film types that have $q > \bar{q}$ choose the same advertising strategy \bar{a}^* , where \bar{q} is determined by $\bar{\xi} = T - u(\bar{q})$. Other film types with $q < \bar{q}$ choose smaller $a^*(q)$, which decreases as q decreases. This weak monotonicity in advertising strategies across all q is due to the same reasoning as in the two-type case, where the incentives to advertise for a lesser q film are less, so they will never spend more on advertising than a higher q film in equilibrium.

What is perhaps more apparent in this generalized version of the model is that films of very high quality are still limited in their potential pool of consumers by the presence of lesser quality films, even in the full information setting where $r = 1$. This limitation is encapsulated by the position of the indifferent reached consumer, given by $\bar{\xi}$, as even the best films are forced to appeal only to the $1 - H(\bar{\xi})$ mass of interested consumers by playing the same advertising strategy as all other films that have $q > \bar{q}$. With increased r , as lesser films with quality $q < \bar{q}$ decrease their advertising, $\bar{\xi}$ will decrease and move closer to $T - u(q = 1)$ (and \bar{q} will increase). However, as shown in the two-type case, lesser quality films can still be incentivized to advertise even with $r = 1$, so the best quality films still will not collect their “fair share” of the theatrical market.

2.5.10 Two Periods

Another natural extension is to consider the film distributor’s advertising decision across the two time periods of its film’s life: the pre-release week ($t = 0$) and the post-release week ($t = 1$). In this case, the distributor of a film with characteristics w and realized quality θ makes advertising decisions a_t for $t = 0, 1$ which represent the pre-release and post-release phase. Assume that across both periods, the distributor faces the same mass M of consumers and the ticket price is constant at T .

The mass of consumers that are effectively reachable by advertising in the post-release period a_1 , given that a_0 was chosen in the pre-release period is limited to consumers who were not reached by advertising in $t = 0$. I assume that consumers who were already reached by advertising at $t = 0$ would not make a second decision if reached again in period $t = 1$. This is because consumers who have already seen the movie will not see the same movie twice, and consumers who were reached but were initially uninterested in the film would remain uninterested when reminded of the film. The mass of effectively reachable consumers is therefore: $M_1 = \eta M_0(1 - \exp\{-\frac{a}{bM_0}\}) \exp\{-\frac{a}{bM_0}\}$, where I also introduce a discount factor $\eta \in [0, 1)$ which captures the idea that some consumers would only consider seeing a movie during the opening week. The two period revenue for the film distributor is given by:

$$\begin{aligned}
E[R(a_0, a_1)|\theta] = & \\
& M(1 - \exp\{-\frac{a_0}{bM}\}) \int_{\bar{\xi}_0}^{+\infty} ((1 - r_0)T + r_0[\mathbb{I}(V_2(\theta) + \xi > 0) \cdot T])h(\xi)d\xi \\
& + \eta M(1 - \exp\{-\frac{a_1}{bM}\})(\exp\{-\frac{a_0}{bM}\}) \int_{\bar{\xi}_1}^{+\infty} ((1 - r_1)T + r_1[\mathbb{I}(V_2(\theta) + \xi > 0) \cdot T])h(\xi)d\xi \quad (2.28)
\end{aligned}$$

where $\bar{\xi}_t$ is the idiosyncratic shock of the indifferent consumer in period t .

The profit maximizing advertising levels for films with $\theta = G$ are:

$$a_{G,0}^* = \max \left(-bM \log \left\{ \frac{b}{T(1 - H(\bar{\xi}_0)) - \eta(1 - \exp(-\frac{a_{G,1}^*}{bM}))(1 - H(\bar{\xi}_1))} \right\}, 0 \right) \quad (2.29)$$

$$a_{G,1}^* = \max \left(-bM \log \left\{ \frac{b}{T\eta(1 - H(\bar{\xi}_1))} \right\} - a_{G,0}^*, 0 \right) \quad (2.30)$$

Similarly, those for films with $\theta = B$ are:

$$a_{B,0}^* = \max \left(-bM \log \left\{ \frac{b}{T(\bar{\Psi}_0(r_0) - \eta(1 - \exp(-\frac{a_{B,1}^*}{bM}))\bar{\Psi}_1(r_1))} \right\}, 0 \right) \quad (2.31)$$

$$a_{B,1}^* = \max \left(-bM \log \left\{ \frac{b}{T\eta(\bar{\Psi}_1(r_1))} \right\} - a_{B,0}^*, 0 \right) \quad (2.32)$$

where $\bar{\Psi}_t(r_t) = (1 - r_t)(1 - H(\bar{\xi}_t)) + r_t(1 - H(-V_2(B)))$ for brevity.

Film distributors' incentives to advertise in both periods depend on the value of η . For sufficiently low η , the model reverts to the single-period case. To establish the existence of an equilibrium in the two-period case, a similar argument invoking Brouwer's fixed-point theorem can be made as with the one-period case with certain assumptions about b and T . Uniqueness cannot be established using the same arguments as before because there are four equilibrium advertising levels instead of two.

2.6 Conclusion

I sought to find evidence of the practice of “hype” in the U.S. movie industry, to address why the practice is feasible, and to assess what impact its feasibility has on the movie industry. I accomplish these goals by examining a panel data set of movie characteristics, box office revenue, advertising expenditure, and critical review outcomes for wide release theatrical movies from 2003-2012. The OLS regression analysis and DID estimation results show that advertising levels are statistically indistinguishable during the pre-release phase across underlying quality levels. In the post-release phase, advertising strategies diverge, with the distributors of high quality films continuing to advertise substantially and those of low quality films abandoning their products. This behavior is consistent with film distributors not knowing their own quality, but I use data on the publication dates of critical reviews and data from the limited release market to show that this is unlikely to be the case. One aspect of the movie industry which could be

responsible for the feasibility of hype is the fact that critical reviews and word of mouth effects are limited during the opening week of a theatrical film.

I build a theoretical model of advertising in the movie industry based on Butters (1977) to explain the empirical results, using the availability of information on film quality as a key feature. After providing conditions for the existence and uniqueness of advertising equilibria in the model, I provide two numerical examples to show the relationship between advertising behavior and availability of information. For the distributors of bad films, as consumers become better informed about movie quality, their incentives to advertise decrease. As the distributors of bad films become less incentivized to advertise, the distributors of good films become more incentivized to advertise in equilibrium, as consumers become more likely to believe that a film is good when they are reached by advertising. This effect actually increases the incentives for the distributors of all film types to advertise, since the distributors of bad films “free ride” on the increased consumers confidence. The force that keeps the distributors of bad films from increasing their advertising levels is the increased verification probability, which causes first-order decrease on their revenue.

The equilibrium findings of the model can be used to support the empirical results in the sense that the pre-release phase can be represented by consumers having limited quality information about films, so that the observed advertising behaviors of the distributors of good and bad films are statistically indistinguishable. The post-release phase can be paired to a version of the model where the verification probability is high enough so that advertising strategies diverge and a separating equilibrium is reached. These results suggest that consumer access to critical review publications could help explain the feasibility of the practice of hype in the U.S. movie industry. The next question to answer is: how much are consumers in the real world impacted by the practice of hype? Estimation of utility demand in this model could lead to counterfactual analysis of the impact on welfare. The theoretical model can also be extended to the DVD/Blu-ray markets, as film distributors rekindle their advertising efforts when films leave theaters and become

available outside of theaters. The model also has possible extensions to the markets of other experience goods, such as books, music, and video games.

Table 2.1: Summary Statistics - Wide Releases

Variable	Obs	Mean	Std Dev	Min	Max
Budget (\$)	1110	52 million	47.3 million	500000	300 million
RT Score ($q \in [0, 100]$)	1110	43.6	25.75	0	99
Pre-release Ads (\$)	1110	15.41 million	7.93 million	47500	49.75 million
Post-release Ads (\$)	1110	1.75 million	2.05 million	0	14.51 million
Opening Weekend Gross(\$)	1110	20.4 million	21.9 million	443901	169 million

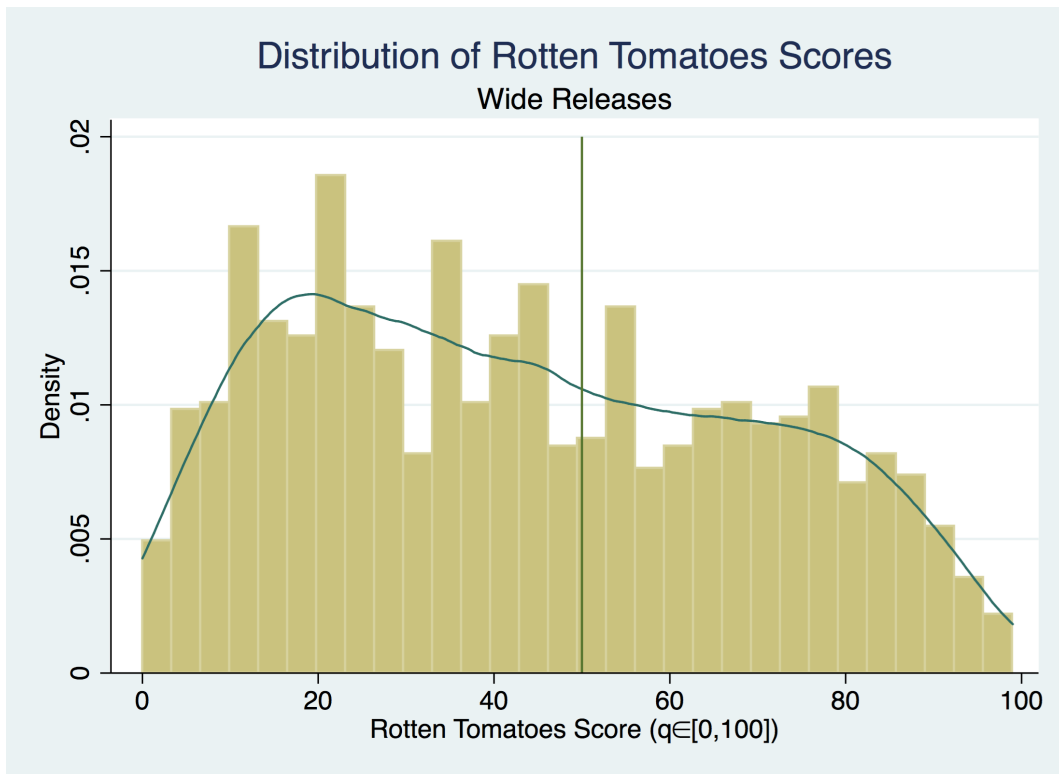
Summary statistics for the sub-sample of movies that had only wide theatrical releases in the domestic market. A movie’s release is considered to be wide if it opens in 600 or more theaters during its first weekend.

Table 2.2: Summary Statistics - Limited Releases (Followed by Wide)

Variable	Obs	Mean	Std Dev	Min	Max
Budget (\$)	118	26.3 million	26.3 million	15000	145 million
RT Score ($q \in [0, 100]$)	118	74.1	20.0	18	98
Pre-release Ads (\$)	118	8.98 million	7.24 million	46900	32.7 million
Post-release Ads (\$)	117	2.75 million	2.93 million	10500	13.83 million
Opening Weekend Gross(\$)	118	7.83 million	11.9 million	763234	102 million

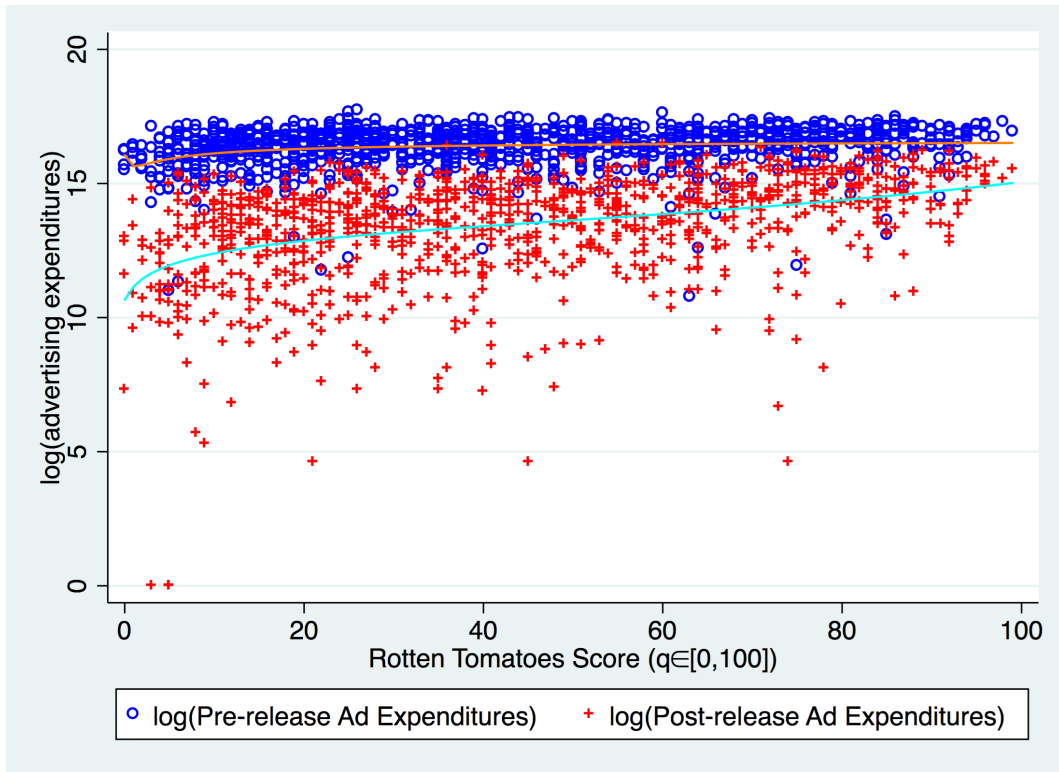
Summary statistics for the sub-sample of movies that initially had a limited release followed by a wide theatrical release in the domestic market. Pre and post-release advertising expenditures measured relative to wide theatrical opening week. This sub-sample of data is very incomplete, as observations that were not matched by web-scraping methods were dropped. Many movies that only had limited releases are not prioritized on search engines.

Figure 2.1: Distribution of Rotten Tomatoes Scores (Wide Releases)



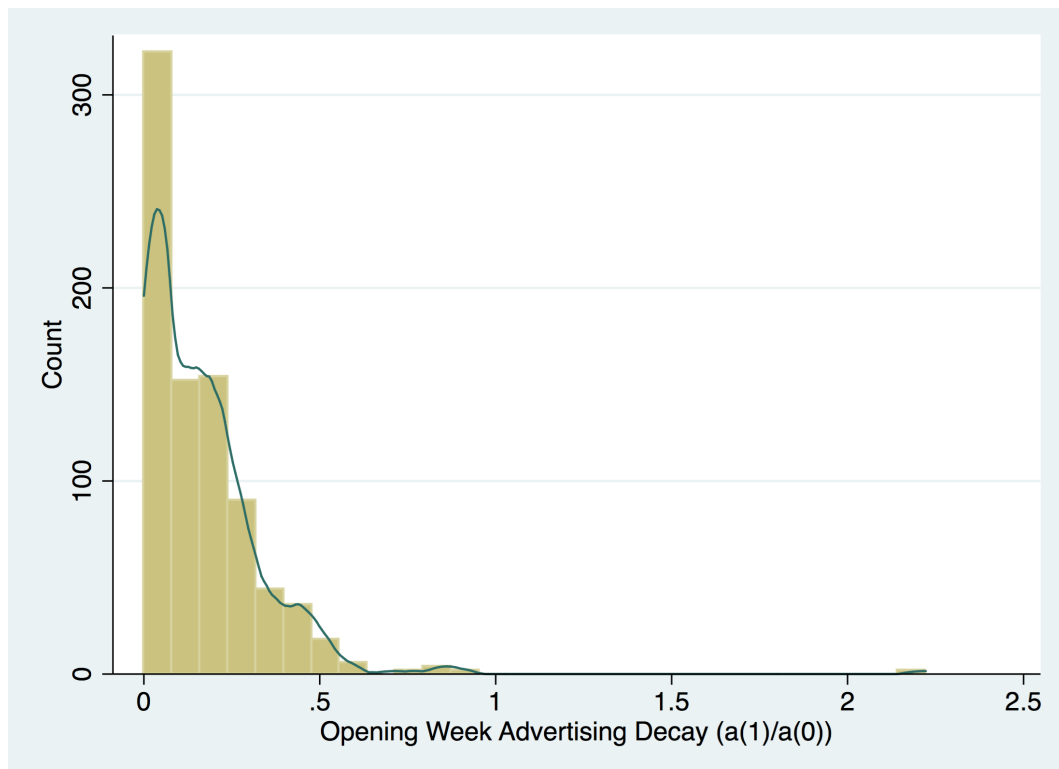
Histogram of Rotten Tomatoes score for all wide release films (2003-2012). Superimposed density function estimated with Epanechnikov kernel. Vertical line drawn at $q = 50$ separates movies labeled as $\theta = 1$ from those as $\theta = 0$.

Figure 2.2: Advertising Expenditures vs. Rotten Tomatoes Score (Wide Releases)



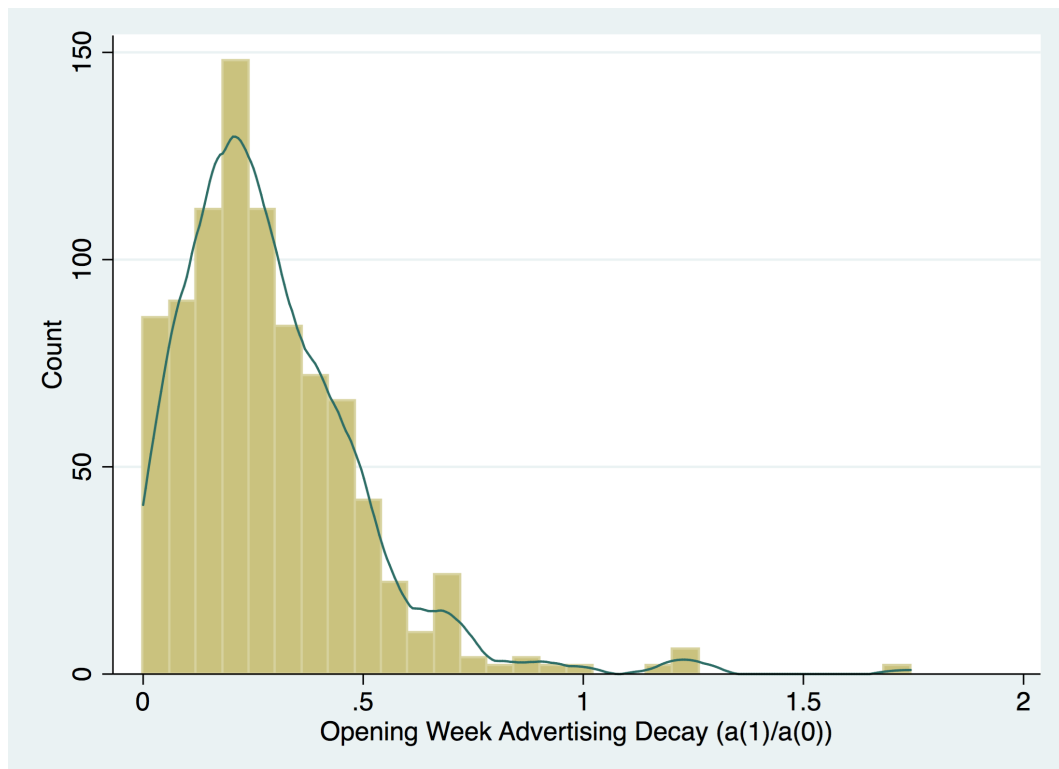
Scatterplots of pre-release and post-release advertising expenditure versus Rotten Tomatoes score for all wide release films in sample from 2003-2012. A movie's release is considered to be wide if it opens in 600 or more theaters during its first weekend. A fractional-polynomial prediction curve is estimated and superimposed for each plot.

Figure 2.3: Distribution of Opening Week Advertising Decay for $q \leq 30$ (Wide Releases)



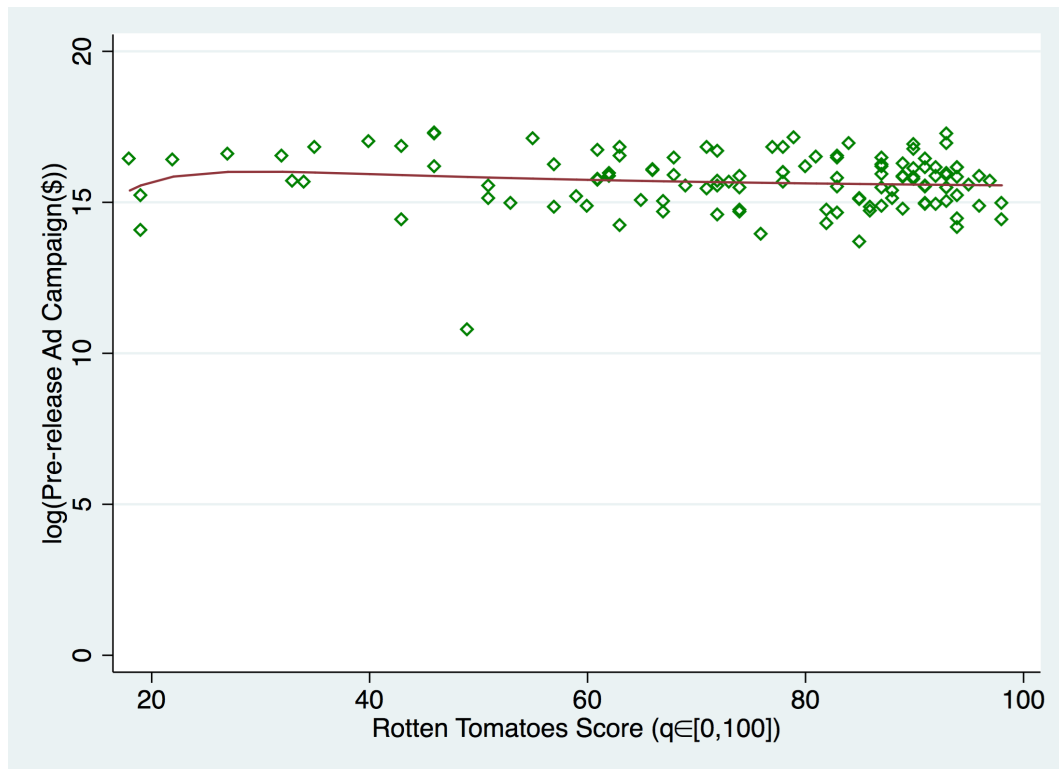
Histogram of decay in weekly advertising expenditures from the opening week to the second week of release for all wide release films in sample from 2003-2012 with Rotten Tomatoes scores below 30. Superimposed density function estimated with Epanechnikov kernel.

Figure 2.4: Distribution of Opening Week Advertising Decay for $q \geq 50$ (Wide Releases)



Histogram of decay in weekly advertising expenditures from the opening week to the second week of release for all wide release films in sample from 2003-2012 with Rotten Tomatoes scores at or above 50. Superimposed density function estimated with Epanechnikov kernel.

Figure 2.5: Pre-release Advertising Expenditures vs. Score (Limited then Wide Releases)



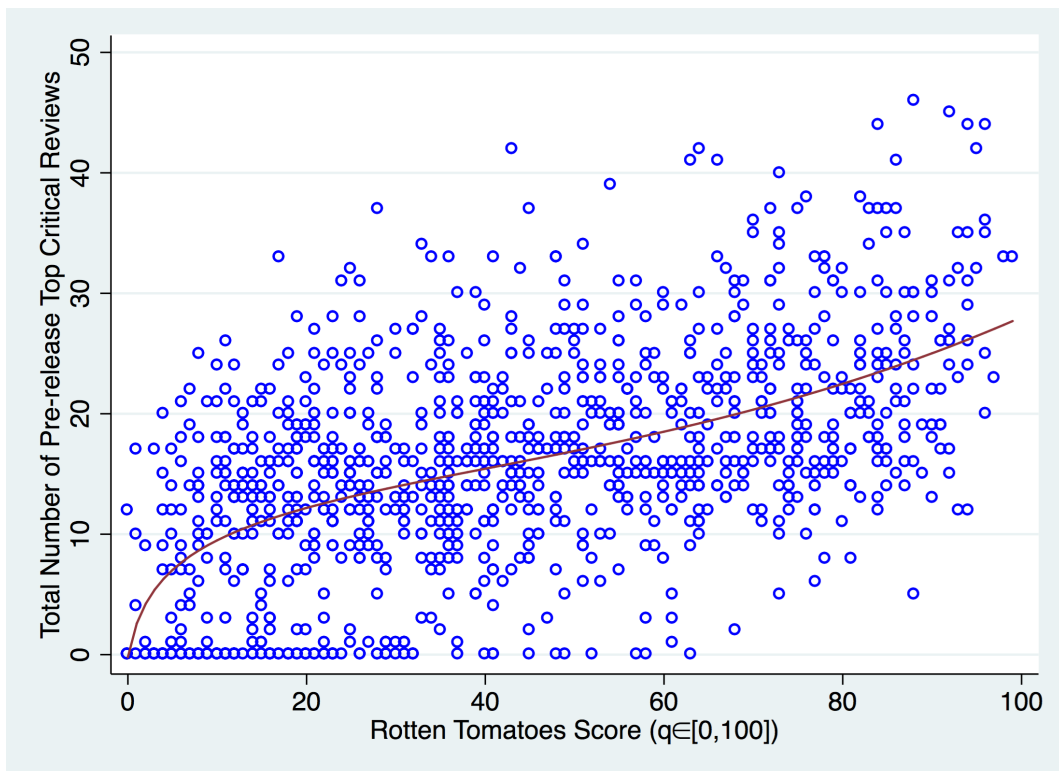
Scatterplot of pre-release advertising expenditures versus Rotten Tomatoes score for films that began with a limited release and then had a wide release at a later date from 2003-2012. A movie's release is considered to be wide if it opens in 600 or more theaters during its first weekend, and limited otherwise. A fractional-polynomial prediction curve is estimated and superimposed for each plot.

Table 2.3: Summary Statistics - Top Critical Review Publications

Variable	Obs	Mean	Std Dev	Min	Max
# Total Reviews	1110	35.5	9.969724	4	54
# Pre-release reviews	1110	15.86036	9.067096	0	46
# Opening day reviews	1110	13.59189	6.108438	0	37
# Post-release reviews	1110	6.047748	3.76382	0	35
$\frac{\# \text{ Pre-release reviews}}{\# \text{ Total reviews}}$	1110	.4138452	.2060107	0	.9361702
$\frac{\# \text{ Opening day reviews}}{\# \text{ Total reviews}}$	1110	.380977	.1515354	0	.8604651
$\frac{\# \text{ Post-release reviews}}{\# \text{ Total reviews}}$	1110	.2051777	.2050495	0	1

Summary statistics for the number and proportion of top critical reviews published during various stages of a film's life for sample of wide releases from 2003-2012. These numbers and proportions do not include reviews or assessments from other critics or the general public.

Figure 2.6: Number of Pre-release Critical Reviews vs. Score (Wide Releases)



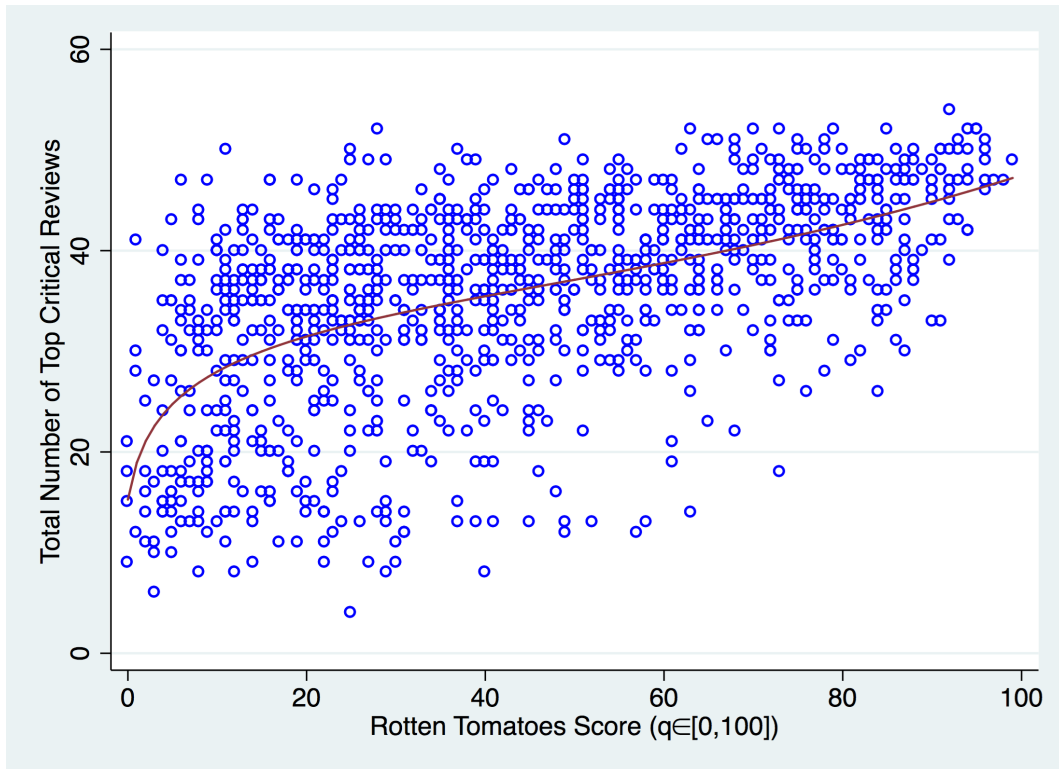
Number of top critical reviews published prior to the date of wide release versus Rotten Tomatoes score for all wide release films in sample from 2003-2012. A movie's release is considered to be wide if it opens in 600 or more theaters during its first weekend. A fractional-polynomial prediction curve is estimated and superimposed for each plot.

Figure 2.7: Proportion of Pre-release Critical Reviews vs. Score (Wide Releases)



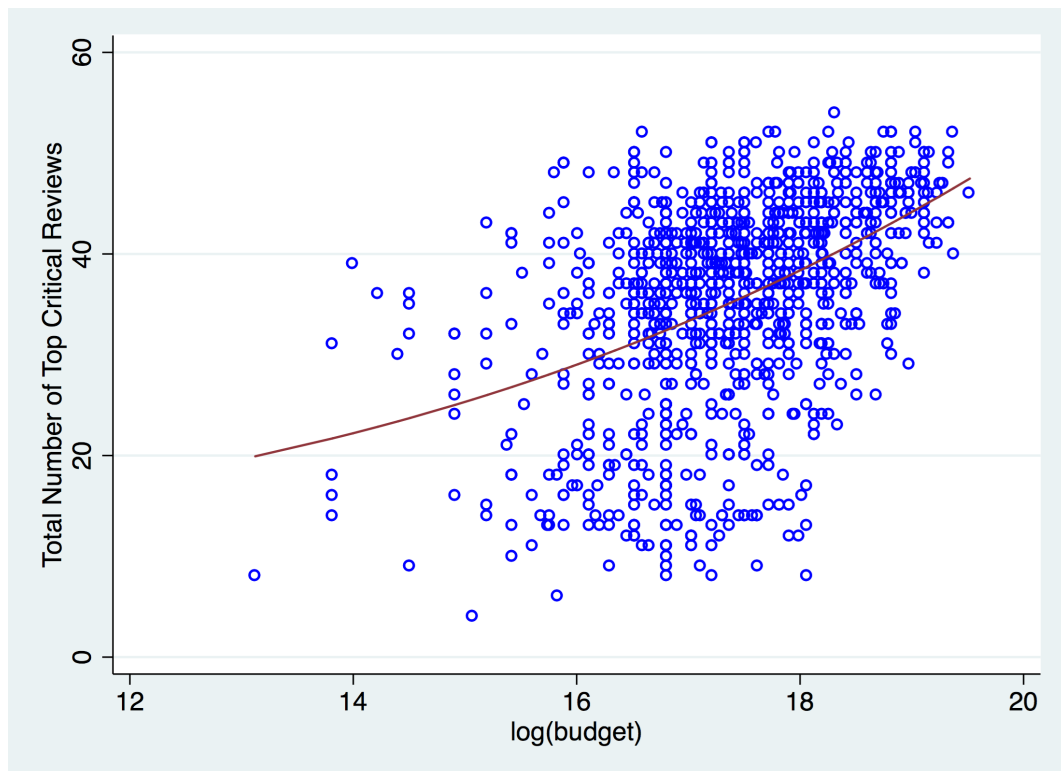
Proportion of top critical reviews that were published prior to the date of wide release versus Rotten Tomatoes score for all wide release films in sample from 2003-2012. A movie's release is considered to be wide if it opens in 600 or more theaters during its first weekend. A fractional-polynomial prediction curve is estimated and superimposed for each plot.

Figure 2.8: Total Number of Critical Reviews vs. Score (Wide Releases)



Scatterplot of number of review publications from top critics versus Rotten Tomatoes score for all wide release films in sample from 2003-2012. A reviewer is considered to be top if he or she is a contributor to publications in the top 10% of circulation. A movie's release is considered to be wide if it opens in 600 or more theaters during its first weekend. A fractional-polynomial prediction curve is estimated and superimposed for each plot.

Figure 2.9: Total Number of Critical Reviews vs. Budget (Wide Releases)



Scatterplot of number of review publications from top critics versus production budget for all wide release films in sample from 2003-2012. A movie's release is considered to be wide if it opens in 600 or more theaters during its first weekend. A fractional-polynomial prediction curve is estimated and superimposed for each plot.

Table 2.4: OLS Regression Results

Dep Var: $\log(a_{j\tau})$	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5	Spec. 6
RT Score ($q \in [0, 100]$)	.0054*** (.0016)	.0010 (.0015)	.0057*** (.0016)	.0023 (.0014)	.0016 (.0015)	.0028* (.0015)
I(After Release)	-3.9437*** (.1143)	-3.9437*** (.1031)	-3.9437*** (.1103)	-3.9437*** (.1013)	-3.9437*** (.1024)	-3.9437*** (.1007)
I(After Release) · Score	.0229*** (.0023)	.0229*** (.0020)	.0229*** (.0022)	.0229*** (.0020)	.0229*** (.0020)	.0229*** (.0020)
$\log(\text{production budget})$	-	.6482*** (.0287)	-	.6104*** (.0302)	.6522*** (.0335)	.6183*** (.0347)
Constant	16.1284*** (.0808)	5.0487*** (.4965)	16.3041*** (.1454)	5.8509*** (.5345)	4.7331*** (.5903)	5.4589*** (.6195)
Movie Fixed Effects?	N	N	N	N	Y	Y
Time Fixed Effects?	N	N	Y	Y	N	Y
Adj. R^2	0.5656	0.6466	0.5960	0.6592	0.6518	0.6627

OLS regression results. Advertising expenditure considered as either pre-release era or post-release era, depending on if the advertising was done during or before the film's opening week or not. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 2.5: IV Regression Analysis Results - Instrumenting for Score

Dep Var: $\log(a_{jT})$	Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5	Spec. 6
RT Score ($q \in [0, 100]$)	.0219*** (.0034)	.0050* (.0029)	.0156*** (.0031)	.0020 (.0028)	.0066* (.0030)	.0031 (.0028)
$\mathbb{I}(\text{After Release})$	-4.9986*** (.2213)	-4.9986*** (.1865)	-4.9986*** (.2034)	-4.9986*** (.1784)	-4.9986*** (.1852)	-4.9986*** (.1775)
$\mathbb{I}(\text{After Release}) \cdot \text{Score}$.0471*** (.0048)	.0471*** (.0041)	.0471*** (.0044)	.0471*** (.0039)	.0471*** (.0041)	.0471*** (.0039)
$\log(\text{production budget})$	-	.5654*** (.0325182)	-	.5556*** (.0328)	.5674*** (.0375)	.5631*** (.0373)
Constant	15.4076*** (0.1565)	6.3148*** (.5493)	15.9036*** (0.1949)	6.8329*** (0.5717)	6.0699*** (0.6500)	6.4725*** (0.6601)
Movie Fixed Effects?	N	N	N	N	Y	Y
Time Fixed Effects?	N	N	Y	Y	N	Y
Adj. R^2	0.4170	0.5856	0.5021	0.6165	0.5889	0.6194

IV regression results, with number of published reviews used as instrument for Rotten Tomatoes score. Advertising expenditure is measured as pre-release or post-release total. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 2.6: IV Regression Analysis Results - Instrumenting for Budget

Dep Var: $\log(a_{j,t})$	Spec. 1	Spec. 2	Spec. 3	Spec. 4
$\log(\text{production budget})$	1.2361*** (.0744)	1.3095*** (.0812)	1.1620*** (.0829)	1.195*** (.0844)
RT Score ($q \in [0, 100]$)	-.0029* (.0017)	-.0017 (.0016)	-.0007 (.0016)	.0003 (.0016)
$\mathbb{I}(\text{After Release})$	-3.9437*** (.1123)	-3.9437*** (.1106)	-3.9437*** (.1081)	-3.9437*** (.1061)
$\mathbb{I}(\text{After Release}) \cdot \text{Score}$.0229*** (.0022)	.0229*** (.0022)	.0229*** (.0021)	.0229*** (.0021)
Constant	-5.0007*** (1.2747)	-6.0754*** (1.3552)	-3.5961** (1.4272)	-4.038*** (1.413)
Movie Fixed Effects?	N	Y	N	Y
Time Fixed Effects?	N	N	Y	Y
Adj. R^2	0.5798	0.5911	0.6075	0.6201

Table 2.7: IV regression results, with number of published reviews used as instrument for film's production budget. Advertising expenditure is measured as pre-release or post-release total. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 2.8: Balance test of covariates of propensity score method

Weighted Variable	Mean Control	Mean Treated	Diff.	t	Pr(T > t)
log(Opening Week Ad Expenditure)	8.196	8.271	0.076	1.90	0.0578*
log(Total Pre-release Ad Expenditure)	16.396	16.477	0.081	1.80	0.0723*
log(Opening Week Revenue)	16.653	16.666	0.013	0.24	0.8122
log(budget)	17.495	17.534	0.039	0.68	0.4980
# Opening Theaters	2788.300	2795.296	6.996	0.15	0.8833
Genre: Action	0.131	0.130	-0.001	0.07	0.9430
Genre: Adventure	0.165	0.173	0.008	0.36	0.7219
Genre: Comedy	0.277	0.294	0.017	0.62	0.5323
Genre: Drama	0.221	0.206	-0.015	0.62	0.5371
Genre: Thriller	0.135	0.132	-0.003	0.13	0.9002
Rating: G	0.032	0.029	-0.003	0.28	0.7775
Rating: PG	0.205	0.202	-0.003	0.14	0.8912
Rating: PG-13	0.396	0.388	-0.008	0.28	0.7774
Rating: R	0.367	0.381	0.015	0.50	0.6172
Month: January	0.054	0.052	-0.003	0.20	0.8421
Month: February	0.068	0.061	-0.007	0.48	0.6299
Month: March	0.085	0.090	0.005	0.30	0.7610
Month: April	0.086	0.087	0.001	0.09	0.9294
Month: May	0.074	0.074	-0.000	0.00	0.9992
Month: June	0.080	0.083	0.003	0.19	0.8526
Month: July	0.073	0.085	0.012	0.73	0.4637
Month: August	0.109	0.108	-0.001	0.07	0.9417
Month: September	0.106	0.092	-0.014	0.78	0.4369
Month: October	0.096	0.108	0.011	0.63	0.5316
Month: November	0.100	0.081	-0.019	1.12	0.2640
Month: December	0.069	0.081	0.011	0.72	0.4731
Year: 2003	0.075	0.076	0.001	0.08	0.9339
Year: 2004	0.089	0.085	-0.004	0.23	0.8183
Year: 2005	0.094	0.094	0.000	0.02	0.9839
Year: 2006	0.097	0.103	0.007	0.37	0.7139
Year: 2007	0.110	0.105	-0.005	0.27	0.7906
Year: 2008	0.105	0.105	0.001	0.04	0.9718
Year: 2009	0.088	0.096	0.009	0.51	0.6091
Year: 2010	0.114	0.103	-0.011	0.57	0.5693
Year: 2011	0.121	0.117	-0.004	0.22	0.8256
Year: 2012	0.108	0.114	0.006	0.31	0.7529

Balancing t-test of the difference in means of the covariates between the control and the treatment groups in the pre-release period. # Opening Theaters is the number of theaters in which the movie was shown during its opening weekend. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 2.9: Propensity Score of Treatment (Receiving $\theta > 50$)

Number of obs	=	1110
LR $\chi^2(30)$	=	103.44
Prob > χ^2	=	0.0000
Pseudo R^2	=	0.0692
Log likelihood	=	-696.12426

Variable	Coeff.	S.E.	z	P> z	95% Conf. Interval	
log(budget)	.2384712**	.105463	2.26	0.024	.0317675	.445175
# Opening Theaters	.0001851	.0001218	1.52	0.128	-.0000536	.0004237
Genre: Action	.2420359	.3077997	0.79	0.432	-.3612404	.8453122
Genre: Adventure	1.213984***	.3445565	3.52	0.000	.5386658	1.889302
Genre: Comedy	.545758**	.2719847	2.01	0.045	.0126778	1.078838
Genre: Drama	1.352327***	.3005705	4.50	0.000	.7632195	1.941434
Genre: Thriller	.7764723***	.2991643	2.60	0.009	.1901211	1.362824
Rating: PG	-.438181	.4528334	-0.97	0.333	-1.325718	.4493561
Rating: PG-13	-.4176297	.4497696	-0.93	0.353	-1.299162	.4639024
Rating: R	.3968301	.4612546	0.86	0.390	-.5072122	1.300872
Month: January	-.5024946	.3283555	-1.53	0.126	-1.14606	.1410702
Month: February	-.5873263*	.3171477	-1.85	0.064	-1.208924	.0342717
Month: April	.1962501	.3044046	0.64	0.519	-.4003719	.7928721
Month: May	-.0166503	.3242146	-0.05	0.959	-.6520992	.6187986
Month: June	.1239696	.3174349	0.39	0.696	-.4981913	.7461306
Month: July	.1087406	.3110717	0.35	0.727	-.5009488	.71843
Month: August	.1449293	.2893278	0.50	0.616	-.4221428	.7120015
Month: September	.1588033	.2997185	0.53	0.596	-.4286341	.7462407
Month: October	.1247108	.2992248	0.42	0.677	-.461759	.7111807
Month: November	-.075619	.3152223	-0.24	0.810	-.6934434	.5422054
Month: December	.2790358	.3290269	0.85	0.396	-.3658452	.9239167
Year: 2004	.0074037	.3048647	0.02	0.981	-.5901201	.6049275
Year: 2005	.1886319	.3053051	0.62	0.537	-.409755	.7870188
Year: 2006	.1572218	.2988003	0.53	0.599	-.4284161	.7428596
Year: 2007	.2570071	.299104	0.86	0.390	-.329226	.8432401
Year: 2008	.177176	.2971969	0.60	0.551	-.4053193	.7596713
Year: 2009	.1404446	.302557	0.46	0.643	-.4525562	.7334453
Year: 2010	.3094302	.3046392	1.02	0.310	-.2876518	.9065121
Year: 2011	.2871562	.2952324	0.97	0.331	-.2914888	.8658011
Year: 2012	.5875353	.3091615	1.90	0.057	-.0184101	1.193481
Constant	-5.838799***	1.707097	-3.42	0.001	-9.184647	-2.49295

Propensity score logit estimation results. Dummies for the MPAA G-rating, the month of March, the year of 2003, and the genre of horror were dropped due to collinearity. # Opening Theaters is the number of theaters in which the movie was shown during its opening weekend. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 2.10: Propensity Score of Treatment (Receiving $\theta > 50$) with Opening Week Box Office Revenue

Number of obs	=	1110
LR $\chi^2(31)$	=	191.82
Prob > χ^2	=	0.0000
Pseudo R^2	=	0.1282
Log likelihood	=	-651.93866

Variable	Coeff.	S.E.	z	P> z	95% Conf. Interval
log(Opening Week Revenue)	1.284005***	.1460445	8.79	0.000	.9977629 1.570247
log(budget)	.1491839	.1094998	1.36	0.173	-.0654317 .3637995
# Opening Theaters	-.0010583***	.0001894	-5.59	0.000	-.0014296 -.0006871
Genre: Action	.3606117	.3227513	1.12	0.264	-.2719691 .9931926
Genre: Adventure	1.343368***	.3614226	3.72	0.000	.6349928 2.051743
Genre: Comedy	.6951069**	.2842806	2.45	0.014	.1379271 1.252287
Genre: Drama	1.310918***	.3124332	4.20	0.000	.6985605 1.923276
Genre: Thriller	1.002537***	.3126379	3.21	0.001	.3897777 1.615296
Rating: PG	-.5211994	.4771949	-1.09	0.275	-1.456484 .4140855
Rating: PG-13	-.9398258**	.4755157	-1.98	0.048	-1.871819 -.0078322
Rating: R	-.0560434	.4864292	-0.12	0.908	-1.009427 .8973403
Month: January	-.6263089*	.3396523	-1.84	0.065	-1.292015 .0393975
Month: February	-.6298854*	.3290097	-1.91	0.056	-1.274732 .0149617
Month: April	.3778482	.3164718	1.19	0.233	-.2424252 .9981215
Month: May	-.0494721	.3380849	-0.15	0.884	-.7121062 .6131621
Month: June	.0549268	.3292872	0.17	0.868	-.5904643 .7003179
Month: July	.0837933	.3237407	0.26	0.796	-.5507267 .7183134
Month: August	.4565973	.303159	1.51	0.132	-.1375834 1.050778
Month: September	.4076157	.3100991	1.31	0.189	-.2001673 1.015399
Month: October	.2654907	.3103285	0.86	0.392	-.3427419 .8737233
Month: November	-.0959973	.3281993	-0.29	0.770	-.7392561 .5472615
Month: December	.5041878	.343455	1.47	0.142	-.1689717 1.177347
Year: 2004	-.0499494	.3159975	-0.16	0.874	-.6692932 .5693944
Year: 2005	.2653811	.3173283	0.84	0.403	-.3565708 .8873331
Year: 2006	.2306686	.3082774	0.75	0.454	-.3735439 .8348812
Year: 2007	.3553111	.3095534	1.15	0.251	-.2514024 .9620247
Year: 2008	.2891829	.3073646	0.94	0.347	-.3132406 .8916063
Year: 2009	.0773851	.3129466	0.25	0.805	-.5359789 .6907492
Year: 2010	.3967424	.3162163	1.25	0.210	-.2230301 1.016515
Year: 2011	.495118	.3067892	1.61	0.107	-.1061778 1.096414
Year: 2012	.6893823	.3209481	2.15	0.032	.0603357 1.318429
Constant	-21.7545***	2.569406	-8.47	0.000	-26.79044 -16.71856

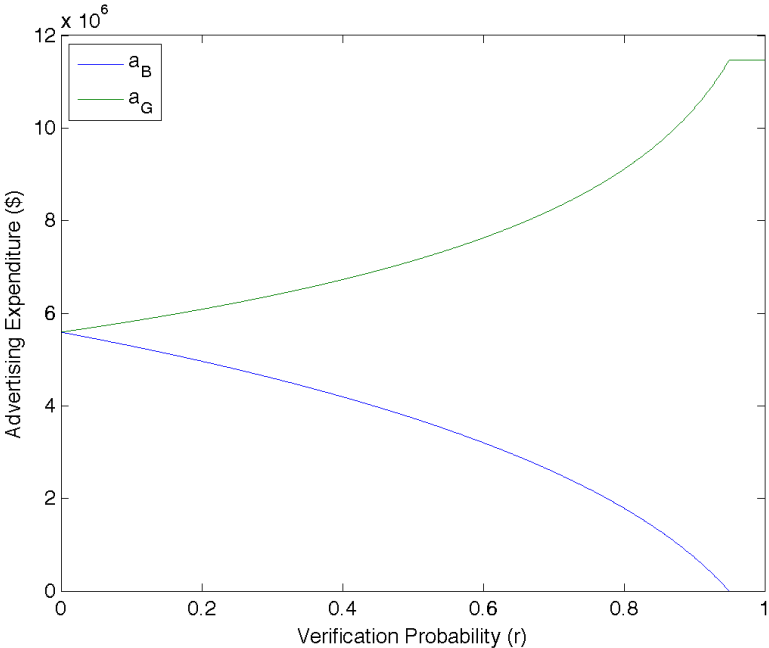
Propensity score logit estimation results. Specification includes opening weekend box office revenue as a matching covariate. Dummies for the MPAA G-rating, the month of March, the year of 2003, and the genre of horror were dropped due to collinearity. # Opening Theaters is the number of theaters in which the movie was shown during its opening weekend. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 2.11: DID Estimation Results

Outcome Var: log(advertising expenditures)		Spec. 1	Spec. 2	Spec. 3	Spec. 4
Pre-Release	Control	8.205 (0.051)	16.412 (0.057)	8.196 (0.049)	16.396 (0.056)
	Treated	8.271 (0.051)	16.477 (0.057)	8.271 (0.049)	16.477 (0.055)
	Diff (Baseline)	0.066 (0.071)	0.065 (0.081)	0.076 (0.069)	0.081 (0.078)
Post-Release	Control	5.838 (0.051)	13.116 0.057	6.123 (0.049)	13.461 (0.056)
	Treated	6.678 (0.051)	14.117 (0.057)	6.678 (0.049)	14.117 (0.055)
	Diff (Follow-up)	0.841*** (0.071)	1.002*** (0.081)	0.555*** (0.069)	0.656*** (0.078)
DID		0.775*** (0.101)	0.937*** (0.114)	0.480*** (0.098)	0.575*** (0.111)
R^2		0.4306	0.5394	0.3982	0.5141
Weekly/Era Ads		Weekly	Era	Weekly	Era
Opening Box Office Returns		N	N	Y	Y

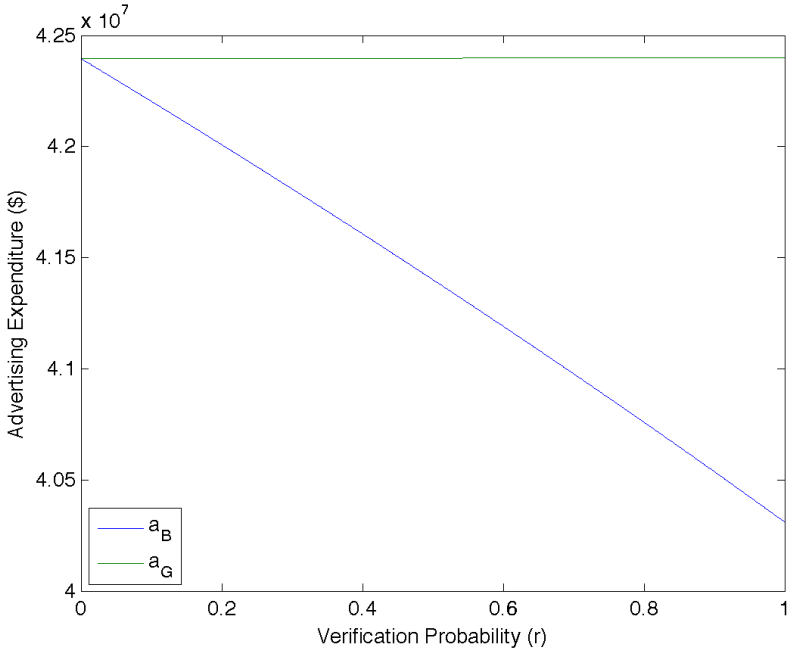
Difference-in-differences estimation results. Cutoff threshold of θ for treatment is 50. Advertising expenditure is measured as opening week or second release of release spending for Specifications 1 and 3. Advertising expenditure is measured as pre-release or post-release total for Specifications 2 and 4. Standard errors are in parentheses. * Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Figure 2.10: Advertising Strategies vs. Verification Probability (Numerical Solution 1)



Graph of numerical solution with $U(\theta = G) = 4.0, U(\theta = B) = 3.74, T = \$5.3, M = 230.7$ million, $b = 0.05, \mu = 0.5$, and a Type I extreme value distribution assumption on ξ . The green and blue lines indicate the equilibrium advertising strategies of the distributors of good and bad films, respectively, as a function of consumer quality verification probability.

Figure 2.11: Advertising Strategies vs. Verification Probability (Numerical Solution 2)



Graph of numerical solution with $U(\theta = G) = 5.48, U(\theta = B) = 5.144, T = \$5.3, M = 230.7$ million, $b = 0.05, \mu = 0.5$, and a Type I extreme value distribution assumption on ξ . The green and blue lines indicate the equilibrium advertising strategies of the distributors of good and bad films, respectively, as a function of consumer quality verification probability.

APPENDIX

A.1 Asymptotic Variance of Two Step Estimator

The moments associated with the two-step estimation described in Chapter 1 are derived from the first order conditions of the log-likelihood functions in both the first and the second steps. They are given as follows:

Step 1:

$$\begin{aligned} & \frac{\partial}{\partial \gamma} \sum_{m=1}^M \sum_{i=1}^{n_m} \log(\Phi(x'_i \gamma)^{y_i} (1 - \Phi(x'_i \gamma))^{1-y_i}) = 0 \\ \rightarrow & \frac{1}{\sum_{m=1}^M n_m} \sum_{m=1}^M \sum_{i=1}^{n_m} \frac{\partial}{\partial \gamma} \log(\Phi(x'_i \gamma)^{y_i} (1 - \Phi(x'_i \gamma))^{1-y_i}) = 0 \end{aligned}$$

Step 2:

$$\begin{aligned} & \frac{\partial}{\partial \theta} \sum_{m=1}^M \sum_{i=1}^{n_m} \log\left(\left(\frac{\exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1 | s_m))}{1 + \exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1 | s_m))}\right)^{y_i} \cdot \right. \\ & \qquad \qquad \qquad \left. \left(\frac{1}{1 + \exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1 | s_m))}\right)^{1-y_i}\right) = 0 \\ \rightarrow & \frac{1}{\sum_{m=1}^M n_m} \sum_{m=1}^M \sum_{i=1}^{n_m} \frac{\partial}{\partial \theta} \log\left(\left(\frac{\exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1 | s_m))}{1 + \exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1 | s_m))}\right)^{y_i} \cdot \right. \\ & \qquad \qquad \qquad \left. \left(\frac{1}{1 + \exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1 | s_m))}\right)^{1-y_i}\right) = 0 \end{aligned}$$

where $\theta = (\beta, \delta)'$, β and γ are column vectors of parameters, δ is a scalar, s_i is the individual state variable, $s_m = (s_1, \dots, s_{n_m})$ is the market state variable, and x_i is a subvector of s_i which includes additional market state variables.

For notational brevity, we let $n = \sum_{m=1}^M n_m$ and renumber the observations to write

these conditions as:

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \gamma} \log(\Phi(x'_i \gamma)^{y_i} (1 - \Phi(x'_i \gamma))^{1-y_i}) = 0$$

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log\left(\left(\frac{\exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1|s))}{1 + \exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1|s))}\right)^{y_i} \left(\frac{1}{1 + \exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1|s))}\right)^{1-y_i}\right) = 0$$

Assuming that all $z_i = (x_i, s_i, y_i)$ are i.i.d., the law of large numbers yields:

$$\mathbb{E}(m(z_i; \gamma_0)) = 0 \tag{A.33}$$

$$\mathbb{E}(g(z_i; \theta_0, \gamma_0)) = 0 \tag{A.34}$$

where

$$\begin{aligned} m(z_i; \gamma_0) &= \frac{\partial}{\partial \gamma} \log(\Phi(x'_i \gamma)^{y_i} (1 - \Phi(x'_i \gamma))^{1-y_i}) \\ &= \frac{\partial}{\partial \gamma} \log(f_1(z_i, \gamma_0)) \\ g(z_i; \theta_0, \gamma_0) &= \frac{\partial}{\partial \theta} \log\left(\left(\frac{\exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1|s_m))}{1 + \exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1|s_m))}\right)^{y_i} \left(\frac{1}{1 + \exp(s'_i \beta + \delta \sum_{j \neq i} \hat{\sigma}(a_j = 1|s_m))}\right)^{1-y_i}\right) \\ &= \frac{\partial}{\partial \theta} \log(f_2(z_i, \theta_0, \gamma_0)) \end{aligned}$$

A feasible estimator for the moment condition given in Equation A.34 is given by:

$$\frac{1}{n} \sum_{i=1}^n g(z_i; \hat{\theta}, \hat{\gamma}) = 0$$

Invoking the mean value theorem, an expansion around the true value θ_0 yields the condition:

$$\frac{1}{n} \sum_{i=1}^n g(z_i; \theta_0, \hat{\gamma}) + \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial g(z_i; \theta^*, \hat{\gamma})}{\partial \theta'}\right) (\hat{\theta} - \theta_0) = 0$$

where θ^* lies between (or on the line connecting) θ_0 and $\hat{\theta}$. Rearranging this expression yields:

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial g(z_i; \theta^*, \hat{\gamma})}{\partial \theta'}\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \hat{\gamma})\right) \tag{A.35}$$

The second term on the right hand side of Equation A.35 can be expanded around the true value γ_0 in a similar way:

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \hat{\gamma}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \gamma_0) + \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial g(z_i; \theta_0, \gamma^*)}{\partial \gamma'} \right) (\hat{\gamma} - \gamma_0) \\ \rightarrow \frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \hat{\gamma}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \gamma_0) + \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial g(z_i; \theta_0, \gamma^*)}{\partial \gamma'} \right) \sqrt{n} (\hat{\gamma} - \gamma_0) \end{aligned}$$

where γ^* lies between γ_0 and $\hat{\gamma}$.

We also have as an estimator for the moment condition given in Equation A.33, which can be expanded by again appealing to the Mean Value Theorem:

$$\frac{1}{n} \sum_{i=1}^n m(z_i; \hat{\gamma}) = 0$$

$$\rightarrow 0 = \frac{1}{n} \sum_{i=1}^n m(z_i; \gamma_0) + \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial m(z_i; \gamma^*)}{\partial \gamma'} \right) (\hat{\gamma} - \gamma_0) \quad (\text{A.36})$$

$$\rightarrow \sqrt{n}(\hat{\gamma} - \gamma_0) = - \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial m(z_i; \gamma^*)}{\partial \gamma'} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n m(z_i; \gamma_0) \quad (\text{A.37})$$

Then, combining Equations A.35, A.36, and A.37 yields:

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &= - \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial g(z_i; \theta^*, \hat{\gamma})}{\partial \theta'} \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \gamma_0) \right) \\ &+ \frac{1}{n} \sum_{i=1}^n \frac{\partial g(z_i; \theta_0, \gamma^*)}{\partial \gamma'} \left(- \left(\frac{1}{n} \sum_{i=1}^n \frac{\partial m(z_i; \gamma^*)}{\partial \gamma'} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n m(z_i; \gamma_0) \right) \end{aligned} \quad (\text{A.38})$$

Assuming that $\hat{\theta}$ and $\hat{\gamma}$ are consistent and under some regularity conditions, we can rewrite this result as:

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &= - \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial g(z_i; \theta_0, \gamma_0)}{\partial \theta'} + o_p(1) \right) \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \gamma_0) \right) \\ &+ \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial g(z_i; \theta_0, \gamma_0)}{\partial \gamma'} + o_p(1) \right) \left(- \left(\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial m(z_i; \gamma_0)}{\partial \gamma'} + o_p(1) \right) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n m(z_i; \gamma_0) \right) \end{aligned} \quad (\text{A.39})$$

Denote:

$$G_\theta = \mathbb{E}\left(\frac{\partial g(z_i; \theta_0, \gamma_0)}{\partial \theta'}\right), G_\gamma = \mathbb{E}\left(\frac{\partial g(z_i; \theta_0, \gamma_0)}{\partial \gamma'}\right)$$

With z_i being i.i.d., the law of large numbers and Slutsky's theorem together yield:

$$\sqrt{n}(\hat{\theta} - \theta_0) = -G_\theta^{-1}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \gamma_0)\right) - G_\theta^{-1}G_\gamma\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (-\mathbb{E}\left(\frac{\partial m(z_i; \gamma_0)}{\partial \gamma'}\right)^{-1})m(z_i; \gamma_0)\right) + o_p(1)$$

Call $\alpha(z_i) = G_\gamma \psi(z_i)$ where $\psi(z_i) = \mathbb{E}\left(\frac{\partial m(z_i; \gamma_0)}{\partial \gamma'}\right)^{-1}m(z_i; \gamma_0)$.

$$\sqrt{n}(\hat{\theta} - \theta_0) = -G_\theta^{-1}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n g(z_i; \theta_0, \gamma_0) + \alpha(z_i)\right) + o_p(1)$$

The central limit theorem yields:

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta_0) &\rightarrow_d N(0, G_\theta^{-1}Var(g(z_i; \theta_0, \gamma_0) + \alpha(z_i))G_\theta^{-1})' \\ &= {}_d N(0, G_\theta^{-1}\mathbb{E}(g(z_i; \theta_0, \gamma_0)g(z_i; \theta_0, \gamma_0)' + g(z_i; \theta_0, \gamma_0)\alpha(z_i)' + \alpha(z_i)g(z_i; \theta_0, \gamma_0)' + \alpha(z_i)\alpha(z_i)')G_\theta^{-1}) \end{aligned}$$

B.2 Compilation of Movie Industry Data

With the exception of the advertising expenditures that were provided by Kantar Media, all of the data used in Chapter 2 were obtained from the internet. The box office revenue numbers were downloaded from The Numbers using a web-scraping tool called *Mozenda*⁹, which can capture web-browsing actions from the user and then repeat them autonomously. The Numbers features extensive historical box office revenue data for films and for distributors alike. For box office performance, the website has a separate page for each weekend at the box office, ranking the top 50 movies by their revenue from that weekend. In addition to box office revenue for each film, the webpage lists the number

⁹Available at <http://www.mozenda.com/>.

of theaters at which the movie played, lifetime box office gross, and revenue decay from the previous week. For a given weekend, the program was instructed to append the box office revenue data to a master data set, navigate to the next weekend, and repeat the process. This would continue until the last week in the sample time period is reached. I merged these data with the advertising expenditure data set, giving careful attention to possible errors or alternatives in naming. The data were matched in a way so that for a given week, the advertising expenditures figure correspond to spending from Monday until Sunday, while the box office revenue correspond to tickets sold from Friday until Sunday.

Next, the Python package “Beautiful Soup”¹⁰ was used to collect movie characteristics, review, and budget data. Python allows the user to create a pseudo-instance of a web browser and navigate websites. The package specializes in parsing HTML and XML code so that their contents can be interpreted and manipulated more easily in coding environments. The first step in this data collection process was to input the titles of each of the films in my data set and navigate to the first search result. For each IMDb movie webpage, the official IMDb movie title, the MPAA rating, the genre, the production budget (if available), the release date, and the names of the director and first three actors listed on the bill were recorded. The same process was done using the Rotten Tomatoes website and its search engine to collect “tomatometer” scores for each film, as well as the number of top critical reviews published before, on the day of, and after the film’s release.

To accurately match titles from the box office revenue/advertising data set with those collected from IMDb and Rotten Tomatoes, the exhibition years were compared as a first test. I then used FuzzyWuzzy, which is another Python package which takes two inputs and returns a matching score with a range from 0 to 100, which represents the package’s confidence that the two inputs refer to the same phrase or name. This was particularly useful in verifying movie name that were widely accepted truncations of official movie titles. Movies that were found to be improperly matched were re-searched

¹⁰Available at <http://www.crummy.com/software/BeautifulSoup/>.

with the movie's year (according to the box office revenue/advertising data set) on the appropriate search bar. For movies that still weren't properly matched, Google's search engine was used.

Lastly, for movies that were missing a budget on IMDb but had matching advertising expenditures, movie characteristics, and scores, the budgets were obtained via automated search, and eventually manual search, from The Numbers, which also has individual webpages for each movie that the website had box office revenue data for. For my main data set, I kept movies that opened with 600 or more theaters in its first appearance in the box office revenue data and movies for which I could find all matching data.

REFERENCES

- [1] Franklin Allen. Reputation and product quality. *RAND Journal of Economics, The RAND Corporation*, 15(3):311–327, 1984.
- [2] Patrick Bajari, Han Hong, John Krainer, and Denis Nekipelov. Estimating Static Models of Strategic Interaction. *SSRN eLibrary*, 2006.
- [3] Patrick Bajari, Han Hong, and Denis Nekipelov. Game theory and econometrics: A survey of some recent research. 2004.
- [4] Heski Bar-Isaac. Imperfect competition and reputational commitment. *Economics Letters*, 89(2):167–173, November 2005.
- [5] Suman Basuroy, Kalpesh Kaushik Desai, and Debabrata Talukdar. An empirical investigation of signaling in the motion picture industry. *Journal of Marketing Research*, 43(2):287–295, 2006.
- [6] Lori S. Benneer and Sheila M. Olmstead. The impacts of the “right to know”: Information disclosure and the violation of drinking water standards. *Journal of Environmental Economics and Management*, 56(2):117 – 130, 2008.
- [7] S. Berry and J. Waldfogel. Product quality and market size. *Journal of Industrial Economics*, 58(1):1–31, 2010.
- [8] Steven T Berry. Estimation of a model of entry in the airline industry. *Econometrica*, 60(4):889–917, July 1992.
- [9] T. Bresnahan and P. Reiss. Empirical models of discrete games. *Journal of Econometrics*, 48:57–81, 1991.
- [10] Alexander L. Brown, Colin F. Camerer, and Dan Lovallo. To review or not to review? limited strategic thinking at the movie box office. *American Economic Journal: Microeconomics*, 4(2):pp. 1–26, 2012.
- [11] Gerard R. Butters. Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies*, 44(3):pp. 465–491, 1977.
- [12] David Dranove, Daniel Kessler, Mark McClellan, and Mark Satterthwaite. Is more information better? the effects of "report cards" on health care providers. *Journal of Political Economy*, 111(3):pp. 555–588, 2003.
- [13] Liran Einav. Not all rivals look alike: Estimating an equilibrium model of the release date timing game. *Economic Inquiry*, 48(2):369–390, 2010.
- [14] Anita Elberse and Bharat N. Anand. The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market. *Information Economics and Policy*, 19(3-4):319–343, October 2007.

- [15] A Franklin and G.R. Faulhaber. Optimism invites deception. *The Quarterly Journal of Economics*, MIT Press, 103(2):397–407, 1988.
- [16] James J. Heckman, Hidehiko Ichimura, and Petra E. Todd. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):pp. 605–654, 1997.
- [17] J. Horner. Reputation and competition. *American Economic Review*, 92:644–663, 2002.
- [18] Pauline M. Ippolito and Alan D. Mathios. Information, advertising and health choices: A study of the cereal market. *The RAND Journal of Economics*, 21(3):pp. 459–480, 1990.
- [19] Ginger Zhe Jin and Phillip Leslie. The effect of information on product quality: Evidence from restaurant hygiene grade cards. *The Quarterly Journal of Economics*, 118(2):409–451, May 2003.
- [20] Ginger Zhe Jin and Alan T. Sorensen. Information and consumer choice: The value of publicized health plan ratings. *Journal of Health Economics*, 25(2):248 – 275, 2006.
- [21] Hailey Hayeon Joo. Social learning and optimal advertising in the motion picture industry. 2009 Meeting Papers 513, Society for Economic Dynamics, 2009.
- [22] Richard E. Kihlstrom and Michael H. Riordan. Advertising as a signal. *Journal of Political Economy*, 92(3):pp. 427–450, 1984.
- [23] R. E. Kranton. Competition and the incentive to produce high quality. *Economica*, 70:385–404, 2003.
- [24] Ryan Lampe. Post-release advertising, movie quality, and word-of-mouth. *mimeo*, May 2015.
- [25] Haiyan Liu. A structural model of advertising signaling and social learning: The case of motion picture industry. *mimeo*, 2015.
- [26] G. Mailath and L. Samuelson. Who wants a good reputation? *Review of Economic Studies*, 68(2):415–41, 2001.
- [27] Richard D. McKelvey and Thomas R. Palfrey. Quantal response equilibria for normal form games. In *Normal Form Games, Games and Economic Behavior*, pages 6–38, 1996.
- [28] Charles C. Moul. *A Concise Handbook of Movie Industry Economics*. Cambridge University Press, New York, NY, 2005.
- [29] Phillip Nelson. Advertising as information. *Journal of Political Economy*, 82(4):pp. 729–754, 1974.

- [30] Nicholas Powers, Allen Blackman, Thomas Lyon, and Urvashi Narain. Does disclosure reduce pollution? evidence from india's green rating project. *Environmental and Resource Economics*, 50:131–155, 2011. 10.1007/s10640-011-9465-y.
- [31] C SHAPIRO. Consumer information, product quality, and seller reputation. *Bell Journal of Economics*, 13:20–35, 1982.
- [32] Tiffany Shih. The effect of quality disclosure regulations on nearby firms.
- [33] J. Sutton. Sunk costs and market structure. *MIT Press.*, 1991.
- [34] Jennifer. Vanderhart and Steven N. Wiggins. Advertising, quality, and motion picture performance. 2001.