Title
Consumer Online Search and New-Product Marketing

Author
Kim, Ho

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Consumer Online Search and New-Product Marketing

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Management

by

Ho Kim

2013

ABSTRACT OF THE DISSERTATION

Consumer Online Search and New-Product Marketing

by

Ho Kim

Doctor of Philosophy in Management

University of California, Los Angeles, 2013

Professor Dominique M. Hanssens, Chair

This dissertation contains three essays that study the implications of online search activity for new-product marketing. Using the U.S. motion picture industry as a test case, the first essay examines the dynamic causal relationship between traditional media, consumers' media generation activity, media consumption activity, and market demand of movies. Consumers' media generation and consumption activities around movies are operationalized by the blog volume and search volume of those movies. I develop three separate models—a pre-launch period model, a post-launch period model, and an opening-week model—and examine the relationship between them separately for each period. As the focal variables are jointly determined, I introduce instruments and explain how to correct for endogeneity bias. I find that consumer searching activity is a key mediator between advertising, consumer blogging activity, and market demand.

The second essay examines the pre-launch advertising effectiveness of new products using online search indexes as the response variable. I model the relationship between the advertising schedule and the online search volume process during the pre-launch period of movies. The model incorporates consumers' willingness-to-search and the time-varying effectiveness of advertising as key elements that influence online search volume at specific times. The model is represented in a Bayesian dynamic linear model framework and applied to the U.S. movie industry. The empirical analysis reveals important features of consumers' pre-launch interest development for new products. First, consumers' pre-launch responses to advertising are substantially influenced by the timing of advertising. Second, advertising effectiveness varies over time as a function of past advertising outlay. Third, the time-to-launch effect and time-varying advertising effectiveness vary substantially across movies. The estimation results are used to suggest a more effective pre-launch advertising schedule.

The third essay provides an explanation of the varying predictive power of online search volume by examining the effects of consumers' quality perception on the search activity for and actual demand of new products. I hypothesize that both the perceived quality and quality uncertainty of a new product increase consumers' search activity for the new product, while only perceived quality positively influences the conversion of search activity into actual demand. Using the U.S. movie industry as a test case, I find that both perceived quality and quality uncertainty are positively associated with the pre-launch search volume of movies, whereas only perceived quality positively influences the conversion of pre-launch searches into opening-week revenue. Similar findings are maintained in the post-launch period analysis. These findings imply that systematic over-/under-prediction of market success may occur if managers use an online search index without considering the effects of quality and quality uncertainty of new products.

The dissertation of Ho Kim is approved.

Randolph E. Bucklin

Andrew Ainslie

Jorge Silva-Risso

Dominique M. Hanssens, Committee Chair

University of California, Los Angeles

2013

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Many people have helped me to complete this dissertation. I would like to thank my great adviser, Mike Hanssens, for his encouraging and insightful comments throughout my academic journey of Ph.D. work. This dissertation could have not been academically rigorous and managerially relevant without his guidance. Many thanks must go to my dissertation committee members, Randy Bucklin, Andrew Ainslie, and Jorge Silva-Risso for their time and helpful comments.

I am grateful to my Professors for their heartwarming support: Duk Bin Jun at KAIST; Shijin Yoo at Korea University; Moon-soo Kim at Hankuk University of Foreign Studies; and Myung-whan Park at Hansung University. Without their help, this journey could have not started.

I have been very fortunate to have Inigo and Rafael as my colleague students. Inigo's great sense of humor has uplifted my spirit. Rafael's academic camaraderie helped me to successfully finish the rigorous coursework of Marketing Ph.D. program at UCLA.

I cannot thank enough to my parents for trusting in me to start my academic journey. Without their unconditional and sacrificial love, my study at UCLA would have not been possible. I thank my wife, Sunyoung Park, whose love and prayers have comforted me. Her support has strengthened me for strenuous work of Ph.D. I am indebted to my younger brother, Taehoon, and his wife, Moonjoo, for their support. In my absence, they have done all the things that the oldest son of a Korean family is supposed to do.

I am blessed to have fellowship with many wonderful Christian brothers and sisters. I am thankful to Cheol-seung hyung and Chang-eon hyungnim for their heartfelt love and guidance to

Christian faith. I was also blessed to know many wonderful Sunday school teachers of Jangchoong Presbyterian Church. It was very meaningful and joyful to serve God and His children in the Sunday school program. I am thankful to Sungho hyung at ETRI, who walked with me while I was going through the most difficult time of my life—so far. It was a God's blessing that I could meet Pastors Woosoon Choe and Soomi Kim of the United Campus Church of Southern California. Their continuing prayers and fellowship have deeply strengthened my faith. They have shown me the example of Christian life that glorifies God. I would like to thank the choir members, the fellowship committee members, and the Wednesday Bible study group of Sage Granada Park United Methodist Church for their prayers and support. Singing in the choir to worship God has brought much joy to my heart. Sage Granada Park UMC is a wonderful congregation that is full of love and kindness. I pray that the congregation glorifies God through their ministries in the San Gabriel Valley area and beyond.

But above all, I thank God. As the ancient Psalmist confessed, God has been my rock, my fortress and my shield—and I know he will be forever. Because of His unfathomable grace and unfailing love, I could have started and finished my Ph.D. work. Therefore, it is not me but God who should be honored and glorified by this achievement. Blessed be the Lord. And, may God's blessings be upon you who has read this.

VITA

1997	B.B.A., Business Administration Yonsei University Seoul, South Korea
1999	M.E., Management Engineering Korea Advanced Institute of Science and Technology Seoul, South Korea
1999 – 2001	Assistant Manager Korea Telecom Infotech Seoul, South Korea
2001 – 2006	Researcher and Consultant Electronics and Telecommunications Research Institute Daejeon, South Korea
2006 – 2008	Graduate Student Researcher Korea Advanced Institute of Science and Technology Seoul, South Korea
2008 – 2013	Teaching and Research Assistant Anderson School of Management University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Moon-soo Kim and Ho Kim (2007), "Is There Early Take-off Phenomenon in Diffusion of IP-Based Telecommunications Services?", *Omega*, 35 (6), 727-739.

Moon-soo Kim and Ho Kim (2004), "Innovation Diffusion of Telecommunications: General Patterns, Diffusion Clusters and Differences by Technological Attributes", *International Journal of Innovation Management*, 8 (2), 223-241.

Ho Kim and Dominique M. Hanssens, "Pre-Launch Advertising Effectiveness of New Products: An Empirical Analysis Using Online Search Volume", paper to be presented at the 2013 Marketing Dynamics Conference, Chapel Hill, North Carolina, May 2013

Ho Kim and Dominique M. Hanssens, “The Dynamics of Commercial Media, Social Media, and Market Outcomes”, paper presented at the 2012 INFORMS Marketing Science Conference, Boston, Massachusetts, June 2012

Ho Kim and Dominique M. Hanssens, “Priming vs. Wearout: Early Prelaunch Advertising, Online Buzz, and New-Product Sales”, paper presented at the 2011 INFORMS Marketing Science Conference, Houston, Texas, June 2011

Ho Kim and Dominique M. Hanssens, “Priming vs. Wearout: Early Prelaunch Advertising, Online Buzz, and New-Product Sales”, paper presented at the 2011 Annual University of Houston Doctoral Symposium, Houston, Texas, April 2011

1. Introduction

1.1. Overview of the Dissertation

This dissertation contains three essays that study the implications of online search activity for new-product marketing. Consumer online search activity is important for marketers for at least three reasons. First, search activity is a media consumption activity, not a media generation activity. Second, collective search intensity for a product may represent consumers' collective interest in the product. Third, search activity is universal online behavior of consumers, as 91% of U.S. adult Internet users use search engines in their everyday lives (The Pew Research Center, 2012). Despite its importance, marketing research to date has not paid enough attention to the implications of consumer online search activity. Perhaps the difficulty in data collection is a reason for the lack of research.

Google Trends, a public service by Google, has changed this. It provides a weekly search index of keyword queries entered into the Google search engine from 2004 to the current time. This tool has enabled marketing researchers to monitor how consumer interest changes on a weekly basis for specific products. The three essays in this dissertation use Google search indices as the main data source and investigate various implications of collective online search activity for new-product marketing. Specifically, the essays attempt to answer the following questions.

- Essay 1: What is the causal relationship between traditional media, consumers' media generation activity, their media consumption activity, and the market outcome of a new product?
- Essay2: How is the advertising effectiveness of a new product determined in its pre-launch period?

- Essay3: When is online search volume a good market predictor and when is it not? How does consumers' quality perception influence their search behavior and actual demand?

To answer the questions, I use empirical approaches. I gather observational data from the U.S. movie industry. For each movie in the data set, I collect information regarding weekly search volume, weekly advertising spending, weekly blog volume, weekly revenue, and various movie characteristics. Then I develop appropriate empirical models and apply them to the data set to answer the research questions.

The first essay examines the causal relationship between advertising, consumer media generation activity, media consumption activity (i.e., consumer search activity), and product consumption. I develop three separate models—a pre-launch period model, a post-launch period model, and an opening-week model—and answer the research question separately for each period. As the focal activities are jointly determined, I suggest variables that are related to exogenous variations in the media generation and media consumption activities and explain how to identify the causal relationship between the jointly determined variables.

Several interesting findings emerge. First and foremost, I find that consumer search activity plays a pivotal role throughout a movie's life cycle. Pre-launch blog volume does not explain the variation in opening-week revenue once pre-launch search volume is controlled for. After opening week, weekly search volume is the key mediator between weekly advertising, blog volume, and revenue: weekly search volume fully mediates the relationship between blog volume and revenue. Second, I find that advertising is the main driver of movie revenue throughout a movie's life cycle. The pre-launch search volume is influenced by pre-launch advertising, but not by pre-launch blog volume. After opening week, the effectiveness of advertising on movie revenue is almost twice as large as that of blog volume.

The second essay examines the pre-launch advertising effectiveness of new products. To this end, I model the relationship between advertising schedule and the online search volume process of new products during their pre-launch periods. I propose consumers’ changing willingness-to-search and time-varying effectiveness of advertising as two critical elements that influence online search volume at a specific pre-launch time. I develop a state-space model that incorporates these two elements. The dependent variable is the weekly online search volume of a movie and the key covariate is the movie’s weekly advertising spending. I estimate the model in the Bayesian dynamic linear model framework.

The empirical analysis reveals important features of consumers’ pre-launch interest development for new products. First, consumers’ pre-launch response to advertising—in the form of keyword search—is substantially influenced by the timing when the advertising is conducted. Pre-launch advertising more efficiently triggers consumer search in the weeks that immediately precede the release time than in the weeks that are far in advance of release. Second, advertising effectiveness is not constant but varies over time. It is influenced by features of the advertising schedule such as the passage of time since the first advertising, the amount of the previous period’s advertising, and the existence of an ad hiatus period. Third, the time-to-launch effect and time-varying advertising effectiveness vary substantially across movies. This heterogeneity is explained by movie characteristics such as director power and whether the move is a sequel. The estimation results are used to suggest a more effective pre-launch advertising schedule.

The third essay aims to explain when collective online activity of consumers is a good predictor for new-product demand versus when it is not. While early studies in this area report that collective online activity of consumers provides excellent predictive performance for new-product demand (Goel et. al., 2010; Kulkarni, Kannan, and Moe 2012; Mao, Counts, and Bollen 2011; Mestyan, Yasseri, and Kertesz 2012; Wu and Brynjolfsson 2009), recent studies report that the excellent predictive performance of collective online activity of users is limited to certain products (e.g., Wong, Sen, and Chiang 2012).

I focus on the role of consumers' quality perception of products—the perceived quality and the perceived uncertainty about the quality of products—in influencing their search activities and actual purchase. I hypothesize that both perceived quality and quality uncertainty of a product increase consumers' search activity for the product, while only perceived quality—and not perceived uncertainty about quality—positively influences the conversion from search to actual demand.

Through an empirical analysis involving theatrically released movies, I find supporting evidence for the hypotheses. First, when it comes to the pre-launch search volume and opening-week revenue of movies, I find that 1) both perceived quality and quality uncertainty about a movie increase the pre-launch search volume of the movie; 2) only the perceived quality positively moderates the conversion of pre-launch search volume into opening-week revenue. I extend the analysis to the post-launch period and find the same results by use of different operationalizations of quality and quality uncertainty. Our findings imply that managers should interpret with caution the market demand predicted with online search volume (e.g., Kulkarni, Kannan, and Moe 2012).

In sum, my dissertation examines managerial implications of collective consumer search activity for new-product marketing. The first essay aims to understand the big picture. It examines how consumer search activity is related to other important variables such as advertising, consumer generated media, and sales. The second essay examines how advertising effectiveness is determined in the pre-launch period of a new product. I model the relationship between advertising and collective search activity of consumers in pre-launch period of movies. The third essay aims to provide an explanation for the varying predictive performance of online search volume. By focusing on the role of perceived quality and quality uncertainty in generating search volume and actual demand, I explain that online search volume can systematically over-/under-predict market outcome.

While the three essays require a search volume metric that is comparable across products, the search index from Google Trends is comparable only across time for a given product. To compare search volume across products, I propose a method that transforms the weekly Google search index to a metric that is comparable across products as well as across time. With the proposed method in chapter 5, researchers will be able to construct data sets that contain cross-sectionally comparable search volume metrics for their own analyses.

I believe that collective online search activity of consumers can provide valuable business intelligence for new-product marketing. My dissertation has aimed to find such implications by applying a few ideas to the U.S. movie industry. I hope my dissertation adds new insights to the marketing area.

1.2. References

Goel, Sharad, Jake M. Hofman, Sebastien Lahaie, David M. Pennock, and Duncan J. Watts (2010), "Predicting Consumer Behavior with Web Search," Proceedings of the National Academy of Sciences of the United States of America (PNAS), 107 (41), 17486-90.

Kulkarni, Gauri, P. K. Kannan, and Wendy Moe (2012), "Using Online Search Data to Forecast New Product Sales," *Decision Support Systems*, 52 (3), 604-11.

Mao, Huina, Scott Counts, and Johan Bollen (2011), "Predicting Financial Markets: Comparing Survey, News, Twitter, and Search Engine Data," arXiv:1112.1051.

Mestyan, Marton, Taha Yasseri, and Janos Kertesz (2012), "Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data," arXiv:1211.0970v1.

The Pew Research Center (2012), *Pew Internet & American Life Project Tracking Surveys*, available at http://pewinternet.org/Trend-Data-(Adults)/Online-Activites-Total.aspx.

Wong, Felix Ming Fai, Soumya Sen, and Mung Chiang (2012), "Why Watching Movie Tweets will not Tell The Whole Story?," arXiv:1203.4642.

Wu, Lynn and Erik Brynjolfsson (2009), "The Future of the Prediction: How Google Searches Foreshadow Housing Prices and Sales," SSRN Working paper.

2. The Dynamics of Consumers' Media Generation, Media Consumption, and New-Product Demand

2.1. Introduction

Consumer online activity around new products generally appears in two distinct forms: media generation (e.g., review posting, rating, and blogging) and media consumption (e.g., searching, reading, and watching). While media generation and consumption may have different effects on market outcomes, research to date has mainly focused on consumers' media generation, leaving the causes and effects of media consumption less studied. The objective of this study is to examine the dynamics between advertising, consumers' media generation, media consumption, and market outcomes around new products.

The motion picture industry is an ideal test case. First, hundreds of new movies are produced each year, enabling researchers to collect a large sample of new movies with diverse characteristics. Second, as a movie is an experiential good, word-of-mouth plays a critical role in making movie-going decisions (Chintagunta, Gopinath, and Venkataraman 2010; Liu 2006). Accordingly, both media generation and media consumption are common in the movie industry. Third, compared to other products, it is relatively easy to measure consumers' media generation and media consumption activities about movies. For media generation activities, there are many online forums, blogs, and review sites from which researches can collect word-of-mouth about movies. For media consumption activities, researchers may use keyword search indices of movies. In our study, as the measure of consumer media generation activity about a movie, we use the number of blog postings about the movie. Blogs are preferred over user reviews for our research questions because blog posts about a movie are written even before the movie is launched. That is, blog postings about a movie are generally available long before the movie is released, enabling researchers to examine the pre-launch relationship between the focal variables. As the measure of media consumption activity about a movie, we use the Google keyword search index.

To answer the research questions, we develop three separate models: a pre-launch period model, a post-launch period model, and an opening-week model. The pre-launch period model is a panel data model that examines the relationship between advertising, blog volume, and online search volume over the pre-launch period. The post-launch period model is a panel data model that examines the relationship between advertising, blog volume, online search volume, and movie revenue over the post-launch period. The opening-week model is a cross-sectional data model that examines the effects of pre-launch blog and search volume on the opening-week revenue of movies. In all models, it is challenging to find causal relationship between the focal variables as they are simultaneously determined. We find variables that are related to exogenous variations in the blogging and search activities and explain how to identify the causal effects.

To the best of our knowledge, this is the first study that examines the relationship between advertising, consumers' media generation, media consumption, and market outcomes. First, we find that consumer search activity plays a pivotal role throughout a movie's life cycle. Pre-launch blog volume does not explain the variation in opening-week revenue if pre-launch search volume is controlled for. After opening week, search activity is the key mediator between advertising, blogging activity, and revenue. Especially, search volume fully mediates the relationship between blog volume and revenue. This means that blog postings do not influence revenue as long as they are not searched for and viewed. This also means that weekly blog volume does not explain variation in weekly revenue if weekly search volume is controlled for.

Second, we find that advertising is the main driver of movie revenue throughout a movie's life cycle. The collective pre-launch searching activity of consumers, which has a significant positive effect on opening-week movie revenue, is influenced by pre-launch advertising, but not by pre-launch blogging activity of consumers. After opening week, the effectiveness of advertising on movie revenue is almost twice as large as that of blog volume. Furthermore, we find that blog postings need to be searched for them to influence revenue, whereas the majority of advertising effect is realized without consumer online search. Third, blogging activity and searching activity have different antecedents and consequences across the pre-launch and post-launch periods. In the pre-launch period, blogging activity and searching activity do not interact with each other; in the post-launch period, the two interact with one another and influence revenue.

The remainder of the study is organized as follows. In the next section, we summarize previous research. We then describe our movie data set. Next, we develop three models to answer our research questions. We apply them to our movie data set and discuss the findings and managerial implications. We then formulate conclusions and areas for future research.

2.2. Relevant Research

To understand the contribution of this study, we briefly review previous studies that examine the antecedents and consequences of consumers' media generation and consumption activity.

2.2.1. Consumer Media Generation

Godes and Mayzlin (2004) use online conversations to measure word-of-mouth (WOM) communication and show that the dispersion of WOM is positively associated with ratings of new television (TV) shows. Chevalier and Mayzlin (2006) use online reviews and ratings at Amazon.com to study the causal relationship between online ratings and book sales. They find that an improvement in book ratings leads to an increase in relative sales.

In the motion picture sector, Liu (2006) finds that WOM volume offers significant explanatory power for both aggregate and weekly box-office revenue, especially in the early weeks after release. However, he does not find explanatory power from WOM valence. Similarly, Duan, Gu, and Whinston (2008a, 2008b) find that box-office sales are significantly influenced by the volume of reviews, but not the valence of reviews. Chintagunta, Gopinath, and Venkataraman (2010), on the contrary, find that valence captured by the average user rating explains designated market area (DMA)-level opening-day box-office revenues, while the volume and variance of ratings do not.

More recently, researchers have started to pay attention to blogs, a new form of online consumer-generated media (CGM). Dhar and Chang (2009) show that future sales of music albums are positively correlated with the volume of blog posts about the albums. Onishi and Manchanda (2012) find that advertising and blogging are synergistic, and cumulative blogs are predictive of market outcomes. Gopinath, Chintagunta, and Venkatraman (2011) find that the volume of blogs has significant effects on opening-week movie sales, but after release it is the valence of blogs that is predictive. Stephen and Galak (2012) examine how earned media—both traditional earned media (e.g., publicity and press mentions) and social earned media (e.g., blogs and online community posts)—affect sales in the microlending market place. They find that, because of greater frequency of social earned media activity, the elasticity of sales to social earned media is substantially greater than that to traditional earned media.

Other streams of research examine the effectiveness of social media versus commercial media in acquiring customers. Villanueva, Yoo, and Hanssens (2008) find that customers who are acquired through WOM add more long-term value to the firm than customers who are acquired through advertising. Trusov, Bucklin, and Pauwels (2009) find that WOM referrals have substantially longer carryover effects than traditional marketing actions in acquiring new members at an Internet social networking site.

When it comes to the antecedents of online CGM, previous literature finds that advertising, the volume of the previous period's CGM, and current and past market outcomes are positively correlated with the volume of the current period's online reviews and blog postings (Duan, Gu, and Whinston 2008a; Duan, Gu, and Whinston 2008b; Liu 2006; Onishi and Manchanda 2012).

In sum, previous research has examined diverse sources of online CGM ranging from newsgroup conversations, online reviews, ratings to blog postings. Various aspects of online CGM have been studied including volume, valence, and dispersion. The product categories that are examined are mainly experiential goods such as TV shows, books, movies, and music albums. Most previous studies have examined sales or revenue as the consequence of online CGM, while some studies have focused on metrics such as customer acquisition and the long-term value of consumers acquired through WOM.

2.2.2. Media Consumption

Contrary to media generation activity, media consumption activity is difficult to measure in the real world. Perhaps for this reason, previous studies on consumers' media consumption predominantly focus on the antecedents of the activity by use of laboratory experiments or surveys (Beatty and Smith 1987; Moorthy, Ratchford, and Talukdar 1997; Punj and Staelin 1983; Srinivasan and Ratchford 1991).

It is rather recently that researchers have begun to examine the relationship between consumers' media consumption and business metrics such as advertising and sales. Kulkarni, Kannan, and Moe (2012) find significant improvement in predicting opening-week box-office movie revenue by including pre-launch online search volume. Examining the relationship between TV advertising and keyword search in the financial services market, Joo, Wilbur, and Zhu (2012) find that TV advertising influences consumers' search activity for branded keywords, but not their search activity for generic keywords. Yang et. al. (2012) develop an individual consumer-level model to examine consumer WOM generation and consumption decision and apply it to survey data collected from the automobile industry. They find that consumers' propensity to generate WOM is positively correlated with their media exposure whereas their propensity of consuming WOM is mixed. Kim (2013) examines how pre-launch advertising influences consumers' pre-launch search activity.

To summarize, prior empirical studies to date have mainly focused on consumers' media generation activities; it is only recently that researchers have begun to pay attention to media consumption activity in real-world business environments. Furthermore, those recent studies do not consider the mutual relationship between media generation and consumption. As a result, the managerial implications of media consumption are not well appreciated. By examining the dynamic relationship between advertising, blog volume, online search volume, and revenue in the U.S. motion picture industry, we aim to provide new insights for new-product managers. Table 2.1 summarizes the relevant previous research mentioned above to help understand the contribution of this study.

Table 2.1: Comparison of Relevant Studies

Study	Variables Examined			
	Traditional Media	Consumer-Generated Media	Media Consumption (Search)	Market Outcome
Chevalier and Mayzlin (2006) Dhar and Chang (2009) Duan, Gu and Whinston (2008a) Duan, Gu and Whinston (2008b) Godes and Mayzlin (2004) Liu (2006) Stephen and Galak (2012)		✓		✓
Chintagunta, Gopinath, and Venkataraman (2010) Gopinath, Chintagunta, and Venkataraman (2011) Onishi and Manchanda (2012) Trusov, Bucklin, and Pauwels (2009) Villanueva, Yoo, and Hanssens (2008)	✓	✓		✓
Beatty and Smith (1987) Punj and Staelin (1983) Srinivasan and Ratchford (1991) Moorthy, Ratchford, and Talukdar (1997)			✓ (drivers of search behavior)	
Joo, Wilbur, and Zhu (2012) Kim (2013)	✓		✓	
Kulkarni, Kannan, and Moe (2012)	✓		✓	✓
Yang et. al. (2012)	✓	✓	✓	
This study	✓	✓	✓	✓

2.3. The Data

Our database consists of 153 movies, most of which were released in 2009. For each of the 153 movies, we collect weekly advertising spending, weekly blog volume, weekly search volume, and weekly box-office revenue from 30 weeks before its release through 10 weeks after the release. To use as instruments in estimating our models, we also collect various movie characteristics, the weekly Google search index of the keyword “opening movie,” and weekly traffic to the five most popular blog sites—i.e., blogger.com, tumblr.com, wordpress.com, squarespace.com, and posterous.com.

2.3.1. Advertising and Box Office Revenue

Our advertising data covers the major media outlets such as television, print, radio, and outdoor expenditure as collected by Nielsen. The average advertising spending of the 153 movies is $21 million with 80% of the advertising budget spent in the pre-launch periods or during the release week. Box-office revenue is collected from The Numbers (www.the-numbers.com). One hundred and forty-two movies were exhibited for at least 10 weeks, and the shortest movie run was five weeks. The median U.S. box-office revenue is $44 million. Figure 2.1(a) shows weekly advertising spending and box-office revenue averaged across the 153 movies.

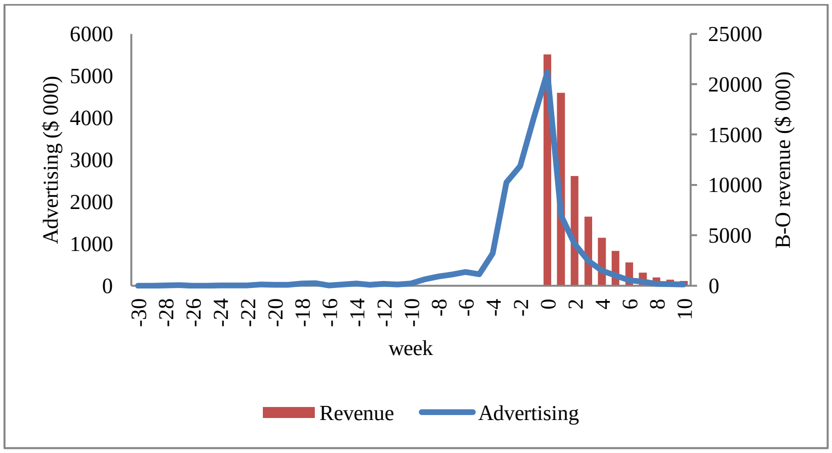
2.3.2. Blog Postings and Search Volume

Weekly blog postings for each movie were collected from the Google blog search engine (www.google.com/blogsearch). To minimize noise in the data collection process, we constrained our search for blogs whose titles contain at least one of the following words: *movies* or *films*. Our general rule of constructing search keywords for blog postings is <movie title> + “movie”.[1] For example, to find blog postings of the movie *Avatar*, we used the keyword “Avatar movie.” For movies with long titles, reduced search keywords were used. For instance, to search for blog postings of the movie *Bad Lieutenant: Port of Call New Orleans*, we searched for the postings that contain the keyword “Bad Lieutenant” in their title. For each week of each movie, we repeated the search practice five times and used the mode or median of the number of blog postings so gathered.

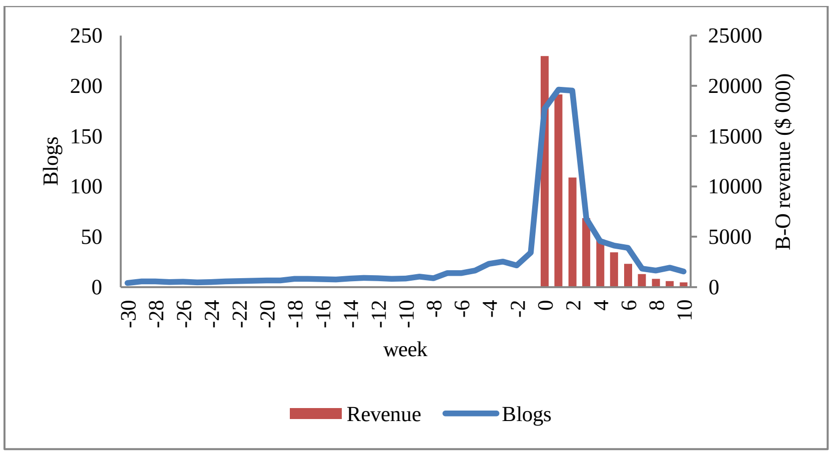
For weekly search volume of movies, we relied on Google Trends. Google Trends shows the weekly search index of the entered keyword. The raw index provided by Google is normalized to conceal the actual search volume of the keyword entered into the Google search engine. This normalization creates a difficulty for analyzing the effect of search volume *across* movies. We propose a method that transforms the raw search index of Google to a cross-sectionally comparable search volume measure. The detailed methodology of collecting weekly Google search indices of movie keywords and transforming them into cross-sectionally comparable measures can be found in chapter 5, the appendix to the dissertation. Figure 2.1(b) and 2.1(c) show how the average blog volume, search volume, and revenue change on a weekly basis over the movie lifecycle.

[1] We developed several different versions of search keywords and found that the keyword <movie title> + “movie” is the most appropriate in the sense that it produced the highest fit with the movie’s observed advertising and launch schedules.

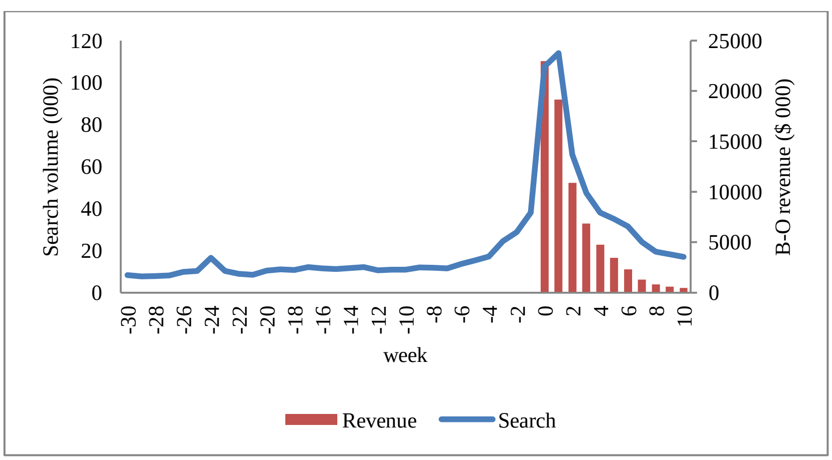
Figure 2.1: Weekly Trends of Advertising, Blog Volume, Search Volume and Box Office Revenue



(a) Average Weekly Advertising Spending and Box Office Revenue (N=153)



(b) Average Weekly Blog Volume and Box office Revenue (N=153)



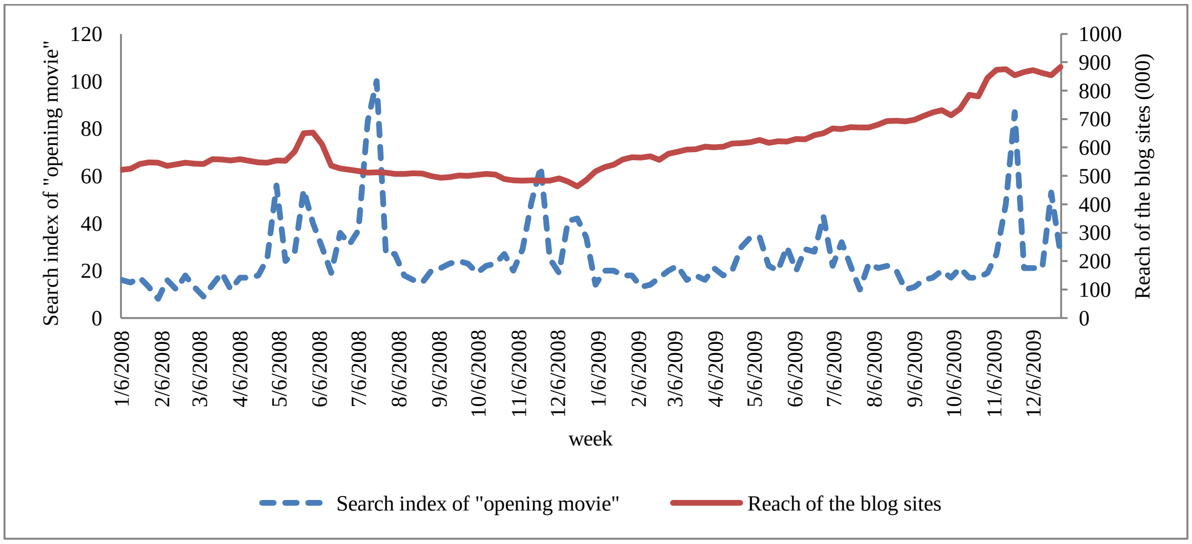
(c) Average Weekly Search Volume and Box office Revenue (N=153)

2.3.3. Other Variables

We collect the weekly Google search index of the keyword "opening movie," weekly traffic to the five most popular blog sites, and various movie characteristics. They are used as instruments in estimating our models, which will be introduced in next sections.

The weekly search index of the keyword "opening movie" was collected from Google Trends. Using the search filters of Google Trends, we consider only the search activities made in the U.S. and in the movie industry. The weekly traffic of the five most popular blog sites (i.e., blogger.com, tumblr.com, wordpress.com, squarespace.com, and posterous.com) was collected from Alexa.com, a subsidiary of Amazon.com. Alexa collects daily traffic information of websites based on a global panel of toolbar users. The panel consists of millions of people using toolbars created by over 25,000 publishers, including Alexa and Amazon. We aggregate the daily reach of the five blog sites at the weekly level to create the weekly total reach of the blog sites. Figure 2.2 exhibits the weekly search index of the keyword "opening movie" and the weekly reach of the five blog sites, from the first week of 2008 to the last week of 2009.

Figure 2.2: Weekly Search Index of the Keyword “Opening Movie” and Weekly Traffic to the Five Blog Sites



We also collect various movie characteristics. They include genre, MPAA rating, monthly seasonality, whether the movie is a sequel, average critic rating, and director power. For the director power, we collect three variables: total revenue of past movies with which the focal movie’s director was involved as either a director, writer, or producer, since 1990 up to one calendar year before the focal movie’s release; average user rating of such movies; and the standard deviation of user ratings of such movies. Descriptive statistics on the movie characteristics can be found in section 4.3. Table 2.2 summarizes the variables and their sources.

Table 2.2: Variables and Data Sources

Category	Variable	Source of Data
Marketing activities	Weekly advertising spending	Nielsen
	Weekly number of screens	The numbers
Focal endogenous variables	Weekly blog postings	Google blog search engine
	Weekly search volume	Google Trends
	Weekly revenue	The-numbers
Movie Characteristics	Genre, MPAA rating, Sequel	IMDb
	Average critic rating [range: 1 – 100]	Metacritic
	Director power variables	IMDb
	Monthly Seasonality: January-April; May-August; September-October, November-December	Einav (2007) Ho, Dhar, Weinberg (2009)
Holiday	National holiday	
Others	Weekly Google search index of the keyword “opening movie”	Google
	Daily reach of the five popular blog sites	Alexa.com

2.4. The Model

Our model consists of three sub-models: the pre-launch period model, the post-launch period model, and the opening-week model. The pre-launch period model examines the causal relationship between advertising, blog volume, and search volume. As such, the pre-launch period model consists of two equations whose dependent variables are weekly blog volume (blog equation) and weekly search volume (search equation). The post-launch period model consists of three equations whose dependent variables are weekly blog volume (blog equation), weekly search volume (search equation), and weekly revenue (revenue equation). The opening-week model examines the effect of pre-launch search volume and pre-launch blog volume on opening-week revenue after controlling for the effects of other relevant variables. As such, the opening-week model is a cross-sectional data model.

Let us first discuss the panel data models. Our specifications for the pre-launch and post-launch period models are based on the following key considerations:

- First, we treat weekly advertising spending as an exogenous variable, following Elberse and Anand (2007); Eliashberg, Elberse, and Leenders (2006); and Onishi and Manchanda (2012).
- Second, covariates of each equation consist of weekly advertising and the weekly endogenous variables that are not the dependent variable of the focal equation. For example, the covariates of the search equation of the post-launch period model consist of weekly advertising, blog volume, and revenue. Determining the lag length of each covariate is discussed later.
- Third, the lagged dependent variables of a focal equation are not included as covariates of the equation. For instance, the lagged search volume is not included on the right-hand-side (RHS) of the search equation. Instead, we include sufficiently many lags of advertising and other endogenous variables. This is a common modeling approach adopted in the representative studies on the movie market (Basuroy, Desai, and Talukdar 2006; Elberse and Eliashberg 2003; Liu 2006).
- Fourth, on the RHSs of each of blog and search equations, we include a variable that is associated with exogenous variations to the dependent variable of the equation. In the search equation, we include the weekly Google search index of the keyword "opening movie"; in the blog equation, we include the weekly reach of the five most popular blog sites.
- Fifth, we include the time-specific effect but not the individual movie-specific effect in each equation. In the panel data model, it is usually recommended to control for unobserved individual or time effects. However, when the lagged endogenous variables are included on the RHS, controlling for individual-specific effects leads to inconsistent estimators (Baltagi 1995). An alternative is to include only the common intercept in each equation. This approach is

supported if cross-sections (i.e., each movie in our case) do not have significantly different individual fixed effects after controlling for the covariates. We conducted Holtz-Eakin (1988) test to examine if there exist significantly different individual fixed effects. The test results showed that there is no supporting evidence for significant individual fixed effects. As such, we include only the common intercept in each equation.

- Sixth, based on unit-root test results, we first-difference the pre-launch period model but not the post-launch period model.

- Seventh, we need to determine the lag lengths of the explanatory variables. There is a trade-off in determining the lag lengths of covariates. More lags are preferred to mitigate the effects of potential model misspecification (Enders 2004). On the other hand, including many lags reduces the degrees of freedom of the model. Also, if the time-series variables in the data set are highly serially correlated, including lagged values can cause the multicollinearity problem. Table 2.3 shows the serial correlation coefficients of variables in the pre- (equations (2-1) and (2-2)) and post-launch (equations (2-3) – (2-5)) models.

Table 2.3: Serial Correlation Coefficients of the Main Variables

(a) The Variables in the Pre-Launch Period Model

	Δ(Advertising)	Δ(Search)	Δ(Blog)
Serial Correlation Coefficient	-0.181	-0.234	-0.235

Δ represents first-differencing.

(b) The Variables in the Post-Launch Period Model

	Advertising	Blogs	Search	Revenue
Serial Correlation Coefficient	0.707	0.505	0.903	0.895

The serial correlation coefficients in Table 2.3 suggest that including lagged explanatory variables in the post-launch period model can cause a serious multicollinearity problem. As such, we include only the contemporaneous explanatory variables in the post-launch period model. The serial correlations of the variables of the pre-launch period models are weak. Therefore, in the pre-launch period model, we include as many as four lags of the explanatory variables to mitigate the misspecification problem that might result from including insufficient number of lags.

- Finally, movie characteristics are not included on the RHSs of the equations. They are reserved for instruments (Elberse and Eliashberg 2003).

With the above considerations, the pre- and post-launch period models are developed as follows.

2.4.1. The Pre-Launch Period Model

The pre-launch period model is specified by equations (2-1) and (2-2).

$$\text{(2-1)}\qquad \begin{aligned}\Delta\ln(\mathrm{Blog}_{it}) &= \sum_{k=0}^{4}\alpha_k^A\Delta\ln(\mathrm{Ad}_{it-k}) + \sum_{k=0}^{4}\alpha_k^S\Delta\ln(\mathrm{Search}_{it-k})\\ &\quad +\alpha^B\Delta\ln(\text{Traffic to blog sites}) + \alpha^c\Delta c_{it} + \Delta\delta_t^B + \Delta u_{it}^B\end{aligned}$$

$$\text{(2-2)}\qquad \begin{aligned}\Delta\ln(\mathrm{Search}_{it}) &= \sum_{k=0}^{4}\beta_k^A\Delta\ln(\mathrm{Ad}_{it-k}) + \sum_{k=0}^{4}\beta_k^B\Delta\ln(\mathrm{Blog}_{it-k})\\ &\quad +\beta^S\Delta\ln(\text{Volume of keyword 'opening movie'}) + \beta^c\Delta c_{it} + \Delta\delta_t^S + \Delta u_{it}^S,\end{aligned}$$

where Δ represents first-differencing. Blog_{it} is the blog volume, Ad_{it} is the advertising spending, and Search_{it} is the search volume of movie i in pre-launch week t. The column vector c_{it} is the set of control variables that might influence the blog volume and search volume of movie i in week t.

In our analysis, c_{it} consists of the holiday dummy variable. δ_t^B and δ_t^S are the week-specific effects of pre-launch week t. We assume that the errors are uncorrelated across equations: $\mathrm{Cov}(u_{it}^B, u_{it}^S) = 0$ for any i and t.

2.4.2. The Post-Launch Period Model

The post-launch period model is specified by equations (2-3) – (2-5).

$$\text{(2-3)} \qquad \begin{aligned} \ln(\mathrm{Blog}_{it}) &= \alpha + \alpha^A \ln(\mathrm{Ad}_{it}) + \alpha^S \ln(\mathrm{Search}_{it}) + \alpha^R \ln(\mathrm{Revenue}_{it}) \\ &\quad + \alpha^B \ln(\text{Traffic to blog sites}) + \alpha^c c_{it} + \delta_t^B + u_{it}^B \end{aligned}$$

$$\text{(2-4)} \qquad \begin{aligned} \ln(\mathrm{Search}_{it}) &= \beta + \beta^A \ln(\mathrm{Ad}_{it}) + \beta^B \ln(\mathrm{Blog}_{it}) + \beta^R \ln(\mathrm{Revenue}_{it}) \\ &\quad + \beta^S \ln(\text{Volume of keyword 'opening movie'}) + \beta^c c_{it} + \delta_t^S + u_{it}^S \end{aligned}$$

$$\text{(2-5)} \qquad \begin{aligned} \ln(\mathrm{Revenue}_{it}) &= \gamma + \gamma^A \ln(\mathrm{Ad}_{it}) + \gamma^B \ln(\mathrm{Blog}_{it}) + \gamma^S \ln(\mathrm{Search}_{it}) \\ &\quad + \gamma^D \ln(\mathrm{Scrns}_{it}) + \gamma^c c_{it} + \delta_t^R + u_{it}^R \end{aligned}$$

Blog_{it} is the blog volume, Ad_{it} is the advertising spending, Search_{it} is the search volume, and $\mathrm{Revenue}_{it}$ is the weekly revenue of movie i in post-launch week t. The vector c_{it} is the variables other than our focal variables that might influence the weekly blog volume, search volume, and revenue. c_{it} consists of the holiday dummy variable. δ_t^B, δ_t^S, and δ_t^R are the week-specific effects of post-launch week t. In the blog and search equations, we include weekly revenue to examine how movie consumption influences consumers' blogging (blog equation) and searching (search equation) activity. As the number of screens is a critical determinant of movie revenue, we control for the screen effect in the revenue equation. We assume that the errors are uncorrelated across equations: $\mathrm{Cov}(u_{it}^B, u_{it}^S) = \mathrm{Cov}(u_{it}^B, u_{it}^R) = \mathrm{Cov}(u_{it}^S, u_{it}^R) = 0$ for any i and t.

2.4.3. The Opening-Week Model

The objective of the opening-week model is to examine the effect of advertising and the pre-launch blog and search volume on the opening-week revenue of movies. It is a cross-section model as only one data point is observed per movie in the opening week. The model is specified as follows.

$$\begin{aligned}\ln(\text{Open\_Revenue}_i) = {} & \phi_0 + \phi_1 \ln(\text{Open\_Ad}_i) + \phi_2 \ln(\text{Open\_Scrns}_i) \\ & + \phi_3 \ln(\text{PreLaunch\_Blog}_i) + \phi_4 \ln(\text{PreLaunch\_Search}_i) \\ & + \phi_5 (\text{Open\_Blog}_i) + \phi_6 (\text{Open\_Search}_i) \\ & + \phi_7 (\text{Holiday}_i) + \phi_8 (\text{Critic\_Review}_i) + u_i^R,\end{aligned} \tag{2-6}$$

where Open_Revenue_i is the opening-week revenue of movie i, Open_Ad_i is the opening-week advertising spending, Open_Scrns_i is the number of opening-week screens, PreLaunch_Blog_i is the blog volume cumulated during the pre-launch period of movie i, $\text{PreLaunch\_Search}_i$ is the search volume cumulated during the pre-launch period of movie i, Open_Blog_i is the opening-week blog volume, Open_Search_i is the opening-week search volume, Holiday_i is the indicator whether movie i’s opening week contains any national holidays, and Critic_Review_i is the average critic rating of movie i as collected from Metacritic.com. Other movie characteristics are reserved for instruments.

2.4.4. Endogeneity and Identification of the Pre- and Post-Launch Period Models

The dependent variables in our panel data models are jointly determined. We rely on covariance restrictions and exclusion restrictions to identify the causal relationship between the jointly determined variables. First, the covariance restrictions use the assumption that the error terms are uncorrelated across equations for each movie and week. They are used to identify the contemporaneous causal effect of revenue on blogging and search activities. Second, the exclusion restrictions are used to identify the effect of blogging and search activities. The exclusion restrictions are achieved if we find exogenous variables that influence only the focal endogenous variable but not the others. For example, in order to identify the causal effect of weekly search volume on weekly revenue (the parameters γ_k^S in equation (2-5)), we need to find at least one exogenous variable that is associated with weekly search volume, but not weekly blog volume and weekly revenue. The exogenous variables should be included on the RHS of the equation of the corresponding endogenous variable. Also, the exogenous variables should be used as instruments in the equations where the corresponding endogenous variable is included as a covariate. In the following subsections, we explain how to identify the model.

Exogenous Variation in the Weekly Blog Volume of a Movie. We propose that the overall weekly blogging activity of the Internet user population—i.e., the blogging intensity across all topics and items—is an exogenous source of variation in the weekly blog volume of individual movies, but not in the weekly search volume and weekly revenue of the movies. The intuition is that the overall blogging activity in a week, which is triggered by various exogenous factors, may be positively related to consumers' blogging activity about specific topics—including individual movies—in that week. If this is the case, the overall blogging activity of Internet users in a week can provide a source of exogenous variation for weekly blog volume about any blogging topic, including the individual movies in our data set. On the other hand, as individual movies' advertising spending will hardly contribute to the overall blogging activity of the entire Internet user population, the weekly advertising spending of individual movies should not be correlated with the weekly overall level of blogging activity of the Internet user population. Also, there will be little, if any, connection between the weekly overall blogging activity of the Internet user population and movie consumers' interest and consumption decisions for specific movies.

Following the above argument, we include the weekly overall blogging activity of the Internet user population in the blog equation and exclude it from the search and revenue equations. Assuming that the weekly reach of popular blog sites represents the overall weekly blogging activity of the Internet user population, we include the weekly reach of the five most popular blog sites in the U.S. (blogger.com, tumblr.com, wordpress.com, squarespace.com, and posterous.com) as a covariate of the blog equation. In the search and revenue equations, we use it as an instrumental variable for the weekly blog volume variable.

Exogenous Variation in the Weekly Search Volume of a Movie. For a source of exogenous variation in the weekly search volume of individual movies, we propose the weekly Google search index of the keyword "opening movie." The search index of the keyword "opening movie" reflects consumers' interests in recently released or soon-to-be released movies. As consumers' generic interests in opening movies may spill over to individual movies that have been released recently or will be released soon, the weekly search index of the keyword "opening movie" should be associated with the search volumes of such individual movies. For example, consumers may first search with the keyword "opening movies" and then narrow down to the movies that are available now or in the near future (Rutz and Bucklin 2011). In contrast, as the search volume of the keyword "opening movie" represents the searchers' generic interest in theatrically released movies without a specific movie under consideration, the advertising spending of an individual movie should not contribute to the search volume of the keyword "opening movie." In the same line of reasoning, there will be no significant association between the weekly blog volume of individual movies and the weekly Google search index of the keyword “opening movie.”

Based on the above argument, we include the weekly search index of the keyword “opening movie” in the search equation and exclude it from the blog and revenue equations. In the blog and revenue equations, we use it as an instrumental variable for the weekly search volume of individual movies.

Identifying the Effects of Revenue on Blog Volume and Search Volume. Unexpected shocks in the supply side (e.g., an unexpected increase or decrease of the weekly number of screens from the consumer perspective) can create an exogenous variation to the weekly revenue of individual movies, but not to the weekly blog volume and search volume of the movies. However, such unexpected shock in the supply side is not readily observable to researchers. To identify the causal effect of the weekly revenue of a movie on its weekly blog volume and search volume, we rely on the assumption that the error term of the revenue equation (2-5) is contemporaneously uncorrelated with the error terms of the blog and search equations (2-3) and (2-4). That is, $\mathrm{Cov}(u_{it}^{R}, u_{it}^{B}) = \mathrm{Cov}(u_{it}^{R}, u_{it}^{S}) = 0$. This assumption is justified if unobserved factors in a post-launch week do not influence the weekly search volume, blog volume, and revenue simultaneously in that week. For example, suppose that in the weeks containing national holidays, more consumers search, blog, and watch movies. In this situation, if the model does not include the holiday dummy variable, then the assumption $\mathrm{Cov}(u_{it}^{R}, u_{it}^{B}) = \mathrm{Cov}(u_{it}^{R}, u_{it}^{S}) = 0$ may not hold. To mitigate any simultaneous effects of unobserved factors on search, blog, and revenue, the model includes the holiday dummy variable as well as the week dummy variables in each equation.

Under the assumption of $\mathrm{Cov}(u_{it}^{R}, u_{it}^{B}) = \mathrm{Cov}(u_{it}^{R}, u_{it}^{S}) = 0$, we can use the residuals from (2-5), $\hat{u}_{it}^{R}$ as an instrument for the log of weekly revenue, ln(Revenue$_{it}$) in (2-3) and (2-4). The intuition is that if the parameters in (2-5) are known, u_{it}^{R} is effectively known. By the covariance restriction, u_{it}^{R} is uncorrelated with u_{it}^{B} and u_{it}^{S}, whereas it is partially correlated with ln(Revenue$_{it}$). Thus, we effectively have u_{it}^{R} as an instrument available for estimating the blog and search equations. The estimation procedure is as follows. First, we estimate (2-5) by an instrumental variable technique and save the residuals, namely $\hat{u}_{it}^{R}$. Then we estimate the blog and search equations using $\hat{u}_{it}^{R}$ as the instrument for ln(Revenue$_{it}$). The fact that $\hat{u}_{it}^{R}$ depends on estimates from a prior stage does not affect consistency of the estimators of the blog and search equations (Wooldridge 2002, p. 207).

2.5. Estimation

2.5.1. Estimation of the Pre- and Post-Launch Period Models

To each equation of the model, we apply a generalized method of moments (GMM) procedure that accommodates the arbitrary serial correlation in the error term. Let i be the index for a movie (i = 1, …, N) where N = 153 and t be the index for time (t = 1, …, T_i). In the pre-launch period model, T_i = 30 for each movie; in the post-launch period model, T_i = 10. Let y_{it} be the dependent variable of the estimation equation, x_{it} be the corresponding row vector of explanatory variables, and z_{it} be the corresponding row vector of instruments. For movie i, let y_i be the $T_i \times 1$ vector of the dependent variable of the focal equation, obtained by stacking y_{it} from t = 1, …, T_i. X_i and Z_i are similarity constructed by stacking x_{it} and z_{it}.

For each equation, the GMM estimation steps are as follows. (i) For the focal equation, apply the 2SLS estimation and obtain residuals. (ii) Use these residuals to obtain the GMM weighting matrix that is robust to arbitrary serial correlation of the error term. (iii) With the weighting matrix, estimate the parameters of the focal equation by GMM. The GMM weighting matrix in step (ii) is in (2-7).

(2-7)
$$W = \left(N^{-1} \sum_{i=1}^{N} Z_i' \hat{u}_i \hat{u}_i' Z_i \right)^{-1},$$

where $\hat{u}_i$ is the $T_i \times 1$ vector of residuals obtained from the 2SLS regression in (i). The GMM estimator and its asymptotic robust covariance matrix are

(2-8)
$$\begin{aligned} \hat{\beta}_{GMM} &= \left(X'ZWZ'X \right)^{-1} X'ZWZ'Y, \\ V(\hat{\beta}_{GMM}) &= \left(\sum_{i=1}^{N} \hat{X}_i' \hat{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \hat{X}_i' \hat{u}_i \hat{u}_i' \hat{X}_i \right) \left(\sum_{i=1}^{N} \hat{X}_i' \hat{X}_i \right)^{-1}, \end{aligned}$$

where X, Z, and Y are obtained by stacking X_i, Z_i, and y_i from i = 1, …, N, $\hat{X}_i = Z_i (Z_i' Z_i)^{-1} Z_i' X_i$, and $\hat{u}_i = y_i - X_i \hat{\beta}_{GMM}$. In the post-launch analysis, we first estimate the revenue equation and obtain $\hat{u}_{it}^{R} = y_{it} - x_{it} \hat{\beta}_{GMM}$, the GMM residual of the revenue equation. Then we use $\hat{u}_{it}^{R}$ as an instrument in estimating the blog and search equations.

The following variables are used as instruments.

- Blog equation: weekly advertising spending, weekly search index of the keyword "opening movie," weekly traffic to the five blog sites, the holiday dummy variable, movie characteristics, the week dummy variables, and residuals from the revenue equation (for the post-launch period model).

- Search equation: weekly advertising spending, weekly search index of the keyword "opening movie," weekly traffic to the five blog sites, the holiday dummy variable, movie characteristics, the week dummy variables, and residuals from the revenue equation (for the post-launch period model).
- Revenue equation: weekly advertising spending, weekly search index of the keyword "opening movie," weekly traffic to the five blog sites, the holiday dummy variable, movie characteristics, and the week dummy variables.

2.5.2. Estimation of the Opening-Week Model

The opening-week model is a cross-section data model. The GMM estimator for this model is similar to (2-7) and (2-8) except that the cross-section heteroskedasticity, instead of the arbitrary serial correlation, is considered to construct the weighting matrix W in (2-7).

The opening-week model contains the following endogenous variables: the opening week's advertising spending, the number of opening screens, the opening-week blog volume, and the opening-week search volume. As such, for instruments, we include the total pre-launch advertising spending of the movie, the Google search index of the keyword "opening movie" in the movie's launch week, the traffic to the five blog sites in the movie's launch week, and various movie characteristics.

2.6. Empirical Results

Table 2.4 shows the results of the pre-launch analysis, comparing the ordinary least squares (OLS) results with the GMM results.[2]

Table 2.4: Estimation Results: Pre-Launch Period (N=153, T=30)

(a) Blog equation (DV: log of weekly blog volume)

	OLS				GMM			
Variable	Coef.	SE	P-val.		Coef.	SE	P-val.	
Ad, same week	0.020	0.009	0.031	***	0.027	0.011	0.010	***
one week ago	0.012	0.007	0.110		0.015	0.009	0.114	
two weeks ago	-0.004	0.007	0.607		-0.011	0.010	0.289	
three weeks ago	0.007	0.009	0.430		-0.001	0.011	0.957	
four weeks ago	0.015	0.010	0.132		0.012	0.009	0.214	
Searching, same week	0.101	0.015	0.000	***	-0.042	0.109	0.701	
one week ago	-0.018	0.013	0.159		0.015	0.113	0.896	
two weeks ago	-0.019	0.011	0.089	*	0.072	0.097	0.458	
three weeks ago	-0.029	0.009	0.001	***	-0.008	0.136	0.951	
four weeks ago	-0.029	0.010	0.003	***	-0.118	0.099	0.234	
Holiday	0.001	0.024	0.975		-0.002	0.024	0.937	
Traffic to the five blog sites	1.564	0.406	0.000	***	1.369	0.392	0.001	***
R^2	0.043				N.A.			
Adj. R^2	0.035				N.A.			
SSR	0.754				0.786			
Corr. coef. between actual and fitted values in level	0.750				0.730			

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

[2] The estimates of week-fixed effects are omitted from Table 2.4 to avoid clutter.

(b) Search equation (DV: log of weekly search volume)

	OLS				GMM			
Variable	Coef.	SE	P-val.		Coef.	SE	P-val.	
Ad, same week	0.041	0.011	0.000	***	0.053	0.014	0.000	***
one week ago	0.036	0.010	0.001	***	0.051	0.016	0.001	***
two weeks ago	-0.016	0.007	0.033	**	-0.007	0.014	0.627	
three weeks ago	-0.005	0.009	0.584		0.005	0.016	0.759	
four weeks ago	-0.006	0.007	0.389		-0.008	0.017	0.646	
Blogging, same week	0.310	0.036	0.000	***	0.020	0.320	0.950	
one weeks ago	0.221	0.031	0.000	***	0.235	0.258	0.363	
two weeks ago	0.064	0.025	0.010	***	-0.157	0.404	0.698	
three weeks ago	0.037	0.024	0.121		0.058	0.316	0.854	
four weeks ago	0.001	0.023	0.969		0.137	0.242	0.573	
Holiday	0.038	0.030	0.205		0.056	0.047	0.238	
Search index of keyword “opening movie”	0.226	0.073	0.002	***	0.237	0.085	0.005	***
R^2	0.064				N.A.			
Adj. R^2	0.057				N.A.			
SSR	1.078				1.121			
Corr. coef. between actual and fitted values in level	0.923				0.680			

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

First of all, in the pre-launch period of movies, consumers’ blogging and searching activities have no mutual effect on one another, as the weekly blog volume and search volume do not influence each other. Rather, consumers’ blogging and searching activities are attributed solely to advertising. The OLS result of a significant contemporaneous association between blog volume and search volume is likely to result from the common upward trend that the two series share during the pre-launch period (see Figure 2.1). The difference between the OLS results and the GMM results shows that, without proper instruments for the blog and search volume, researchers are likely to conclude that pre-launch blogging activity triggers consumers’ pre-launch searching activity, and vice versa.

Second, in the pre-launch periods of movies, advertising has only immediate effects on blogging activity, whereas it has carryover effects on searching activity. The effect of advertising on search activities lasts one week, according to the GMM results. During the pre-launch periods of movies, the total elasticity of blog volume to advertising is 0.027, and that of search volume to advertising is 0.104. Note that compared to the GMM results, the OLS results underestimate the importance of advertising (0.1 versus 0.06 for generating search activities; 0.03 versus 0.02 for generating blog postings).

Third, holidays in the pre-launch period have no influence on blogging and searching activities of consumers. Fourth, the OLS results lack face validity as some lagged effects are estimated to be negative. For example, according to the OLS results, the weekly pre-launch blog volume is negatively associated with the lagged weekly search volume; also, according to the OLS results, advertising two weeks before the current week decreases the current week's search activities. Lastly, it is empirically supported that the reach of the five blog sites and the search index of the keyword "opening movie" are significantly associated with their corresponding endogenous variables—i.e., the weekly blog volume and weekly search volume of individual movies.

Now consider the post-launch analysis results. Table 2.5 shows the OLS and GMM estimation results for the post-launch period model. To show how the inclusion of search (blog) volume reveals new findings, we present in Table 2.6 (Table 2.7) the GMM estimation results of the nested post-launch period model that omits search (blog) volume.[3] Figure 2.3 juxtaposes statistically significant relationships between the endogenous variables in the three versions of the post-launch period model.

[3] The estimates of week-fixed effects are omitted from the tables to avoid clutter.

Table 2.5: Estimation Results: Post-Launch Period (N=153, T=10)

(a) Blog equation (DV: log of weekly blog volume)

	OLS				GMM			
Variable	Coef.	SE	P-val.		Coef.	SE	P-val.	
Advertising	0.074	0.036	0.040	**	0.040	0.044	0.360	
Searching	0.372	0.071	0.000	***	0.674	0.152	0.000	***
Revenue	0.096	0.062	0.121		-0.032	0.076	0.677	
Holiday	-0.120	0.086	0.165		-0.166	0.070	0.017	**
Traffic to the five blog sites	1.944	0.552	0.000	***	2.258	0.565	0.000	***
R^2	0.296				N.A.			
Adj. R^2	0.288				N.A.			
SSR	1.338				1.372			

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

(b) Search equation (DV: log of weekly search volume)

	OLS				GMM			
Variable	Coef.	SE	P-val.		Coef.	SE	P-val.	
Advertising	0.055	0.033	0.096	*	0.067	0.032	0.034	**
Blogging	0.164	0.043	0.000	***	0.477	0.075	0.000	***
Revenue	0.455	0.059	0.000	***	0.245	0.068	0.000	***
Holiday	0.129	0.059	0.027	**	0.100	0.048	0.037	**
Search index of keyword “opening movie”	0.288	0.171	0.092	*	0.315	0.180	0.079	*
R^2	0.535				N.A.			
Adj. R^2	0.530				N.A.			
SSR	0.981				1.089			

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

(c) Revenue equation (DV: log of weekly revenue)

	OLS				GMM			
Variable	Coef.	SE	P-val.		Coef.	SE	P-val.	
Advertising	0.205	0.014	0.000	***	0.201	0.015	0.000	***
Blogging	0.051	0.025	0.038	**	0.084	0.055	0.127	
Searching	0.195	0.033	0.000	***	0.250	0.067	0.000	***
Holiday	0.123	0.036	0.001	***	0.100	0.035	0.004	***
Screens	0.755	0.030	0.000	***	0.629	0.055	0.000	***
R^2	0.913				N.A.			
Adj. R^2	0.912				N.A.			
SSR	0.565				0.588			

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

Table 2.6: Estimation Results: The Post-Launch Period Model Omitting Search Volume

(a) Blog equation (DV: log of weekly blog volume)

	OLS			GMM		
Variable	Coef.	SE	P-val.	Coef.	SE	P-val.
Advertising	0.097	0.035	0.006 **	0.252	0.036	0.000 ***
Searching	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Revenue	0.281	0.063	0.000 ***	-0.090	0.097	0.352
Holiday	-0.072	0.089	0.416	-0.092	0.074	0.215
Traffic to the five blog sites	1.637	0.563	0.004 ***	1.479	0.558	0.008 ***
R^2	0.241			N.A.		
Adj. R^2	0.234			N.A.		
SSR	1.388			1.469		

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

(b) Revenue equation (DV: log of weekly revenue)

	OLS			GMM		
Variable	Coef.	SE	P-val.	Coef.	SE	P-val.
Advertising	0.237	0.016	0.000 ***	0.216	0.016	0.000 ***
Blogging	0.093	0.026	0.001 ***	0.194	0.045	0.000 ***
Searching	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Holiday	0.174	0.040	0.000 ***	0.135	0.035	0.000 ***
Screens	0.821	0.030	0.000 ***	0.731	0.047	0.000 ***
R^2	0.902			N.A.		
Adj. R^2	0.901			N.A.		
SSR	0.600			0.627		

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

Table 2.7: Estimation Results: The Post-Launch Period Model Omitting Blog Volume

(a) Search equation (DV: log of weekly search volume)

	OLS				GMM			
Variable	Coef.	SE	P-val.		Coef.	SE	P-val.	
Advertising	0.068	0.035	0.055	*	0.117	0.035	0.001	**
Blogging	N.A.	N.A.	N.A.		N.A.	N.A.	N.A.	
Revenue	0.501	0.058	0.000	***	0.289	0.069	0.000	***
Holiday	0.102	0.060	0.090	*	0.099	0.053	0.063	*
Search index of keyword “opening movie”	0.309	0.172	0.073	*	0.250	0.152	0.101	
R^2	0.509				N.A.			
Adj. R^2	0.504				N.A.			
SSR	1.007				1.046			

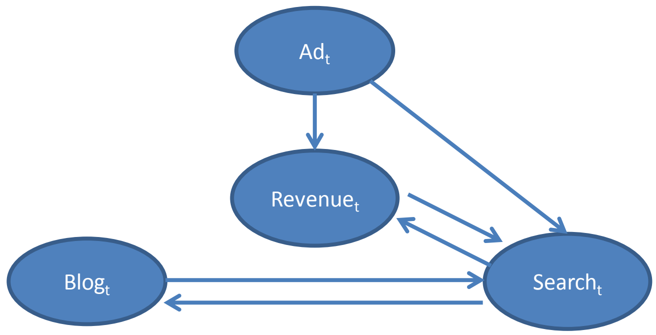
Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

(b) Revenue equation (DV: log of weekly revenue)

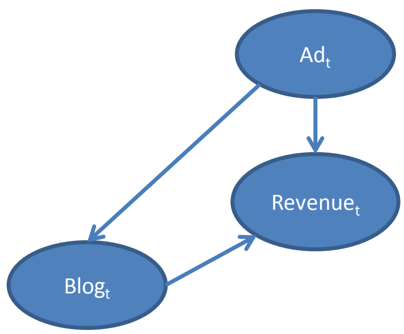
	OLS				GMM			
Variable	Coef.	SE	P-val.		Coef.	SE	P-val.	
Advertising	0.209	0.014	0.000	***	0.199	0.016	0.000	***
Blogging	N.A.	N.A.	N.A.		N.A.	N.A.	N.A.	
Searching	0.214	0.034	0.000	***	0.346	0.063	0.000	***
Holiday	0.113	0.037	0.002	***	0.061	0.038	0.105	
Screens	0.757	0.031	0.000	***	0.599	0.057	0.000	***
R^2	0.912				N.A.			
Adj. R^2	0.911				N.A.			
SSR	0.569				0.605			

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

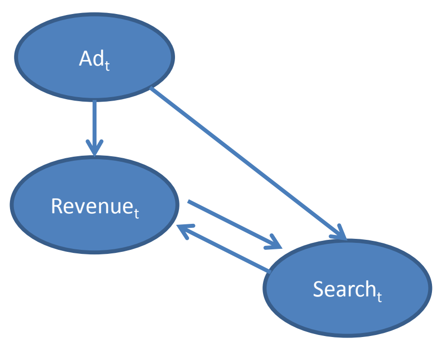
Figure 2.3: Comparison of Statistically Significant Relationship in the Three Versions of the Post-Launch Period Model



(a) The Full Model



(b) The Model Omitting Search Activity



(c) The Model Omitting Blogging Activity

First of all, the estimation results reveal the role of search activity in post-launch period of movies. Consumer searching activity works as a mediator between advertising and revenue, advertising and blogging, and blogging and revenue. Importantly, by comparing Figure 2.3(a) and 2.3(b), we find that search volume fully mediates the effect of blog volume on revenue. In other words, if there is no searching activity by consumers, blogging activity does not influence revenue. Furthermore, the more searching is done for a movie, the larger effect the blog postings have on movie revenue. To summarize, by adding online search volume in the model, we show that the positive effect of WOM volume on sales—that is found in the previous research (Dhar and Chang 2009; Duan, Gu, and Whinston 2008a; Duan, Gu, and Whinston 2008b; Liu 2006)—is good as long as WOM is exposed to consumers. This result supports Onishi and Manchanda (2012) who find that not the number of blogs generated but the number of blogs viewed matters for market success of new products.

Second, advertising is the main driver of the financial outcome of movies. Advertising's direct elasticity of weekly movie revenue is 0.201 and its indirect elasticity of weekly movie revenue through consumer search is 0.017 (= 0.067×0.250), amounting to the total elasticity of 0.218. Note that the indirect effect of advertising on revenue through consumer search activity—0.017—is almost negligible in comparison with its direct effect on revenue—0.201. In other words, advertising can generate nearly the same level of revenue without consumer searching activity. The story changes when it comes to the effect of blogging activity on revenue. The effect of blogging on revenue is realized only indirectly, namely through consumer searching activity. Note that the elasticity of revenue to blog volume is 0.119 (= 0.477×0.250), which is substantially lower than the advertising's elasticity of revenue.

Third, in terms of the direct effectiveness on movie revenue, consumer searching activity is at least as effective as advertising. The searching elasticity of revenue is 0.250, while the advertising elasticity of revenue is 0.201. Note that blogging activity has no direct effect on revenue if advertising and searching activity are controlled for. Therefore, if the purpose is to predict weekly revenue, monitoring and forecasting weekly search volume is sufficient—i.e., once search volume is known, blog volume is unnecessary for prediction of revenue.

Fourth, Table 2.5(b) shows that consumers' generic interest for opening movies in a certain week—as measured by the weekly Google search index of the keyword "opening movie"—influences online search volume of individual movies that are theatrically running in the week. Perhaps, consumers' generic interest for opening movies in a week spills over to the interest for individual movies that are theatrically running in the week (Rutz and Bucklin, 2011). The estimation result indicates that one percent increase in the search index of keyword "opening movie" leads to 0.315 percent increase in consumers' searching activity for a theatrically running movie. This implies that, the advertising spending being the same, an ad of a movie aired when the generic interest is higher than usual generates more consumer search than the same ad aired when the generic interest level is lower than usual. Examining when the weekly search index of "opening movie" is high during a year may help studio managers allocate advertising budgets.

Fifth, consumers' searching activity is influenced through multiple channels: advertising, online WOM, and movie consumption. That is, consumer interest in a theatrically running movie (i.e., consumer search activity) is developed by not only commercial communication but also peer consumers' WOM and movie consumption. In terms of the effectiveness for developing consumer interest for individual movies, WOM is the most effective (elasticity: 0.477) and advertising is the least effective (elasticity: 0.067) with movie consumption in the middle (elasticity: 0.245).

Sixth, the estimation results of the full model (Figure 2.3(a)) indicate that movie consumption influences blogging activity–indirectly through consumer search activity. As such, blog postings in post-launch period of movies are likely to contain consumer opinion about movies.

Seventh, there are a few notable differences in the pre- and post-launch relationships of the focal variables. The first is the mutual relationship between blogging and searching activity. Blogging and searching activity influence each other in the post-launch period, whereas they do not in the pre-launch period. The second difference is the impact of holidays. In the post-launch period, holidays increase consumers' searching activity as well as movie revenue, while, in pre-launch period, holidays do not influence consumers' searching activity.

Lastly, the GMM results that correct for endogeneity are substantially different from the OLS results that do not correct for endogeneity. The OLS results find more significant relationships between variables than the GMM results, perhaps because the focal variables share a common downward trend in the post-launch period. This analysis shows the importance of correcting for endogeneity.

Lastly, consider the opening-week analysis in Table 2.8.

Table 2.8: Estimation Results: Opening-Week (N=153)

	OLS			GMM		
Variable	Coef.	SE	P-val.	Coef.	SE	P-val.
Advertising	0.170	0.067	0.012 **	0.217	0.108	0.047 **
Screens	0.649	0.032	0.000 ***	0.673	0.061	0.000 ***
Pre-launch blog volume	0.040	0.053	0.454	0.018	0.117	0.875
Pre-launch search volume	0.081	0.074	0.276	0.271	0.126	0.034 **
Opening-week blog volume	-0.011	0.061	0.854	-0.010	0.152	0.949
Opening-week search volume	0.205	0.084	0.016 **	-0.111	0.186	0.553
Holiday	0.215	0.132	0.104	0.344	0.142	0.017 **
Critical Review	0.423	0.189	0.027 **	0.473	0.219	0.032 **
R^2	0.931			N.A.		
Adj. R^2	0.927			N.A.		
SSR	0.560			0.592		

Note. * P-val <0.1, ** P-val < 0.05, *** P-val < 0.01.

The most notable finding is the importance of pre-launch online search volume. Note that pre-launch search volume is significantly associated with opening-week revenue but pre-launch blog volume is not. For managers, this implies that the pre-launch searching activity of consumers, not pre-launch blogging activity, is the metric that they need to monitor for better prediction of opening- week movie revenue.

2.7. Implications

Our findings reveal several important implications for studio executives as well as researchers. First of all, while advertising and blogging are the drivers of post-launch movie revenue, there are big differences between the two. The first is the effectiveness of the two in generating movie revenue. The elasticity of weekly movie revenue to weekly advertising spending is almost twice as large as that to weekly blog volume. The second difference is related to the role of searching activity. Advertising can generates movie revenue without consumers' online search activity—perhaps because movie advertisements reach consumers through diverse media including offline as well as online channels; in contrast, blog postings cannot generate revenue without consumer online search activity.

Second, studio managers will be able to better predict the opening week revenue of movies by monitoring the pre-launch search volume of movies. To this end, the Google search index can serve as a readily available data source. Note that the pre-launch blog volume will be unnecessary to predict the opening-week revenue of movies, as long as the pre-launch search volume is known.

Third, the finding that pre-launch search activity influences opening-week revenue implies the importance of managing the pre-launch advertising schedule properly. The time-series of pre-launch search volume of a movie can be used to examine the pre-launch advertising effectiveness of the movie.

For researchers, this study exhibits that the online keyword search index and website traffic can provide a source of exogenous variations in certain online actions of consumers. Figure 2.4 summarizes key influencers of movie revenue and according managerial implications.

Figure 2.4: Key Influencers on Revenue

	Opening-week revenue	Weekly revenue after opening week
Key influencers	• Pre-launch search volume • Opening-week advertising	• Weekly advertising • Weekly online search activity
Implications	• Monitoring pre-launch search activity can help managers better predict opening-week revenues of movies. • The time-series of pre-launch search volume of a movie can be used to develop efficient pre-launch advertising schedule for the movie.	• For the purpose of predicting weekly revenue, knowing weekly advertising and search volume is sufficient; weekly blog volume does not contribute to the predictive performance. • Advertising influences revenue without the help of consumer search; online WOM need consumer search activity for it to influence movie revenue.

2.8. Conclusions

Despite the prevalence of consumers' media consumption activity, researchers have not paid as much attention to it as consumers' media generation activity. Taking the movie industry as an empirical case, this study has examined the dynamics between advertising, consumers' collective blogging activity, their collective online search activity, and market outcome.

Several important findings emerge. First, there is an important difference as to how advertising and blogging activity influence movie revenue. We find that blog postings need consumer search for them to influence movie revenue. Advertising, on the other hand, influences revenue without consumers' online search activity. In fact, in the post-launch period of a movie, the indirect effectiveness of advertising on revenue through consumer searching activity is so small that advertising generate almost the same level of movie revenue without consumer searching activity.

Second, advertising is the main driver of movie revenue throughout the movie life cycle. The opening-week revenue of a movie is influenced by the opening-week advertising and pre-launch search volume of the movie. But the pre-launch search volume of a movie is influenced mainly by advertising. As such, advertising is the dominant cause of opening-week revenue of movies. In the post-launch period, both advertising and consumer blogging activity influence the weekly movie revenue. But, advertising is almost twice more effective than the volume of blog posts in generating weekly movie revenue.

Third, once online search volume of a movie is controlled for, blog volume of the movie does not improve the performance in predicting movie revenue. For the purpose of predicting the market success of a movie, managers should focus on search volume, not blog volume.

Fourth, by adding online search volume in the post-launch period model, we were able to find that movie consumption influences consumer blogging activity in the post-launch period. As such, blog postings in the post-launch period are likely to contain consumers' opinion on the theatrically running movies.

This study provides implications for studio managers. First of all, to predict opening-week revenue of a movie, pre-launch search volume, not pre-launch WOM volume, is the metric that managers should monitor. Also to predict the post-launch weekly revenue of a movie, weekly online search volume of the movie should be monitored, not its weekly blog volume. Second, almost 80% of movie advertising is executed in pre-launch period or during opening-week. Therefore, finding an efficient pre-launch advertising schedule is an important task. Our findings that pre-launch advertising is the main driver of pre-launch search activity and that the pre-launch search activity has substantial effect on opening-week revenue suggest that studio managers may use the time-series of pre-launch search volume to measure the effectiveness of pre-launch movie advertising.

This study is subject to several limitations. First, we use blog volume to examine the effect of consumers' media generation. While blog postings are the only WOM of consumers before a new-product's launch, in the post-launch period consumers express their opinions through various channels including review sites. Second, we do not consider the valence of blog postings. This may be acceptable in the pre-launch period because the new product is not yet available, and as such there should be no valence information. But in the post-launch period, the WOM valence can influence movie-going decisions. Perhaps, one way of controlling for the WOM valence is to collect movie review data and include it in the model. Third, this study is conducted in the movie industry. For the generalization of the findings, we need to extend this study to other industries such as video games, music albums, and books.

Several research opportunities remain in this field. The content of blogs may influence consumer search behavior. For example, search may be greater when there is strong disagreement among consumers' opinions. Second, given that pre-launch search volume influences post-launch sales, determining the optimal allocation of a pre-launch advertising budget to maximize search volume is important. Lastly, consumer search activity may lead them to the related products' websites. Examining the relationship between search activity and traffic to product websites can be an interesting topic.

2.9. References

Baltagi, Badi H. (1995), *Econometric Analysis of Panel Data*. John Wiley & Sons, New York.

Basuroy, Suman, Kalpesh Kaushik Desai, and Debabrata Talukdar (2006), "An Empirical Investigation of Signaling in the Motion Picture Industry," *Journal of Marketing Research,* 43 (May), 287-95.

Beatty, Sharon E. and Scott M. Smith (1987), "External Search Effort: An Investigation Across Several Product Categories," *Journal of Consumer Research*, 14 (1), 83-95.

Chevalier, Judith and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-54.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," *Marketing Science*, 29 (5), 944-57.

Dhar, Vasant and Elaine A. Chang (2009), "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing*, 23 (4), 300-7.

Du, Rex Yuxing and Wagner A. Kamakura (2012), "Quantitative Trendspotting," *Journal of Marketing Research*, forthcoming.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008a), "The Dynamics of Online Word-of-Mouth and Product Sales – An Empirical Investigation of the Movie Industry," *Journal of Retailing*, 84 (2), 233-42.

______ (2008b), "Do Online Reviews Matter? – An Empirical Investigation of Panel Data," *Decision Support Systems*, 45 (4), 1007-16.

Einav, Liran (2007), "Seasonality in the U.S. Motion Picture Industry," *RADN Journal of Economics*, 31 (1), 329-54.

Elberse, Anita and Bharat Anand (2007), "The Effectiveness of Pre-Release Advertising for Motion Pictures: An Empirical Investigation Using a Simulated Market," *Information Economics and Policy*, 19, 319-43.

______ and Jehoshua Eliashberg (2003), "Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures," *Marketing Science*, 22 (3), 329-54.

Eliashberg, Jehoshua, Anita Elberse, and Mark A.A.M. Leenders (2006), "The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions," *Marketing Science*, 25 (6), 638-61.

Enders, Walter (2004), *Applied Econometric Time Series*, John Wiley & Sons, Hoboken, NJ.

Godes, David and Dina Mayzlin (2004), "Using Online Conversation to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545-60.

Gopinath, Shyam, Pradeep K. Chintagunta, and Sriram Venkataraman (2011), "Blogs and Local-market Movie Box-office Performance," SSRN working paper.

Ho, Jason Y.C., Tirtha Dhar, and Charles B. Weinberg (2009), "Playoff payoff: Super Bowl Advertising for Movies," *International Journal of Research in Marketing*, 26, 169-79.

Holtz-Eakin, Douglas (1988), "Testing for Individual Effects in Autoregressive Models," *Journal of Econometrics*, 39, 297-307.

Joo, Mingyu, Kenneth C. Wilbur, and Yi Zhu (2012), "Television Advertising and Online Search," SSRN Working Paper.

Kim (2013), “Pre-Launch Advertising Effectiveness of New Products: An Empirical Analysis Using Online Search Volume,” Unpublished Manuscript.

Kulkarni, Gauri, P. K. Kannan, and Wendy Moe (2012), “Using Online Search Data to Forecast New Product Sales,” *Decision Support Systems*, 52 (3), 604-11.

Liu, Yong (2006), “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing*, 70 (July), 74-89.

Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), “Consumer Information Search Revisited: Theory and Empirical Analysis,” *Journal of Consumer Research*, 23 (4), 263-77.

Punj, Girish N. and Richard Staelin (1983), “A Model of Consumer Information Search Behavior for New Automobiles,” *Journal of Consumer Research*, 19 (4), 366-80.

Onishi, Hiroshi and Puneet Manchanda (2012), “Marketing Activity, Blogging, and Sales,” *International Journal of Research in Marketing*, 29 (3), 221-34.

Rutz, Oliver J. and Randolph E. Bucklin (2011), “From Generic to Branded: A Model of Spillover in Paid Search Advertising”, *Journal of Marketing Research*, 48 (February), 87-102.

Ratchford, Brian T. and Narasimhan Srinivasan (1993), “An Empirical Investigation of Returns to Search,” *Marketing Science*, 12 (1), 73-87.

Srinivasan, Narasimhan and Brian T. Ratchford (1991), “An Empirical Test of a Model of External Search for Automobiles,” *Journal of Consumer Research*, 18 (2), 233-42.

Stephen, Andrew T. and Jeff Galak (2012), “The Effects of Traditional and Social Earned Media on Sales: A Study of a Microlending Marketplace,” *Journal of Marketing Research*, 49 (October), 624-39.

Trusov, Michael, Randolph E. Bucklin, and Koen Pauwels (2009), “Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site,” *Journal of Marketing*, 73 (September), 90-102.

Villanueva, Julian, Shijin Yoo, and Dominique M. Hanssens (2008), “The Impact of Marketing-Induced Versus Word-of-Mouth Customer Acquisition on Customer Equity Growth,” *Journal of Marketing Research*, 45 (February), 48-59.

Wooldridge, Jeffrey M (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press, Cambridge, MA.

3. Pre-Launch Advertising Effectiveness of New Products: An Empirical Analysis Using Online Search Volume

3.1. Introduction

Eighteen months before Ford introduced Fiesta, the company started a pre-launch marketing campaign called Fiesta Movement. Videos related to the campaign generated 6.5 million views on YouTube. When the car was released in late 2010, over 10,000 cars were sold in the first six days (McKinsey Quarterly 2012). Fiesta Movement is an example of pre-launch advertising, which has become the rule rather than the exception in many industries including automobiles, movies, computer games, smart phones and tablet PCs. Indeed, pre-launch advertising is vital for many reasons: short life cycle of new products requires good sales results quickly; pre-launch advertising can make consumers look forward to the new product launch and adjust their purchase planning accordingly; and managers can use pre-launch advertising to create buzz or to preempt competition.

This study aims to examine the pre-launch advertising effectiveness of new products, an important but under-researched topic. Taking the U.S. motion picture industry as the test case, we develop a model that relates the advertising schedule of new movies to the time-series of online search volume—i.e., collective online search activity of consumers—of the movies. Two key elements of our model are (i) consumers' willingness-to-search (WTS) that may change over the course of pre-launch period and (ii) the advertising effectiveness that may also vary over time. First, consumers' willingness-to-search for new-movie information may increase as the new movie's availability dates come closer. The intuition is that as the launch date of a movie comes closer, the more interested consumers will be in the movie and the higher willingness-to-search they will have for the movie information. Second, advertising effectiveness may vary over time as a function of advertising schedule. For example, Naik, Mantrala, and Sawyer (1998) find that advertising effectiveness at a specific time is determined by, among other things, the passage of time since the first execution of the advertising campaign, past advertising spending and the existence of ad hiatus. The two elements are incorporated in our model, which is applied to an empirical movie data set.

Our data set consists of the weekly advertising spending of 106 movies, weekly Google search indices of keywords of the movies, and the movies' characteristics. As the raw search index from Google Trends is comparable only across time within a time series, we devise a method that transforms the raw Google search index into cross-sectionally comparable search volume measures. This enables us to compare the advertising effectiveness across different movies. The parameters are drawn by a Gibbs sampler, which embeds the data-augmentation step to deal with the left-truncated property of the Google search index (Chib 1992), the forward-filtering/backward-sampling step to estimate time-varying parameters (West and Harrison 1997), and the hierarchical Bayesian approach to estimate heterogeneity across movies.

We compare our model with alternative models and show that the proposed model is superior to the alternative models in forecasting the following week's search volume of movies. The empirical analysis reveals interesting features of consumers' pre-launch interest development for new products. First, consumers' pre-launch response to advertising—in the form of keyword search—is substantially influenced by the timing of advertising. That is, pre-launch advertising more efficiently triggers consumer search in the weeks that immediately precede the release time than in the weeks that are far in advance of release. In other words, time-to-launch is a key variable that influences consumer search in the pre-launch period. Second, pre-launch advertising effectiveness varies over time. It is influenced by advertising schedules such as the passage of time since the first execution of the advertising campaign, the previous period's advertising intensity, and the existence of ad hiatus periods. Third, the time-to-launch effect and time-varying advertising effectiveness vary substantially across movies. The heterogeneity is explained by movie characteristics such as director power and whether the movie is a sequel. The findings imply that the same advertising schedule of different movies can generate different levels of consumer search. In the managerial implications section, we solve a constrained nonlinear optimization problem to find a more efficient allocation of an advertising budget over the pre-launch period.

The remainder of the paper is organized as follows. In the model section, we review relevant research and identify critical features of consumers' pre-launch search behavior. Then we develop a model of search volume process by incorporating the effect of the advertising outlay and time-to-launch. We then briefly introduce the data. In the empirical analysis section, we estimate the model in the Bayesian dynamic linear model (DLM) framework and discuss the managerial implications. We show how the model can be used to suggest a better allocation of a given advertising budget across time. We then formulate conclusions and areas for future research.

3.2. The Model

We propose that past advertising schedule and consumers' WTS are the two key elements that determine the online search volume of a new product at a specific pre-launch time of the product. Previous research has established that a different advertising schedule can generate a different market outcome because advertising effectiveness at a specific point in time is influenced by advertising outlay up to that time (Bass et al. 2007; Bruce 2008; Bruce, Zhang and Kolsarici 2012; Naik, Mantrala and Sawyer 1998; Pechmann and Stewart 1990). In addition to the advertising schedule effect, we propose that, over the course of a new product's pre-launch period, consumers have different levels of WTS for the new-product's information. The WTS, in combination with advertising effectiveness, determines consumers' search intensity at a specific time of pre-launch period. The intuition is that, regardless of how much advertising a consumer is exposed to, he or she would not search for the advertised product unless motivated to do so. Therefore, we represent online search volume at a specific point in time as a multiplicative function of marketing communication up to that time and consumers' willingness-to-search at that time.

Let index i denote product and index t denote time. The online search volume of product i—i.e., collective online search activity of consumers for product i—at t is represented as a function of advertising outlay up to t, consumers' willingness-to-search for product i's information at t, and the error term.

(3-1) $$S^{*}_{it} = \exp\left(G_{it}\right) \cdot \lambda_{it} \cdot \exp\left(v_{it}\right), \text{ where } v_{it} \sim N(0, V_i)$$

where S^{*}_{it} is the online search volume of product i at t, G_{it} is the effect of advertising schedule up to t and λ_{it} is the WTS that is independent of past advertising outlay. Exp(v_{it}) is the error term where v_{it} follows a normal distribution with mean 0 and variance V_i. The composite term $\exp(G_{it}) \cdot \lambda_{it}$ corresponds to the advertising goodwill of product i at t.

3.2.1. Modeling the Effect of Advertising Schedule

G_{it} summarizes the effect of current and past advertising outlay of product i on its search volume, net of the effect of WTS. Following previous literature, we assume that G_{it} is influenced by the current advertising, A_{it}, as well as the carryover from past advertising.

$$G_{it} = q_{i,t-1} g(A_{it}) + (1-\delta_i) G_{i,t-1} + w_{it}^G, \text{ where } w_{it}^G \sim N(0, W_i^G), \tag{3-2}$$

A_{it} is the advertising intensity of product i at t, $q_{i,t-1}$ is advertising effectiveness of product i at t.[4] Note that $q_{i,t-1}$ is allowed to vary over time. $1-\delta_i$ is the carryover rate of advertising. w_{it}^G is the normally distributed error with mean 0 and variance W_i^G.

Consistent with previous literature, we assume that advertising effectiveness varies over time due to the following three factors: the previous period's advertising intensity, the restoration of advertising effectiveness when the ad is off and the natural deterioration of advertising effectiveness due to the passage of time. The rationale for the three factors is well reviewed in Naik, Mantrala and Sawyer (1998) and studies that follow it (e.g., Bass et al. 2007). By incorporating those factors, the change in advertising effectiveness at t is represented in (3-3).

$$\Delta q_{i,t-1} = -\alpha(A_{i,t-1}) q_{i,t-2} + \delta_i I(A_{i,t-1} = 0)(1 - q_{i,t-2}) + w_{it}^q, \tag{3-3}$$

$$\text{where } \alpha(A_{i,t-1}) = c_i + w_i \cdot a(A_{i,t-1}) \text{ and } w_{it}^q \sim N(0, W_i^q).$$

The parameter c_i represents the change of advertising effectiveness due to the passage of time and w_i is the effect of the previous period's advertising intensity on the current period's

[4] We intentionally use subscript t-1 to represent the advertising effectiveness at time t. The reason becomes clear when we transform the model into a state-space model framework. For all other variables, subscript t-1 represents the variable's corresponding quantity at t-1. For example, $A_{i,t-1}$ is the advertising spending at t-1, not t.

advertising effectiveness. δ_i is the restoration of advertising effectiveness when the advertising is off in the previous period, i.e., when $I(A_{i,t-1} = 0)$. The restoration of advertising effectiveness occurs due to consumers' forgetting about the specifics of previous advertising content (Grass and Wallace 1969).

3.2.2. Modeling the Effect of Willingness-to-Search

Consumers do not search for a new product just because they are exposed to an advertisement of the product. Instead, the advertisement will generate search only when consumers think their search activities are sufficiently beneficial to them. In other words, there exists a concept termed "willingness-to-search" that moderates the effect of advertising on collective search activity of consumers.

Previous research on consumer search behavior implies that consumers' WTS for product i at t is proportional to the benefit of searching for that information at that time (Beatty and Smith 1987; Punj and Staelin 1983; Srinivasan and Ratchford 1991; Weitzman 1979). We propose three determinants of WTS: the development of product interest over time, the decay of consumer memory, and product characteristics. The first two elements apply after a new product's first advertising through the launch time (i.e., during the pre-launch period since the first advertising). We will show that they allow the pre-launch WTS to gradually increase as a new-product launch is approaching. The last element, product characteristics, applies regardless of advertising schedule and explains variation in search volume across different products.

Development of Product Interest over Time. Consumer interest in a new product develops as the launch time of the new product approaches, perhaps because consumers perceive higher personal relevance for events that will happen sooner than later. For example, in the early pre-launch period of a new movie, interest for the movie may not be well developed because there are many other alternative movies that are either available or soon-to-be available to consumers. In the same vein, literature on intertemporal decision theory finds that individuals often place higher value on a near-future reward than on a distant-future reward, even when the distant-future reward is larger (Trope and Lieberman 2003). These arguments imply that consumers' interest for a new product and thus their WTS for the product tends to increase as the release time of the product approaches. Let us assume that consumer interest decreases by a factor of k ($k > 1$) during a unit time interval of a pre-launch period and normalize the interest level in the release week by one. Further, let us represent the pre-launch time by negative numbers (-1, -2, …), the launch time by zero and the post-launch time by positive numbers (1, 2, …). Then, the interest level t periods before a new-product launch can be represented as $1/k^{|t|}$.

Decay of Consumer Memory over Time. Human memory decays over time and so does the probability of recalling the advertised content. If a consumer expects that the new-product information he acquires at a specific pre-launch time of the product will be gradually forgotten—due to decay of memory, the perceived benefit of acquiring the new-product information in the pre-launch time may be affected by that expectation. In other words, consumers' WTS will decrease as a function of expected information loss between the time of information acquisition and the time of new-product availability. Suppose that consumers forget the information about a new product at a constant rate d ($0< d <1$) during a unit time interval in the pre-launch period. Then, the probability that the information that was collected at t periods before launch is still active in consumers' memory at the launch time will be $(1- d)^{|t|}$, where $|t|$ measures the time distance between the search time t and the new-product launch time denoted by zero. The smaller the decay rate d is, the more willing consumers will be to search new-product information across the pre-launch periods.

Product Characteristics. Consumers are inclined to search more for products with certain characteristics than others. For example, Kim (2013) finds that consumers search more for movies of higher quality and higher quality-uncertainty. Therefore, consumer WTS may vary across products as a function of product characteristics. Let μ_i summarize the effect of the characteristics of product i. Then, μ_i can be represented in a multiplicative form as follows.

$$\mu_i = X_i^{\beta_1} e^{D_i \beta_2}, \tag{3-4}$$

where X_i is the vector of numerical characteristics and D_i is the vector of categorical characteristics of product i.

To summarize, λ_{it}, the WTS for product i at time t, is represented as a multiplicative function of the above factors.

$$\lambda_{it} = k_i^{-|t|\cdot I(f_i \le t \le 0)} \left(1 - d_i\right)^{|t|\cdot I(f_i \le t \le 0)} \mu_i = e^{\psi_i |t|\cdot I(f_i \le t \le 0)} \mu_i, \text{ where } e^{\psi_i} = (1 - d_i)/k_i, \tag{3-5}$$

where f_i represents the time of the first advertising of movie i. Note that time-to-launch plays an important role in determining the search volume in pre-launch period. Also, note that 1-d and k are not separately identified due to the lack of variation that distinguishes d and k. In the empirical analysis, we estimate the composite effect of the two, i.e., ψ_i.

3.2.3. Heterogeneity

We allow the time-to-launch parameters (ψ_i,), the carryover rate (1-δ_i) and the parameters of advertising effectiveness (c_i, w_i) to vary across products as a function of product characteristics. Let us denote $\left(\psi_i, \delta_i, c_i, w_i\right)'$ by ϕ_i and let M_i be the collection of the characteristics of product i. The matrix M_i is appropriately arranged to conform to matrix multiplication: $M_i = I_4 \otimes (X_i', D_i')$ where X_i and D_i are vectors of product characteristics in (3-4). Then the heterogeneity is modeled as in (3-6).

(3-6) $$\phi_i \equiv \left(\psi_i, \delta_i, c_i, w_i\right)' = M_i \gamma + \xi_i \text{, where } \xi_i \sim N(0, \Sigma_\xi).$$

3.3. The Data

We apply the model to a movie data set. Our data set mostly consists of movies that were widely released in 2009 in the U.S. For each movie, we collect weekly advertising spending and weekly online search index from twenty weeks before its launch until five weeks after the launch. For the online search index, we use Google Trends. We also collect various movie characteristics. For our estimation, we remove some movies from the set. First, we discard movies whose first pre-launch advertising started less than six weeks before their launch weeks. Those movies have too short a period to estimate parameters. Second, we do not include movies whose titles are too general to affect the validity of the search index from Google. The final data set contains 106 movies. Because the time of the first advertising varies movie by movie, the panel data is unbalanced. We account for this unbalanced structure when estimating our model. Table 3.1 summarizes our variables and their sources.

Table 3.1: Variables and Data Sources

Category	Variable	Source of Data
Dependent variable	Weekly search volume	Google
Marketing variable	Weekly advertising spending	Nielsen
Movie characteristics	Total U.S. box-office revenue of movies with which the focal movie's director was involved as a director, writer, or producer since 1990	IMDb
	Average rating of past movies that the director of with which the focal movie was involved as a director, writer, or producer since 1990 [range: 1 – 10]	
	Average rating of past movies that the director of with which the focal movie was involved as a director, writer, or producer since 1990	
	Genre	
	MPAA rating	
	Sequel: whether a movie is a sequel or not	

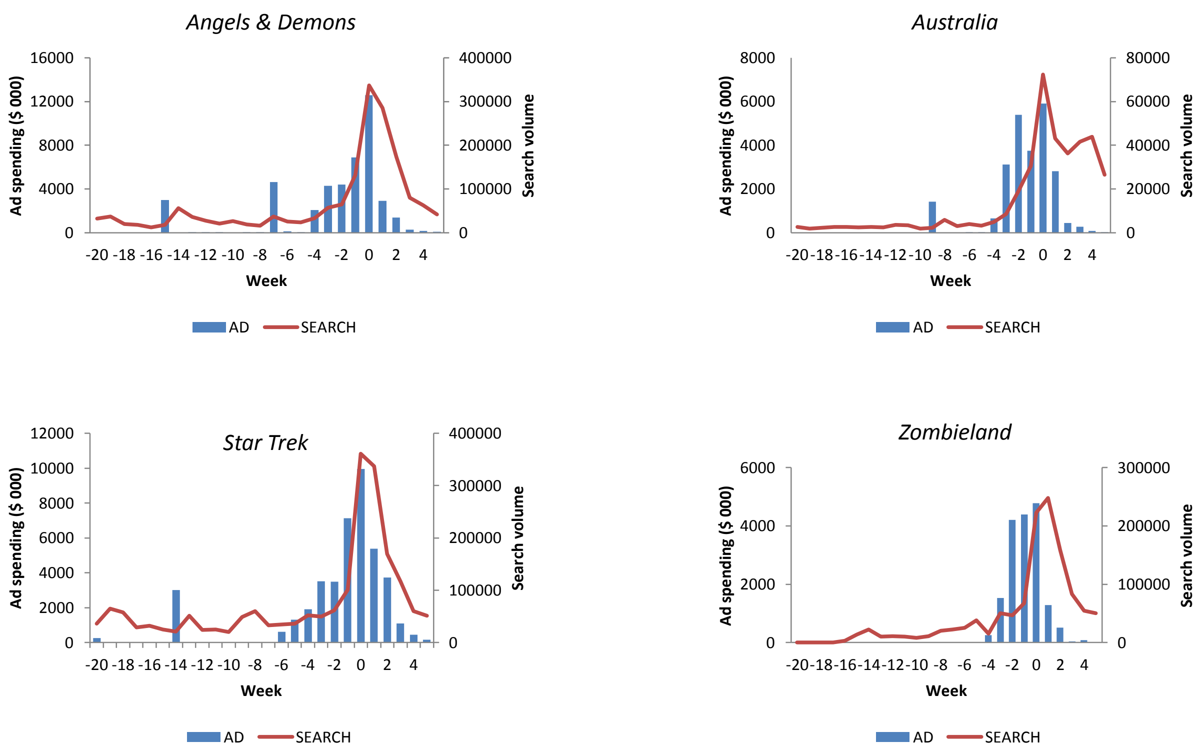
3.3.1. Advertising

The advertising data is provided by Nielsen and covers all major media outlets including television, print, radio Internet, and outdoor channels. We discard the weeks prior to the twenty weeks before release ($t < -20$) because advertising is exceedingly sparse before that time. The average advertising spending of the 106 movies is \$24.2 million (Std. Dev. = \$11.2 million) with 80% of the advertising budget executed in the pre-launch periods or during the release week. On average, the movies in the data set started their first advertising about 12.5 weeks before their releases (Std. Dev. = 6.1 weeks before release).

3.3.2. Online Search Volume

Google Trends provides a weekly search index of keyword queries entered into the Google search engine. Because the index is normalized to conceal the actual search volume of the keyword, researchers cannot compare the search volumes across different keywords if the raw search index is used. We avoid this problem with a method that transforms the raw search indices from Google into cross-sectionally comparable search volume measures. The detailed methodology of collecting weekly Google search indices of movie keywords and transforming them into cross-sectionally comparable measures can be found in chapter 5, the appendix to the dissertation. Figure 3.1 exhibits the real advertising and search volume of four movies.

Figure 3.1: Examples of Weekly Advertising and Search Volume



3.3.3. Movie Characteristics

Movie characteristics are collected from IMDb.com. We include three director power variables. The first is the total U.S. box-office revenue of movies in which the director of each focal movie has been involved either as a director, a writer or a producer since 1990 until one calendar year before the focal movie's release. The second variable is the average user rating of the past movies in which the director of each focal movie has been involved since 1990 until one calendar year before the focal movie's release. The last variable is the average of standard deviations of user ratings of each past movie in which the focal movie's director has been involved since 1990 until one calendar year before the focal movie's release. The total number of past movies directed/written/produced by the directors of the 106 focal movies and whose U.S. box-office revenue is reported to IMDb is 949. On average, a director was involved with 8.95 movies since 1990 until the one calendar year before the focal movie's release. The average of the total U.S. gross revenues of the past movies is $905 million, and the standard deviation is $1,040 million. The average user rating of the 949 movies is 6.74 (10-point scale). The average of standard deviations of user ratings of the 949 movies is 1.96. Table 3.2 summarizes descriptive statistics of variables.

Table 3.2: Descriptive Statistics of the Variables (N = 106)

	Mean	Median	Std. Dev.	Min	Max
Advertising expenditure ($000)	24,222	25,075	11,230	0.9	50,781
Search volume (000)	1,223	491	2,977	16	21,617
Total U.S. box-office revenue of past movies of focal directors ($ 000)	905,000	505,000	1,040,000	0.2	6,520,000
Average user ratings of past movies of focal directors	6.74	6.81	0.65	5.22	8.71
SD of user ratings of past movies of focal directors	1.96	1.95	0.21	1.55	2.80
Major Genre (%)	Action: 22.6, Comedy: 27.4, Drama: 17.0				
MPAA Rating (%)	G: 1.9, PG: 23.6, PG13: 44.3, R: 30.2				
Sequel	10 movies are sequels.				

3.3.4. A Preliminary Analysis

To check the effect of time-to-launch on search volume generation, we estimate a log-transformed model of eq (3-1). Because the objective of this analysis to quickly check the effect of the time-to-launch variable, we simplify our analysis by substituting the advertising goodwill G_{it} with the current and past advertising efforts as in (3-7).

$$G_{it} = \alpha_0 \log(A_{it}) + \ldots + \alpha_k \log(A_{i,t-k}) \tag{3-7}$$

Then, (3-1) is log-transformed to (3-8).

$$\log(S_{it}) = \alpha_0 \log(A_{it}) + \ldots + \alpha_k \log(A_{i,t-k}) + \psi \cdot |t| \cdot I(t \leq 0) + \beta_1 \log(X_i) + \beta_2 D_i + \varepsilon_{it}. \tag{3-8}$$

The variable $|t| \cdot I(t \leq 0)$ represents the time-to-launch at pre-launch week t. Google keyword search index is left-truncated at an unknown threshold. Therefore, we estimate a Tobit model with the specification of (3-8). Note that the unbalanced panel structure is not considered in this preliminary analysis. Table 3.3 shows the result.

Table 3.3: Preliminary Analysis: Tobit Model Estimation
(Left-censored observations: 623, Total observations: 3,193)

Covariate	Coefficient	SE	P val.
Constant	-22.08	1.95	0.00
Advertising at t	0.12	0.03	0.00
Advertising at t-1	0.09	0.04	0.02
Advertising at t-2	0.03	0.04	0.51
Advertising at t-3	0.05	0.04	0.23
Advertising at t-4	0.06	0.03	0.07
Time-to-launch	-0.12	0.01	0.00
Total U.S. B-O revenue of past movies of the focal director	0.32	0.03	0.00
Average rating of past movies of the focal director	9.73	0.82	0.00
SD of ratings of past movies of the focal director	9.03	0.75	0.00
Genre: Action	1.40	0.16	0.00
Genre: Comedy	-0.89	0.15	0.00
Genre: Drama	0.70	0.18	0.00
MPAA: R	-1.72	0.42	0.00
MPAA: PG	-1.07	0.42	0.01
MPAA: PG13	-1.72	0.41	0.00
Sequel	3.19	0.19	0.00
Log-likelihood			-9032.71
AIC			4.74
BIC			4.77

The estimation result of this simple analysis shows that time-to-launch plays an important role in generating search volume in the pre-launch period. If all else are equal, the willingness-to-search decreases by a factor of 0.89 (= exp(-0.12)) each week during the pre-launch period. In other words, certain amount of advertising that is done at two weeks before release generates only 89%

of search volume that could have been generated if the same amount of advertising is done at one week before release.

3.4. Estimation

There are several considerations when estimating the model (3-1) – (3-6). First, the model contains time-varying parameters, namely G_{it} and q_{it}. Second, the weekly Google search index is left-truncated. Third, we need to estimate the heterogeneity in (3-6). Finally, the week of first advertising differs movie by movie. The Bayesian DLM method (West and Harrison, 1997) is an excellent framework that addresses all the considerations. As such, we represent the equations (3-1) – (3-6) into a Bayesian DLM framework and estimate the model by a Gibbs sampling method. The log-linearized form of (3-1) is expressed in (3-9).

(3-9) $$\log(S_{it}^{*}+1)=G_{it}+\psi_i\,|t|\cdot I(f_i\le t\le 0)+x_i'\beta+v_{it}=G_{it}+\underbrace{\begin{pmatrix}|t|\cdot I(f_i\le t\le 0) & x_i'\end{pmatrix}}_{q_{it}'}\underbrace{\begin{pmatrix}\psi_i\\ \beta\end{pmatrix}}_{\kappa_i}+v_{it},$$

where $x_i'=(X_i',D_i')$ and $\beta=\left(\beta_1' \quad \beta_2'\right)'$. Note that what we observe from Google Trends is a left-truncated version of S_{it}^{*} because Google Trends reports the search index only when the corresponding search volume exceeds a certain unknown threshold, τ. Let S_{it} be what we observe from Google Trends. Then it holds that

(3-10) $$S_{it}=\begin{cases}S_{it}^{*} & \text{if } S_{it}^{*}\ge\tau\\ 0 & \text{if } S_{it}^{*}<\tau\end{cases}$$

Equation (3-10) with equations (3-9) and (3-2) – (3-3) can be represented in a state-space framework as follows.

(3-11) $$\underbrace{\log(S_{it}+1)}_{y_{it}} = \begin{cases} \log(S^*_{it}+1) & \text{if } \log(S^*_{it}+1) \geq \log(\tau+1) \\ 0 & \text{if } \log(S^*_{it}+1) < \log(\tau+1) \end{cases}$$

(3-12) $$\underbrace{\log(S^*_{it}+1)}_{y^*_{it}} = \underbrace{[1 \quad 0]}_{F_i'} \underbrace{\begin{bmatrix} G_{it} \\ q_{it} \end{bmatrix}}_{\theta_{it}} + q'_{it}\kappa_i + v_{it}.$$

(3-13) $$\underbrace{\begin{bmatrix} G_{it} \\ q_{it} \end{bmatrix}}_{\theta_{it}} = \underbrace{\begin{bmatrix} 1-\delta_i & g(A_{it}) \\ 0 & \pi(A_{it}) \end{bmatrix}}_{H_{it}} \underbrace{\begin{bmatrix} G_{i,t-1} \\ q_{i,t-1} \end{bmatrix}}_{\theta_{i,t-1}} + \underbrace{\begin{bmatrix} 0 \\ \delta_i I(A_{it}=0) \end{bmatrix}}_{u_{it}} + \underbrace{\begin{bmatrix} w^G_{it} \\ w^q_{it} \end{bmatrix}}_{w_{it}},$$

where $\pi(A_{it}) = 1-\alpha(A_{it}) - \delta_i I(A_{it}=0) = 1-c_i - w_i g(A_{it}) - \delta_i I(A_{it}=0)$.

We substitute the unknown threshold τ with the minimum order statistic of positive S_{it}, i.e., $\hat{\tau} = \min\{S_{it}; S_{it} > 0\}$ which is the maximum likelihood estimator of τ (Zuehlke 2003). By adopting vector-matrix notation and incorporating the left-truncated nature of the search index, (3-11), (3-12), and (3-13) can be compactly represented by (3-11'), (3-12') and (3-13').

(3-11') $$y_{it} = y^*_{it} \cdot I(y^*_{it} \geq \log(\hat{\tau}+1)),$$

(3-12') $y^*_{it} = F_i'\theta_{it} + q'_{it}\kappa_i + v_{it}$, where $v_{it} \sim N(0, V_i)$,

(3-13') $\theta_{it} = H_{it}\theta_{it-1} + u_{it} + w_{it}$, where $w_{it} \sim N(0, W_i)$.

By stacking the above equations across movies and adding the heterogeneity equation (3-6), our DLM is ready to be estimated:

(3-11'') $$y_t = A_t y_t^*,$$

(3-12'') $$y_t^* = F'\theta_t + Q_t\kappa + v_t,$$

(3-13'') $$\theta_t = H_t\theta_{t-1} + u_t + w_t,$$

(3-14'') $$\phi = M\gamma + \xi,$$

where

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix}, A_t = \begin{bmatrix} I(y_{1t}^* \geq \log(\hat{\tau}+1)) & 0 & \cdots & 0 \\ 0 & I(y_{2t}^* \geq \log(\hat{\tau}+1)) & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & I(y_{Nt}^* \geq \log(\hat{\tau}+1)) \end{bmatrix}, y_t^* = \begin{bmatrix} y_{1t}^* \\ y_{2t}^* \\ \vdots \\ y_{Nt}^* \end{bmatrix},$$

$$F' = I_N \otimes F_i', \theta_t = \begin{bmatrix} \theta_{1t} \\ \theta_{2t} \\ \vdots \\ \theta_{Nt} \end{bmatrix}, Q_t = \begin{bmatrix} q_{1t}' & 0 & \cdots & 0 \\ 0 & q_{2t}' & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & q_{Nt}' \end{bmatrix}, \kappa = \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_N \end{bmatrix}, v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Nt} \end{bmatrix},$$

$$H_t = \begin{bmatrix} H_{1t} & 0 & \cdots & 0 \\ 0 & H_{2t} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & H_{Nt} \end{bmatrix}, u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Nt} \end{bmatrix}, w_t = \begin{bmatrix} w_{1t} \\ w_{2t} \\ \vdots \\ w_{Nt} \end{bmatrix},$$

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix}, M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_N \end{bmatrix}, \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix}, v_t \sim N(0, V \cdot I_N), w_t \sim N(0, I_N \otimes W), \text{ and } \xi \sim N(0, I_N \otimes \Sigma_\xi).$$

Note that the DLM is constructed assuming that all movies in the data set starts its first advertising all in the same week. This is not true as some movies start first advertising earlier than others. This unbalanced panel structure is considered during the estimation by modifying the matrix H_{it} and the vectors F_i and u_{it} as follows. Before the first advertising of movie i, H_{it} is replaced by the identity matrix of size two, and F_i and u_{it} are replaced by the zero column vector of size two. This modification represents that, when time t is before the first advertising of movie i (i.e., $t < f_i$), only random errors influence advertising goodwill (that is net of WTS) and advertising effectiveness. That is, $\theta_{it} = \theta_{i,t-1} + w_{it}$. Also, this modification implies that the search volume of a movie before its first advertising is accounted for only by its attributes ($y_{it}^* = x_i'\beta + v_{it}$). When estimating the DLM, we use these modified versions of H_{it}, F_i and u_{it}.

The model is estimated by a Gibbs sampler that embeds the forward-filtering/backward-sampling (FF/BS) step to estimate the time-varying paper θ_{it} (West and Harrison 1997). The Gibbs sampler also embeds the MCMC step proposed by Chib (1992) to augment the censored latent variable $\left\{y_{it}^*; y_{it}^* < \log(\hat{\tau}+1)\right\}$ from the truncated normal distribution with support $(-\infty, \log(\hat{\tau}+1))$. In the following subsection, we briefly describe the MCMC step for our Tobit model. The complete algorithm for our final DLM is described in subsection 3.8.

3.4.1. Data-Augmentation Step for Truncated Search Volume

The truncated data can be simulated by a Gibbs sampler proposed by Chib (1992). It applies the data augmentation technique by Tanner and Wong (1987). In what follows, we introduce the data augmentation step of Chib (1992). Let y_1 be the nonzero observations of y_{it} and let y_0 be the censored observations—i.e., every element of vector y_0 is zero. Suppose that along with the censored observations y_0, corresponding latent data z is available. By definition, all elements of z should be smaller than the threshold $\log(\hat{\tau}+1)$. Then, it must hold that the joint posterior distribution of parameters conditional on y_0, y_1, and z is equivalent to the joint posterior distribution of the parameters conditional only on y_1 and z. That is, parameters | y_0, y_1, z ~ parameters | y_1, z, where ~ denotes equality in distribution. Although z is not available, we can simulate it from its distribution. The distribution of z_{it}, an element of z, is a truncated normal distribution with support $(-\infty, \log(\hat{\tau}+1))$ and conditional pdf

$$(3\text{-}15)\quad f(z_{it} \mid y_{it}, F_i'\theta_{it} + q_{it}'\kappa_i, V_i) = \frac{1}{\sqrt{V_i}}\phi\left(\frac{z_{it} - F_i'\theta_{it} - q_{it}'\kappa_i}{\sqrt{V_i}}\right) \Big/ \Phi\left(\frac{\log(\hat{\tau}+1) - F_i'\theta_{it} - q_{it}'\kappa_i}{\sqrt{V_i}}\right),$$

$$-\infty < z_{it} < \log(\hat{\tau}+1), \quad (i,t) \in \Gamma,$$

where $\Gamma = \{(i, t); S_{it} = 0\}$, the set of indexes of movies and weeks for which Google search index is zero. In each iteration of the Gibbs sampler, we draw random numbers from (3-15) and replace the truncated data points in the original data set. Based on this augmented data set, the model parameters are estimated. This process is iterated sufficiently many times to guarantee the convergence of the Markov chain.

3.5. Empirical Analysis

The model (3-11’’) – (3-14’’) and (3-15) is applied to our movie data set. For numerical stability, we scale down the search volume measure by 10^3. That is, y_{it} in (3-11) is $\log((S_{it} + 1) / 1000)$. The threshold of the left-truncated search volume is scaled accordingly.

3.5.1. Model Comparison

We compare the performance of the proposed model with alternative models. Model 1 assumes that the advertising effectiveness is constant over time. That is, $q_{it} = q_i$ for all t. Model 2 removes the time-to-launch effect, namely $\psi_i = 0$ in Model 2. This is a heterogeneity-extended version of Naik, Mantrala and Sawyer (1998). Model 3 assumes that the time-invariant parameters are homogenous across movies: i.e., $\phi_i = \phi$ for all i in Model 3. Model 4 does not consider the left-truncated nature of the Google search index. Model 5 and Model 6 are the proposed models. Model 6 relates the heterogeneity of time-invariant parameters with movie characteristics while Model 5 relates the heterogeneity with a common intercept for each parameter. Table 3.4 summarizes the performance of the models in terms of the negative log-likelihood value. The log-likelihood of DLM measures the one-week ahead forecasting performance (West and Harrison 1997, pp. 326–329). Note that the proposed model outperforms alternative models in terms of one-week ahead forecasting accuracy of online search volume.

Table 3.4: Model Comparison

Model	Model Description				-Log-like
	Ad effectiveness	Willingness-to-search	Heterogeneity	Left-censored search index	
M1	Constant	Yes	Yes	Modeled	3970.90
M2	Time-varying	No	Yes	Modeled	4933.80
M3	Time-varying	Yes	No	Modeled	3254.70
M4	Time-varying	Yes	Yes	Not modeled	3892.60
M5	Time-varying	Yes	Intercept only	Modeled	3050.90
M6	Time-varying	Yes	Movie characteristics	Modeled	3106.90

Several things are worthy of attention. First, comparing Model 1 and the propose model reveals that advertising effectiveness is time-varying over the course of movie life cycle. Second, the inferior performance of Model 2 implies that incorporating WTS is critical. Especially, note that Model 2 is the worst, suggesting that WTS is the most important element in modeling the pre-launch online search volume. Third, comparing Model 3 and the proposed model reveals that advertising effectiveness and WTS effect are substantially heterogeneous across movies. Lastly, the inferior performance of Model 4 shows that modeling the left-truncated nature of weekly Google search index is important.

3.5.2. Parameter Estimates

Table 3.5 reports the posterior medians and 95% highest probability density interval (HPDI) of the parameters in Model 5. Table 3.6 reports the posterior medians and 95% HPDI of the estimates of parameter vector γ of (3-6) in Model 6, namely, the impacts of movie characteristics on the time-to-launch effect and advertising schedule effect.

Table 3.5: Parameter Estimates

Covariate	Median Estimate	2.5th HPDI	97.5th HPDI
Time-to-launch effect, ψ	-0.48	-0.44	-0.41
Carryover rate, $1 - \delta$	0.94	0.94	0.94
Deterioration of ad effectiveness due to time passage, c	0.75	0.79	0.83
Effect of the previous week's advertising, w	0.07	0.08	0.08
Effect of movie characteristics on search volume, β			
Sum of past revenues of the director	0.14	0.18	0.22
Average rating of past movies of the director	0.16	0.42	0.67
Standard deviation of ratings of past movies of the director	1.73	2.32	2.89
Genre: Action	0.50	0.69	0.88
Genre: Comedy	-0.60	-0.41	-0.22
Genre: Drama	0.28	0.49	0.70
MPAA rating: R	-2.72	-2.18	-1.65
MPAA rating: PG	-2.86	-2.32	-1.79
MPAA rating: PG13	-2.66	-2.13	-1.61
Sequel	2.31	2.55	2.79
Post-launch week effect			
One week after launch	0.80	1.00	1.21
Two weeks after launch	0.51	0.71	0.92
Three weeks after launch	0.32	0.52	0.73
Four weeks after launch	0.17	0.37	0.58
Five weeks after launch	0.04	0.25	0.45
Observation variance, V	1.05	1.11	1.18
System variance on goodwill, W^G	3.61e-10	1.01e-09	1.10e-06
System variance on ad effectiveness, W^q	1.06e-11	4.77e-11	5.64e-08
Heterogeneity variance, Σ_ξ			
Variance for ψ	2.06e-2	2.67e-2	3.54e-2
Variance for δ	3.70e-4	4.78e-4	6.33e-4
Variance for c	3.42e-2	4.43e-2	5.88e-2
Variance for w	2.24e-4	2.90e-4	3.83e-4

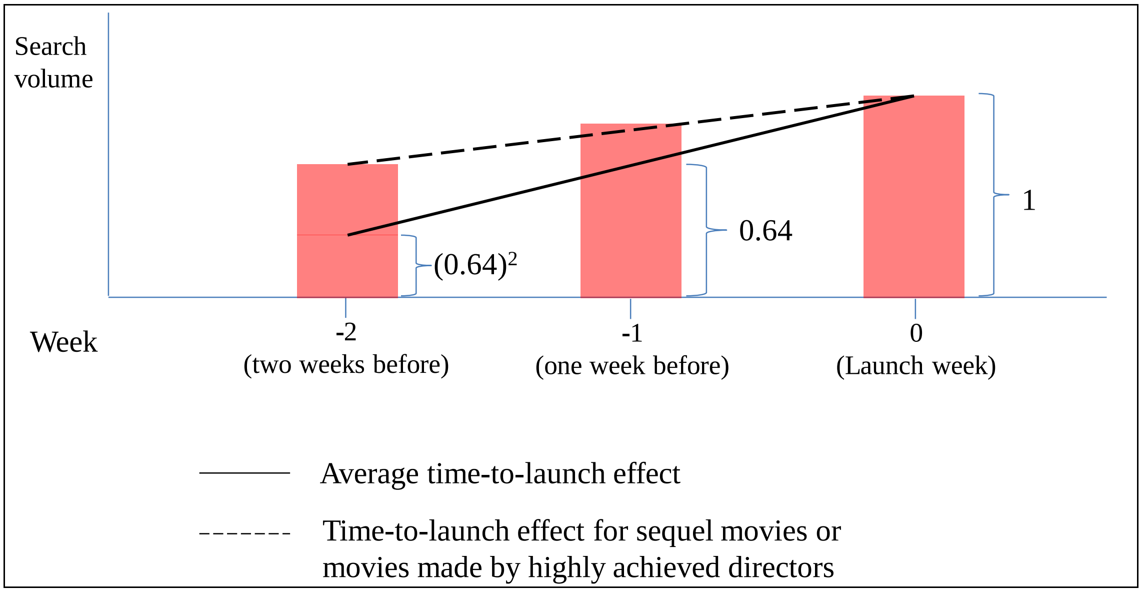
Table 3.6: Effects of Movie Characteristics on Heterogeneous Ad Effectiveness

Movie Characteristics	Median Estimate	2.5th HPDI	97.5th HPDI
Effect of movie characteristics on Time-to-launch effect (ψ)			
Intercept	-2.09	-1.87	-1.65
Sum of past revenues of the director	0.03	0.04	0.04
Average rating of past movies of the director	0.72	0.81	0.91
Standard deviation of ratings of past movies of the director	-0.11	-0.03	0.06
Genre: Action	0.19	0.21	0.23
Genre: Comedy	0.11	0.13	0.15
Genre: Drama	0.25	0.27	0.29
MPAA rating: R	-0.15	-0.10	-0.05
MPAA rating: PG	-0.05	0.00	0.05
MPAA rating: PG13	-0.18	-0.13	-0.08
Sequel	0.03	0.05	0.07
Effect of movie characteristics on 1 minus carryover rate (i.e., δ)			
Intercept	0.23	0.31	0.39
Sum of past revenues of the director	0.00	0.00	0.00
Average rating of past movies of the director	-0.14	-0.10	-0.07
Standard deviation of ratings of past movies of the director	0.08	0.11	0.14
Genre: Action	-0.03	-0.02	-0.01
Genre: Comedy	-0.03	-0.03	-0.02
Genre: Drama	-0.07	-0.06	-0.06
MPAA rating: R	-0.15	-0.13	-0.11
MPAA rating: PG	-0.16	-0.14	-0.12
MPAA rating: PG13	-0.14	-0.12	-0.10
Sequel	0.02	0.03	0.04
Effect of movie characteristics on c			
Intercept	-1.42	-0.42	0.56
Sum of past revenues of the director	0.02	0.03	0.05
Average rating of past movies of the director	0.25	0.67	1.09
Standard deviation of ratings of past movies of the director	0.48	0.86	1.25
Genre: Action	-0.03	0.05	0.13
Genre: Comedy	0.03	0.10	0.18
Genre: Drama	0.00	0.09	0.17
MPAA rating: R	-0.83	-0.62	-0.40
MPAA rating: PG	-0.73	-0.52	-0.31
MPAA rating: PG13	-0.80	-0.59	-0.38
Sequel	-0.60	-0.50	-0.40
Effect of movie characteristics on w			
Intercept	-0.40	-0.32	-0.23
Sum of past revenues of the director	-0.01	-0.01	-0.01
Average rating of past movies of the director	0.10	0.14	0.17
Standard deviation of ratings of past movies of the director	-0.05	-0.02	0.02
Genre: Action	0.02	0.02	0.03
Genre: Comedy	0.00	0.00	0.01
Genre: Drama	-0.02	-0.01	-0.01
MPAA rating: R	0.09	0.11	0.13
MPAA rating: PG	0.11	0.13	0.15
MPAA rating: PG13	0.08	0.10	0.12
Sequel	0.10	0.11	0.11

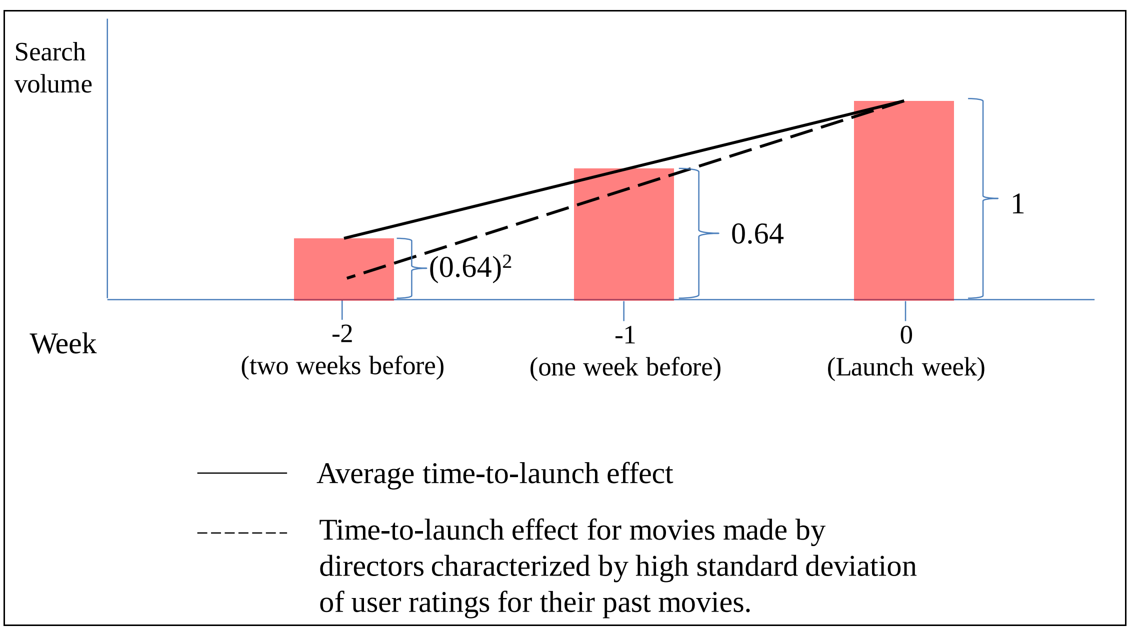
First, the time-to-launch has a negative effect on consumer search intensity. This means that, ceteris paribus, the same advertising spending allocated in a pre-launch week generate less search volume, the further apart the week is from the launch time of a movie in its pre-launch period. The effect size is substantial. On average, the weekly search volume decreases by a factor of 0.64 (= exp(-0.44)) as the advertising is done one-week earlier from the launch week. That is, if a certain amount of ad generates 1,000 search activities in the launch week, the same amount of ad would generate only 640 (= $1,000 \times 0.64$) search activities if the ad were aired one week before the release week—all else being equal.

Table 3.6 shows how movie characteristics influence the time-to-launch effect. Notably, consider the past performance of directors and the sequel nature of the movie—i.e., whether the movie is a sequel or not. The more successful history a movie director has in terms of his or her total past revenue and the average past user ratings earned, the less vulnerable the movie is to the time-to-launch effect in attracting consumer interests across the whole pre-launch weeks. In other words, given that the same amount of advertising is executed, consumers are more willing to search for movies made by highly achieved directors than movies made by their less successful peers over the entire pre-launch period. The same pattern is observed for sequel movies versus original movies (Figure 3.2(a)). Perhaps, consumers relate such characteristics to movie quality (Basuroy, Desai and Talukdar 2006). As movie quality is associated with the benefit of movie consumption, movies with such characteristics enjoy more consumer interests (Beatty and Smith 1987; Punj and Staelin 1983; Srinivasan and Ratchford 1991) and thus more search volume over the entire pre-launch period, compared with movies without such characteristics.

Figure 3.2: Illustration of the Time-to-Launch Effect



(a) Movies made by highly achieved directors



(b) Movies made by directors of high standard deviation of user ratings for their past movies

On the other hand, the standard deviation of past user ratings of the focal movie's director exacerbates the time-to-launch effect (Figure 3.2(b)). Consumers tend to interpret a high standard deviation of director ratings as high uncertainty about the quality of the movies made by the director. While previous research finds that uncertainty increases search activities (Moorthy, Ratchford and Talukdar 1997), our finding implies that the search activities tend to be reserved until the weeks that immediately precede the movie release. For managers, these findings imply that early pre-launch advertising is more effective for high-quality movies (versus low-quality movies) and for movies of low quality uncertainty (versus movies of high quality uncertainty).

Second, consider how advertising effectiveness is influenced by the passage of time since the first advertising. The estimated significant positive effect of time passage on advertising effectiveness indicates that, from the first week of advertising, the advertising effectiveness gradually decays over time. Note that this steady decrease of advertising effectiveness happens regardless of the past advertising intensity or existence of ad hiatus.

Table 3.6 shows that the passage-of-time effect is moderated by movie characteristics. The gradual decay of ad effectiveness due to the passage of time is severer for movies made by highly achieved directors—in terms of past gross revenue and past average user rating—and movies made by directors of controversial past performance—in terms of the standard deviation of past user ratings. The advertising effectiveness of sequel movies is less vulnerable to the passage of time that that of original movies.

Third, consider how advertising effectiveness is influenced by the past advertising intensity. We find that the advertising effectiveness in the current week is negatively associated with the previous week's advertising intensity. That is, the more advertising budget was allocated in the previous week, the fewer search activities are generated by a given amount of advertising in the current week. The estimation results show that doubling the previous week's advertising spending decreases the current week's advertising effectiveness by 8% of the previous week's advertising effectiveness[5]—all else being equal. Perhaps, the previous week's advertising reduces the number of potential movie consumers that the current week's advertising can influence because some consumers have already searched for the advertised movies as a response to the previous week's advertising.

Table 3.6 shows that movie characteristics moderate the relationship between the previous week's advertising intensity and the current week's advertising effectiveness. Advertising for sequel movies—versus that for original movies—are more susceptible to the negative impact of the previous week's advertising intensity. Movies made by directors who are known for higher user ratings for their past movies are more vulnerable to this substitutive relationship between consecutive weeks' advertising.

Fourth, advertising carryover is within the previously reported range (Bruce 2008), and is influenced by movie characteristics. Notably, movies made by directors whose past movies enjoyed higher user ratings are associated with a higher carryover rate than movies made by directors whose past movies experienced lower user ratings. As carryover rate is related to the lasting effect of past advertising, this finding suggests that such movies—i.e., movies made by directors whose past movies enjoyed higher user ratings—enjoy a longer-lasting effect of advertising than original movies.

---

[5] $\frac{dq_{it}}{dA_{it}/A_{it}} = -w \cdot q_{i,t-1} = -0.08 \cdot q_{i,t-1}$.

Finally, let us consider how overall search volume of a movie is associated with the movie's characteristics as presented in Table 3.5. Because movie characteristics in the main model are included regardless of advertising schedule, the effects of movie characteristics in Table 3.5 show how movie characteristics influence consumers' general tendency to search for movie information across the entire movie life cycle. Consistent with implications from previous research (Beatty and Smith 1987; Punj and Staelin 1983; Srinivasan and Ratchford 1991), we find that consumers' search intensity is higher for movies made by high-profile directors—in terms of their past revenues and user ratings—than those made by less achieved directors. The standard deviation of user ratings of past movies of the focal director is also positively associated with online search volume of the new movie made by the focal director. Consumes may interpret high standard deviation of past user ratings of a director as a signal for high quality uncertainty of new movies that will be made by the director. As such, consumers may tend to search more for the new movies made by such directors to reduce the quality uncertainty about the new movies. Sequel movies tend to receive more search actions from consumers than original movies, because consumers may think that sequel movies are better in quality than original movies (Basuroy, Desai and Talukdar 2006).

3.6. Reallocation of Advertising Budget

Given that the pre-launch search volume of a new movie reflects the pre-launch interest level of consumers for the new movie, studio managers may want to use the estimation results to plan efficient allocations of a given advertising budget across pre-launch weeks. For that matter, first note that diverse factors, including past advertising outlay, time to launch, advertising carryover and the initial ad effectiveness, influence the effect of pre-launch advertising on pre-launch search volume. As such, it is hard, if not impossible, to suggest a general prescription that can be applicable to every movie. However, a few implications can be garnered from the model and estimation results.

First, the large time-to-launch effect implies that, all else being equal, the less is search volume generated by the same amount of advertising, the further the advertising campaign is in advance of release. Therefore, early pre-launch advertising should be avoided if the sole objective of advertising is to generate as much consumer interest as possible.

Second, if early advertising should be conducted for some reason (e.g., signaling to competitors or securing distribution at launch), it is recommended to have hiatus periods (i.e., periods with no advertising) between the early advertising and the main advertising. This is because ad effectiveness restores during the hiatus periods. In fact, restoration of ad effectiveness during ad hiatus period implies that early pre-launch advertising that aims to influence distributors' screen allocation decision is justified if sufficiently long hiatus period exists between the early advertising and the main advertising. The worst pre-launch-advertising schedule, in terms of efficient generation of consumer interest, is to start advertising early in the pre-launch period and never have a hiatus period before the launch week.

Third, the negative effect of past advertising intensity on the current ad effectiveness implies that the same amount of current search volume can be achieved with less advertising spending across weeks. Note that the current search volume is the sum of the search volume that is generated by the current advertising and a part of previous search volume that carries over from the previous period. As the two components of current search volume have a substitutive relationship, it is possible to reduce the advertising spending across entire weeks and maintain the same level of search volume in the final week.

Fourth, because willingness-to-search effect varies by movie characteristics, early pre-launch advertising can be more effective for a certain type of movies than others. Specifically, early pre-launch advertising can more efficiently generate consumer search for sequel movies and movies made by highly achieved directors.

More formally, we can set up a constrained nonlinear optimization problem to suggest efficient advertising schedules for a given advertising budget. The parameter estimates in the previous section are used to define the relationship between an advertising schedule and the search volume generated by the schedule. Our objective is to generate as much search volume as possible in the launch week through an efficient allocation of a given budget across pre-launch weeks. Let $E(S_{i0})$ be the expected search volume of movie i in the launch week. Then, the optimization problem is stated in (3-16).

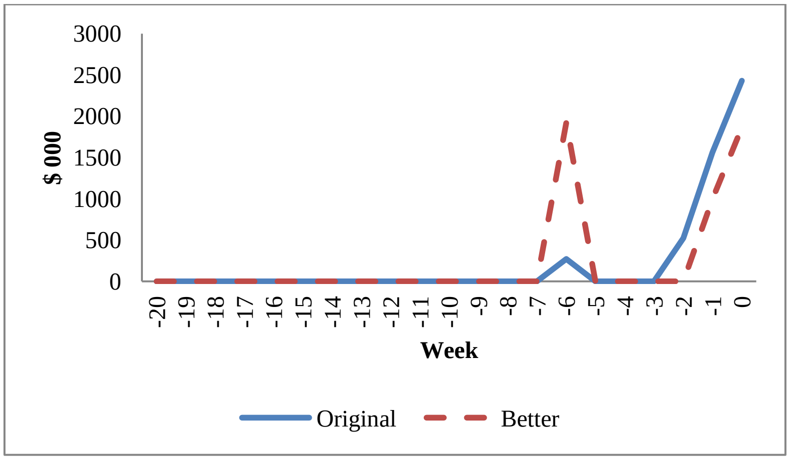
$$
\begin{aligned}
&\max_{A_{it}} \quad E(S_{i0}) \\
&\text{s.t.} \\
&\qquad \sum_t A_{it} \le b_i, \qquad A_{it} \ge 0, \\
&\text{where} \\
&\qquad E(S_{it}) = e^{G_{it}} \lambda_{it}, \\
&\qquad G_{it} = (1-\delta_i) G_{i,t-1} + q_{i,t-1} \log(A_{it}+1), \\
&\qquad q_{it} = (1 - c_i - w_i \log(A_{it}+1)) q_{i,t-1} + \delta_i I(A_{it}=0)(1-q_{i,t-1}),
\end{aligned}
\tag{3-16}
$$

A_{it} is the advertising budget of movie i at the pre-launch week t, and b_i is the pre-launch advertising budget allowed to movie i. Note that the full managerial problem is more complex than (3-16) because we need to determine when to start the first advertising, when to temporarily stop the advertising and how much to allocate in each period.[6] Also, more constraints may be added in the budget allocation process. For example, there may be a limit on the media space (e.g., airtime on TV) that a firm can purchase during a unit time interval. Such constraints are not considered in (3-16).

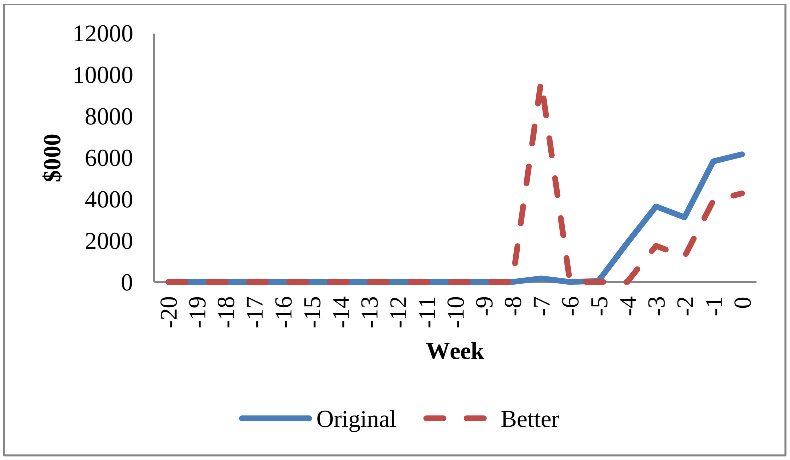
The solution to (3-16) is not necessarily the global optimum as the objective function is not globally concave. However, a solution to (3-16) can suggest an improvement over the original advertising schedule used by management. We solve (3-16) for three selected movies: *Gran Torino*, *Paul Blart: Mall Cop* and *Twilight*. We use each movie's original advertising schedule as the initial solution for the optimization problem. The original advertising schedules of these movies are characterized by small early advertising spending, ensuing ad hiatus, and then a main advertising campaign that follows the hiatus. Figure 3.3 compares the original advertising schedule with a superior one for each of the three movies. The two schedules execute the same amount of pre-launch advertising budget but the superior allocation schedules generate more cumulated search volume in the launch week by 2,668%, 1,504% and 226%, respectively, for the three movies.

[6] For the complexity of the problem, see Naik, Mantrala, and Sawyer (1998) pp. 228-229.

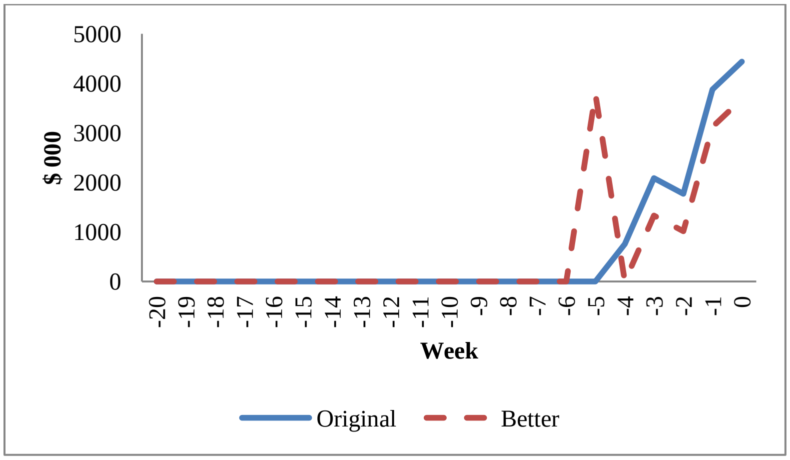
Figure 3.3: The Original Advertising Schedule and a Superior Schedule



(a) Gran Torino



(b) Paul Blart: Mall Cop



(c) Twilight

A common feature of the three movies is that their initial advertising effectiveness is high. *Gran Torino*'s initial advertising effectiveness is 0.48 and those of *Paul Blart: Mall Cop* and *Twilight* are 0.51 and 0.27, respectively. The high initial advertising effectiveness justifies high spending in the early weeks to generate high early search volume. It also contributes to high search volume in the subsequent weeks through the carryover effect of initial advertising. However, high early spending hurts the advertising effectiveness in the later periods. As such, the solution recommends a hiatus period after the initial large spending to restore the advertising effectiveness. The restored effectiveness will work to generate search volume efficiently in later periods. Note that, in the weeks that immediately precede the launch time, the suggested allocation schedules spend less than the original schedules.

3.7. Conclusions

This study examines how advertising generates consumer interest for new products in their pre-launch periods. To this end, we develop a model that relates the advertising schedule to Google keyword search volume. The two pillars of the model are the time-varying advertising effectiveness and consumers' changing willingness-to-search. We assume that the advertising effectiveness is influenced by specifics of the advertising schedule such as the passage of time since the first execution of advertising, the previous period's advertising intensity and the existence of ad hiatus periods. The willingness-to-search is specified by the time-to-launch and product attributes. We apply the model to a movie data set that consists of weekly advertising spending, weekly search volume and movie characteristics. The final model specification for estimation incorporates the unbalanced structure of the panel data and the left-censored feature of the Google search index. The model is estimated in the Bayesian DLM framework. The estimated parameters are input to a constrained nonlinear optimization problem to find a more efficient advertising schedule during the pre-launch period.

We obtain several interesting findings from the empirical analysis. First, keyword search volume in the pre-launch period is substantially influenced by the time-to-launch effect. Second, the advertising effectiveness for generating keyword search volume varies over time as a function of the advertising schedule. Third, the time-to-launch effect and time-varying advertising effectiveness are significantly influenced by movie characteristics. Thus, a same advertising schedule for different movies generates different search volume processes.

We suggest new data to measure the pre-launch advertising effectiveness. The Google search index we use is readily available public data that is observed before a new product is launched. Compared with conventional data such as consumer awareness and purchase intent, an online search index does not require a consumer survey and can be traced for an extended pre-launch period. Compared with the virtual stock price such as the one on the Hollywood Stock Exchange, the online search volume is not restricted to the movie industry; it can be collected for any product, event or human. Also, as the superior predictive ability of online search volume suggests, online search volume is a good reflector of consumer interest.

There are promising future research directions. First, while online search volume represents the level of awareness among the consumer population, it does not indicate the valence of consumer opinion. If a product is experiencing negative word-of-mouth versus positive word-of-mouth, higher search volume for the product may not necessarily mean higher sales of the product. In other words, the online search volume may interact with the prevailing valence of word-of-mouth to influence sales. Therefore, examining the interaction between search volume and word-of-mouth valence may reveal interesting insights into the relationship between search volume and sales. Second, pre-launch consumer interest will be converted into sales when the new products become available. Whether the pre-launch advertising has a permanent versus temporary effect on sales is an important question. The pre-launch search volume may be used to answer this question.

3.8. Appendix to Chapter 3

3.8.1. The Dynamic Linear Model

The model we estimate is specified by equations (3-11’’) – (3-14’’).

(3-11’’) $$y_t = A_t y_t^*,$$

(3-12’’) $$y_t^* = F'\theta_t + Q_t\kappa + v_t,$$

(3-13’’) $$\theta_t = H_t\theta_{t-1} + u_t + w_t,$$

(3-14’’) $$\phi = M\gamma + \xi,$$

where

$$y_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix}, A_t = \begin{bmatrix} I(y_{1t}^* \geq \log(\hat{\tau}+1)) & 0 & \cdots & 0 \\ 0 & I(y_{2t}^* \geq \log(\hat{\tau}+1)) & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & I(y_{Nt}^* \geq \log(\hat{\tau}+1)) \end{bmatrix}, y_t^* = \begin{bmatrix} y_{1t}^* \\ y_{2t}^* \\ \vdots \\ y_{Nt}^* \end{bmatrix},$$

$$F' = I_N \otimes F_i', \theta_t = \begin{bmatrix} \theta_{1t} \\ \theta_{2t} \\ \vdots \\ \theta_{Nt} \end{bmatrix}, Q_t = \begin{bmatrix} q_{1t}' & 0 & \cdots & 0 \\ 0 & q_{2t}' & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & q_{Nt}' \end{bmatrix}, \kappa = \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \vdots \\ \kappa_N \end{bmatrix}, v_t = \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Nt} \end{bmatrix},$$

$$H_t = \begin{bmatrix} H_{1t} & 0 & \cdots & 0 \\ 0 & H_{2t} & \cdots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & \cdots & H_{Nt} \end{bmatrix}, u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{Nt} \end{bmatrix}, w_t = \begin{bmatrix} w_{1t} \\ w_{2t} \\ \vdots \\ w_{Nt} \end{bmatrix},$$

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_N \end{bmatrix}, M = \begin{bmatrix} M_1 \\ M_2 \\ \vdots \\ M_N \end{bmatrix}, \xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix}, \ v_t \sim N(0, V \cdot I_N), w_t \sim N(0, I_N \otimes W), \text{ and } \xi \sim N(0, I_N \otimes \Sigma_\xi).$$

Let us define additional variables for clarity.

i: index for movie.

t: index for week.

t_i^0: index for the week when the first execution of movie i's advertising campaign starts.

N: number of movies.

T: number of time points (weeks)

T_i : number of observations of movie i since its first advertising. $T_i = T - t_i^0 + 1$.

$\Gamma = \{(i, t); S_{it} = 0\}$: the set of index of movies and weeks whose Google search index is zero.

$y_{it} = \log(S_{it}+1)$: log of observed search volume. Note that $y_{it} = 0$ if $S_{it} = 0$.

$y_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$: vector of log observed search volume time-series of movie i.

$y = (y_1', y_2', \ldots, y_N')'$: vector of log observed search volume of all movies.

n_0: the number of zero search volume across all movies and time.

$n_1 = N*T - n_0$: the number of nonzero search volume across all movies and time.

y_0: vector of zero search volume of size n_0. $y_0 = \{y_{it}, (i,t) \in \Gamma\}$.

y_1: vector of nonzero search volume of size n_1. $y_1 = \{y_{it}, (i,t) \notin \Gamma\}$. $y = \{y_1, y_0\}$.

$z = \{z_{it}; (i,t) \in \Gamma\}$: vector of corresponding simulated latent data for y_0.

$y^{(z)}$: vector of log search volume with y_0 replaced by z. $y^{(z)} = \{y_1, z\}$.

3.8.2. The Gibbs Sampling Algorithm

1) Draw z_{it}

For each $(i,\ t) \in \Gamma$,

$z_{it} \mid y_{it}, F_i'\theta_{it} + q_{it}'\kappa_i, V \sim TN(F_i'\theta_{it} + q_{it}'\kappa_i, V) \cdot I(z_{it} < \log(\hat{\tau}+1))$, where TN denotes the truncated normal distribution. $I(\cdot)$ is an indicator function.

Truncated Normal pdf for z_{it}:

$$f(z_{it} \mid y_{it}, F_i'\theta_{it} + q_{it}'\kappa_i, V) = \frac{1}{\sqrt{V}} \phi\left(\frac{z_{it} - F_i'\theta_{it} - q_{it}'\kappa_i}{\sqrt{V}}\right) \Big/ \Phi\left(\frac{\log(\hat{\tau}+1) - F_i'\theta_{it} - q_{it}'\kappa_i}{\sqrt{V}}\right),$$

where $-\infty < z_{it} < \log(\hat{\tau}+1), \quad (i,t) \in \Gamma,$

z_{it} drawn from the above distribution replaces the corresponding element of y_0 to create $y^{(z)}$.

2) Draw θ_t

We use the forward-filtering/backward-sampling steps to draw the time-varying parameter θ_t.

- Forward filtering

(a) Posterior at t-1: $(\theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1})$.

(b) Prior at t: $(\theta_t \mid D_{t-1}) \sim N(a_t, R_t)$, where $a_t = H_t m_{t-1} + u_t$, $R_t = H_t C_{t-1} H_t' + W$.

(c) One-step ahead forecast of $\tilde{y}_t^{(z)} = y_t^{(z)} - Q_t \kappa$: $\tilde{y}_t^{(z)} \mid D_{t-1} \sim N(f_t, B_t)$, where

$f_t = F' a_t$, $B_t = F' R_t F + V$.

(d) Posterior at t: $(\theta_t \mid D_t) \sim N(m_t, C_t)$, where $m_t = a_t + R_t F B_t^{-1}(\tilde{y}_t^{(z)} - f_t)$, $C_t = R_t - R_t F B_t^{-1} F' R_t$.

- Backward sampling

at t=T: $(\theta_T \mid D_T) \sim N(m_T, C_T)$.

at t=T-1, …, 0: $(\theta_t \mid \theta_{t+1}, D_t) \sim N(g_t, K_t)$, where

$g_t = m_t + C_t H'_{t+1} R^{-1}_{t+1} (\theta_{t+1} - a_{t+1}),\ K_t = C_t - C_t H'_{t+1} R^{-1}_{t+1} H_{t+1} C_t$.

In each iteration, the initial values m_0 is updated by g_0.

3) Draw $\beta = (\beta'_1 \quad \beta'_2)'$

- Regression equation: $y^{(z)}_{it} - F'_i \theta_{it} - \psi_i \mid t \mid = x'_{it} \beta + v_{it}$

By stacking the above equation across time for movie i, we get Y = Xβ + ε. Then,

- Likelihood: $\beta \sim N\left(\left(X' \left(Cov(\varepsilon) \right)^{-1} X \right)^{-1} X' \left(Cov(\varepsilon) \right)^{-1} Y,\quad \left(X' \left(Cov(\varepsilon) \right)^{-1} X \right)^{-1} \right) \equiv N(d_\beta, S_\beta)$.

- Prior distribution: $\beta \sim N(\alpha, \Sigma_\beta)$.

- Posterior distribution: $\beta \sim N\left((\Sigma_\beta^{-1} + S_{\beta_i}{}^{-1})^{-1} (\Sigma_\beta^{-1} \alpha + S_\beta^{-1} d_\beta),\quad (\Sigma_\beta^{-1} + S_\beta^{-1})^{-1} \right)$.

4) Draw $\phi_i = (\psi_i, \delta_i, c_i, w_i)'$

Note that the data range varies movie by movie, because the first execution time of advertising campaign differs movie by movie. The data before the first advertising of movie i is discarded to estimate ϕ_i because these parameters are relevant only after the first advertising of movie i started.

- Regression equation:

$$
\begin{aligned}
& y_{it}^{(z)} - F_i'\theta_{it} - x_{it}'\beta = \psi_i \,|\, t \,| + v_{it}, \\
& \Delta G_{it} - q_{i,t-1} g(A_{it}) = -\delta_i G_{i,t-1} + w_{it}^{G}, \\
& \Delta q_{it} - \delta_i I(A_{it} = 0)(1 - q_{i,t-1}) = -c_i q_{i,t-1} - w_i g(A_{it}) q_{i,t-1} + w_{it}^{q} \\
& \qquad = \underbrace{\begin{bmatrix} -q_{i,t-1} & -g(A_{it}) q_{i,t-1} \end{bmatrix}}_{p_{i,t-1}} \begin{bmatrix} c_i & w_i \end{bmatrix}' + w_{it}^{q}.
\end{aligned}
$$

Let t_i^0 be the week of the first pre-launch advertising of movie i. For each movie, stack the above equations across time.

$$
\underbrace{\begin{bmatrix}
y_{i,t_i^0}^{(z)} - F_1'\theta_{i,t_i^0} - x_{i,t_i^0}'\beta \\
\vdots \\
y_{iT}^{(z)} - F_1'\theta_{iT} - x_{iT}'\beta \\
\Delta G_{i,t_i^0} - q_{i,t_i^0-1} g(A_{i,t_i^0}) \\
\vdots \\
\Delta G_{iT} - q_{iT-1} g(A_{iT}) \\
\Delta q_{i,t_i^0} - \delta_i I(A_{i,t_i^0} = 0)(1 - q_{i,t_i^0-1}) \\
\vdots \\
\Delta q_{iT} - \delta_i I(A_{iT} = 0)(1 - q_{i,T-1})
\end{bmatrix}}_{Y_i}
=
\underbrace{\begin{bmatrix}
|t_i^0| & & \\
\vdots & & \\
|0| & & \\
 & -G_{i,t_i^0-1} & \\
 & \vdots & \\
 & -G_{i,T-1} & \\
 & & p_{i,t_i^0-1} \\
 & & \vdots \\
 & & p_{i,t_i^0-1}
\end{bmatrix}}_{X_i}
\phi_i +
\underbrace{\begin{bmatrix}
v_{i,t_i^0} \\ \vdots \\ v_{iT} \\ w_{i,t_i^0}^{G} \\ \vdots \\ w_{iT}^{G} \\ w_{i,t_i^0}^{q} \\ \vdots \\ w_{iT}^{q}
\end{bmatrix}}_{\varepsilon_i}
, \text{ or } Y_i = X_i\phi_i + \varepsilon_i .
$$

Then,

- Likelihood:

$$\phi_i \sim N\left(\left(X_i' \left(\mathrm{Cov}(\varepsilon_i) \right)^{-1} X_i \right)^{-1} X_i' \left(\mathrm{Cov}(\varepsilon_i) \right)^{-1} Y_i, \quad \left(X_i' \left(\mathrm{Cov}(\varepsilon_i) \right)^{-1} X_i \right)^{-1} \right) \equiv N(d_{\phi_i}, S_{\phi_i}).$$

- Prior distribution: $\phi_i \sim N(M_i\gamma, \Sigma_\xi)$ as in (3-6).

- Posterior distribution: $\phi_i \sim N\left((\Sigma_\xi^{-1} + S_{\phi_i}^{-1})^{-1} (\Sigma_\xi^{-1} M_i\gamma + S_{\phi_i}^{-1} d_{\phi_i}), \quad (\Sigma_\xi^{-1} + S_{\phi_i}^{-1})^{-1} \right)$.

5) Draw γ

- Regression equation: $\phi_i = M_i\gamma + \xi_i$ as in (3-6).

By rearranging the regression equation we get $\phi^{(k)} = M^{(k)}\gamma^{(k)} + \xi^{(k)}$ where $\phi^{(k)}$ is the vector of k^{th} parameter of size ($N \times 1$). By stacking this equation over k (and by abusing the notations), we get $\phi = \left(I_4 \otimes M\right)\gamma + \xi$. Then,

- Likelihood:

$$\gamma \sim N\left(\left(\left(I_4 \otimes M\right)' \Xi^{-1}\left(I_4 \otimes M\right)\right)^{-1}\left(I_4 \otimes M\right)'_k \Xi^{-1}\phi, \quad \left(\left(I_4 \otimes M\right)' \Xi^{-1}\left(I_4 \otimes M\right)\right)^{-1}\right) \equiv N(d_\gamma, S_\gamma),$$

where Ξ is the corresponding covariance matrix of the error term.

- Prior: $\gamma \sim N(\gamma_0, \Sigma_\gamma)$.

- Posterior: $\gamma \sim N\left((\Sigma_\gamma^{-1} + S_\gamma^{-1})^{-1}(\Sigma_\gamma^{-1}\gamma_0 + S_\gamma^{-1}d_\gamma), \quad (\Sigma_\gamma^{-1} + S_\gamma^{-1})^{-1}\right)$.

6) Draw Σ_ξ

- Regression equation: $\phi_i = M_i\gamma + \xi_i$ as in (3-6).

- Prior: $\Sigma_\xi \sim IW(v_\xi, S_\xi)$.

- Posterior: $\Sigma_\xi \sim IW\left(v_\xi + N, \quad S_\xi + \sum_{t=1}^{N}(\phi_i - M_i\gamma)(\phi_i - M_i\gamma)'\right)$

7) Draw V

- Regression equation: $y_{it}^{(z)} = F_i'\theta_{it} + q_{it}'\kappa_i + v_{it}$ for all i and t.

- Prior: $V \sim IG(v_V/2, S_V/2)$.

- Posterior: $V \sim IG\left(\left(v_v + NT\right)/2, \quad \left(S_v + \sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}^{(z)} - F_i'\theta_{it} - q_{it}'\kappa_i)(y_{it}^{(z)} - F_i'\theta_{it} - q_{it}'\kappa_i)'\right)\Big/2\right)$.

8) Draw W

Let $W_i = \text{var}(w_{it}) = \begin{bmatrix} W_i^G & r_i\sqrt{W_i^G W_i^q} \\ r_i\sqrt{W_i^G W_i^q} & W_i^q \end{bmatrix}$ and $W_i^G = \text{var}(\omega_{it}^G)$ and $W_i^q = \text{var}(\omega_{it}^q)$.

Stack the equation $\theta_{it} = H_{it}\theta_{it-1} + u_{it} + w_{it}$ across movies and time. Then for each movie, discard the data points that belong to the time periods where the advertising of the movie has not yet started. Then,

- Regression equation: $\theta_{it} = H_{it}\theta_{it-1} + u_{it} + w_{it}$ for i = 1, …, N and t = t_i^0, …, T.
- Prior distribution: $W \sim IW(v_w, S_w)$.
- Posterior distribution:

$$W \sim IW\left(v_W + \sum_{i=1}^{N} T_i,\quad S_W + \sum_{i=1}^{N}\sum_{t=t_i^0}^{T}(\theta_{it} - H_{it}\theta_{it-1} - u_{it})(\theta_{it} - H_{it}\theta_{it-1} - u_{it})'\right)\text{, where}$$

$T_i = T - t_i^0 + 1$, the number of time points of movie i.

3.9. References

Bass, Frank M., Norris Bruce, Sumit Majumdar, and B. P. S. Murthi (2007), “Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of Advertising-Sales Relationship,” Marketing Science, 26 (2), 179-95.

Basuroy, Suman, Kalpesh Kaushik Desai, and Debabrata Talukdar (2006), “An Empirical Investigation of Signaling in the Motion Picture Industry,” Journal of Marketing Research, 43 (May), 287-95.

Beatty, Sharon E. and Scott M. Smith (1987), “External Search Effort: An Investigation across Several Product Categories,” Journal of Consumer Research, 14 (1), 83-95.

Bruce, Norris I. (2008), “Pooling and Dynamic Forgetting Effects in Multitheme Advertising: Tracking the Advertising Sales Relationship with Particle Filters,” Marketing Science, 27 (4), 659-73.

Bruce, Norris I., Natasha Zhang, and Ceren Kolsarici (2012), “Dynamic Effectiveness of Advertising and Word of Mouth in Sequential Distribution of New Products,” Journal of Marketing Research, 49 (August), 469-86.

Calantone, Roger J and Alan G. Sawyer (1978), “The Stability of Benefit Segments,” Journal of Marketing Research, 15 (August), 395-404.

Chib, Siddhartha (1992), “Bayes Inference in the Tobit Censored Regression Model,” Journal of Econometrics, 51, 79-99.

Grass, Robert C. and Wallace H. Wallace (1969), “Satiation Effects of TV Commercials,” Journal of Advertising Research, 9 (September), 3-8.

Kim (2013), “Is Search Volume a Good Market Predictor? Product Quality and Predictive Performance of Search Volume,” Available at SSRN: http://ssrn.com/abstract=2209537 or http://dx.doi.org/10.2139/ssrn.2209537.

McKinsey Quarterly (2012), “Demystifying Social Media,” April 2012, Available at http://bit.ly/HRe8jR.

Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), “Consumer Information Search Revisited: Theory and Empirical Analysis,” Journal of Consumer Research, 23 (4), 263-77.

Naik, Prasad, A., Murali K. Mantrala, and Alan G. Sawyer (1998), “Planning Media Schedules in the Presence of Dynamic Advertising Quality,” Marketing Science, 17 (3), 214-35.

Onishi, Hiroshi and Puneet Manchanda (2012), "Marketing Activity, Blogging, and Sales," International Journal of Research in Marketing, 29 (3), 221-34.

Pechmann, Cornelia and David Stewart (1990), "Advertising Repetition: A Critical Review of Wearin and Wearout," MSI Report, No. 90-106, Cambridge, MA.

Punj, Girish N. and Richard Staelin (1983), "A Model of Consumer Information Search Behavior for New Automobiles," Journal of Consumer Research, 19 (March), 366-80.

Srinivasan, Narasimhan and Brian T. Ratchford (1991), "An Empirical Test of a Model of External Search for Automobiles," Journal of Consumer Research, 18 (September), 233-42.

Tanner, Martin A and Wing-Hung Wong (1987), "The Calculation of Posterior Distribution by Data Augmentation (with discussion)," Journal of the American Statistical Association, 82, 528-50.

Trope, Yaacov and Nira Liberman (2003), "Temporal Construal," Psychological Review, 110 (3), 403-421.

Weitzman, Martin L. (1979), "Optimal Search for The Best Alternative," Econometrica, 47 (3), 641-54.

West, Mike and Jeff Harrison (1997), Bayesian Forecasting and Dynamic Models, Springer-Verlag, New York.

Zuehlke, Thomas W. (2003), "Estimation of A Tobit Model with Unknown Censoring Threshold," Applied Economcs, 35, 1163-69.

4. When Is Search Volume a Good Predictor of Demand? Product Quality and the Predictive Performance of Search Volume

4.1. Introduction

This study aims to provide an explanation on the varying performance of online search volume in predicting new-product demand. While early studies in this area report that collective online activity of consumers provides excellent predictive ability for market outcomes (Goel et. al., 2010; Kulkarni, Kannan, and Moe 2012; Mao, Counts, and Bollen 2011; Mestyan, Yasseri, and Kertesz 2012; Wu and Brynjolfsson 2009), recent studies report that excellent predictive performance of collective activity of online users may be limited to certain products (e.g., Wong, Sen, and Chiang 2012).

Online search volume of a product, which measures the collective online search activity of consumers, is a highly cited metric for its excellent predictive performance as well as ready availability. Focusing on the online search volume of motion pictures, we answer the question of when the online search volume of a movie reflects the consumers' collective purchase intent for the movie well, and when it does not. To answer the research question, we focus on the differential effects of two quality constructs—the perceived quality and perceived uncertainty about the quality—in influencing consumer search activity and converting the search activity into actual demand.

We hypothesize that both the perceived quality and quality uncertainty of a movie increase consumers' search activity for the movie, while it is only perceived quality—and not perceived uncertainty about quality—that positively influences the conversion from search into demand. The intuition is that any consumers want to reduce the risk of consuming an unknown product, a means of which is to search for information about the product. As such, the perceived quality uncertainty as well as the perceived quality of a movie can generate consumer search. On the other hand, the search activity motivated by quality uncertainty may not be converted into demand unless the perceived quality of the movie is sufficiently high to persuade consumers to watch it. As such, the perceived quality uncertainty of a movie will have no, if not negative, influence on the conversion from search into demand. The negative effect of quality uncertainty on the conversion process may be observed if consumers' tendency to avoid consumption risk is sufficiently high.

Quality is an abstract concept, and there exist different states of quality of a product over its life cycle (Golder, Mitra, and Moorman 2012). As such, we utilize different observable variables to operationalize consumers' perceived quality and quality uncertainty about a movie over its life cycle. Then we relate the observable quality variables with the online search volumes and revenues of movies in our model. The model consists of two systems of simultaneous equations, one for the opening week and the other for the subsequent weeks. The model is applied to a movie data set.

We find supporting evidence for our hypotheses. First, on the effect of quality and quality uncertainly on pre-launch search volume and opening-week revenue, we find that 1) both perceived quality and quality uncertainty about a movie increase pre-launch search volume of the movie; 2) only the perceived quality positively moderates the conversion of the pre-launch search volume into opening-week revenue. Second, we extend our analysis to the post-launch period and find the same results by use of different operationalizations of quality and quality uncertainty. Our findings imply that managers should interpret with caution the market demand predicted with online search volume (e.g., Kulkarni, Kannan, and Moe 2012).

Our contribution is both substantive and methodological. Substantively, we provide an explanation for why online search volume, which is receiving tremendous attention for its ability to provide business intelligence, may not be a good predictor of market demand for certain products from the same industry. New-product managers, especially movie studio managers, may want to use our findings to better forecast the market demand of their products. Methodological contributions are two fold. First, we propose a practical way of measuring the pre-launch quality of movies which can be used to better predict their pre-launch search volume. Second, we devise a novel method of constructing cross-sectionally comparable search volume measures from the readily available Google search indices.

The rest of the paper is organized as follows. We review relevant research and develop our hypotheses. Then we introduce our movie data set. We develop two systems of dynamic simulation equations and estimate them to test our hypotheses. We then provide managerial implications and formulate conclusions.

4.2. Relevant Literature and Theory Development

4.2.1. Quality Perception, Search Volume, and Demand

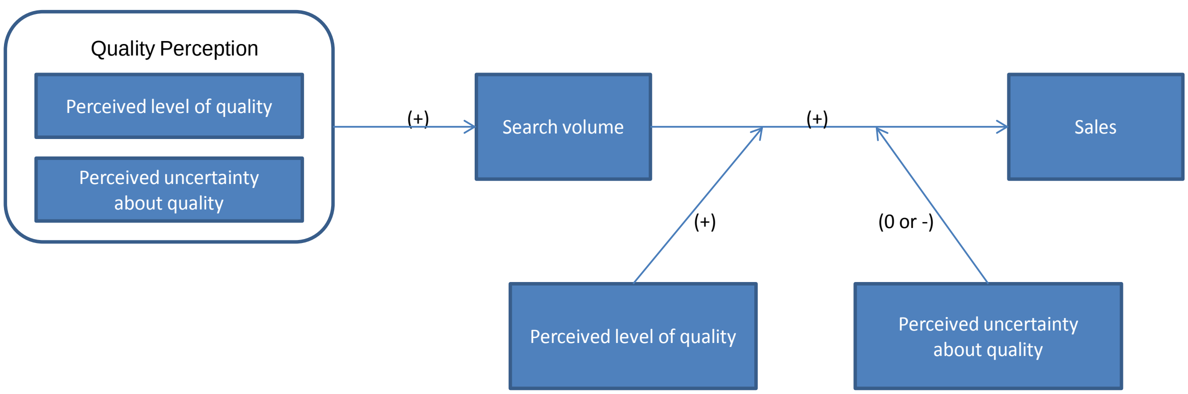
Perceived Quality, Quality Uncertainty and Search Volume. Previous research implies that the online search volume of a product increases with consumers' perceived quality of the product (Beatty and Smith 1987; Punj and Staelin 1983; Srinivasan and Ratchford 1991). This is because, all else being equal, higher quality products generate more interest among consumers, which leads to more information-seeking activities for that product. Therefore, if consumers can infer or perceive the quality of a product from some observable sources, research may use those sources to predict the search volume of the product. There will be a positive association between the perceived quality and the search volume of a product, all else being equal.

Previous research also implies that perceived uncertainty about a product's quality can influence the search volume of the product. Moorthy, Ratchford, and Talukdar (1997) find that perceived uncertainty of a brand increases the amount of search. Hess (1982)'s normative model of cost and benefit of search proposes that uncertainty about the payoffs from alternatives increases the expected returns of search, which in turn increases the amount of search effort. Chatterjee (2001) finds that consumers who are interested in a less popular product (such as niche products) are likely to search more. Because less popular products are less familiar and thus more uncertain to the consumer population, her finding proposes that more searches will be observed for products of higher uncertainty. In the same vein, Zhu and Zhang (2010) find that online reviews are more effective in influencing the purchases of less popular products because consumers are more likely to seek quality information to minimize the purchase risk. Therefore, if consumers can infer the quality uncertainty of a product from observable sources, those sources may be used to predict the search volume of the product. There will be a positive association between the perceived quality uncertainty and the search volume of a product, all else being equal.

Conversion of Search into Demand. Provided that consumers' search activities reflect their interest in the searched products (Kulkarni, Kannan, and Moe 2012), a certain portion of search activities will be converted into actual demand. The conversion of search into demand will be influenced by the quality information, as consumers will purchase the product only if they conclude that the quality of the product is sufficiently high. The perceived quality and quality uncertainty will have differential effects on the conversion of searching activity to actual purchase. The information about the level of quality will positively influence the conversion, whereas the information regarding the quality uncertainty will have either negative—if consumers regard high quality uncertainty as a signal of risk—or no influence the conversion.

In sum, consumer search activity works as a *mediator* between quality perception and actual purchase. Both the perceived quality and uncertainty about quality of a product positively influence the search volume of the product. In the conversion of search into demand, the information on the level of quality positively influences the conversion process, whereas the uncertainty has either negative or no effect on the process. Figure 4.1 illustrates our theory.

Figure 4.1: The Effects of Quality and Quality Uncertainty on Search Volume and Sales



4.2.2. Perception of Movie Quality

In the movie industry, where price differentiation is not a common marketing tool, the quality of movies is one of the most influential factors for market success. As Golder, Mitra, and Moorman (2012) argue, quality is a complex construct, and there exist different states of quality. Namely, in the production stage of a product, the quality of a product is described by the produced attributes; in the experience stage where firms deliver the product to consumers, the attributes of the product are subjectively perceived through consumers' own lens; finally in the evaluation stage, consumers generate a summary statement of their experience with the product in the form of reviews and ratings.

The study of Golder, Mitra, and Moorman (2012) proposes that consumers may use different sources of information to form their perception on movie quality along a movie's life cycle. First, in the pre-launch period of movies where neither experienced quality nor evaluated quality exists, consumers may form a perception on the quality of a movie based on movie characteristics. For example, in the pre-launch period of a movie, consumers may consider the participation of high-profile stars and directors (Basuroy, Desai, and Talukdar 2006; Hennig-Thurau, Houston, and Sridhar 2006) and whether the movie is a sequel (Basuroy, Desai, and Talukdar 2006) to form their quality perception on the movie.

Second, the experiential nature of movie consumption makes consumers seek other people's opinions to infer quality and make purchase decision (Bolton, Katok, and Ockenfels 2004; Clemons, Gao, and Hitt 2006; Pavlou and Gefen 2004). As such, when professional reviews become available, consumers will consider the review information to update their *a priori* assessment about movie quality that was formed solely from movie characteristics. Likewise, when user reviews become available, consumers will take them into account to form their perception on movie quality. Because consumers and professional critics emphasize different criteria when determining movie quality (Holbrook 1999), consumers may weigh more on other consumers' experiences than professional critics' opinions when both are available.

4.2.3. Operationalization of Movie Quality Perception

As quality and uncertainty about quality are abstract constructs, we operationalize them with observable variables for our empirical analysis. As consumers may rely on different information sources depending on the availability of the sources, we use different variables to operationalize the quality and quality uncertainty of movies over the movie life cycle.

Pre-Launch Period and the Opening Week. In the pre-launch period of a movie, consumers have neither user opinions nor professional reviews[7]. In this period, consumers have no other option but to rely on movie characteristics to infer movie quality. In particular, high-profile directors of a movie can signal to consumers that the movie is of high quality (Basuroy, Desai, and Talukdar 2006). For example, consumers may expect that a movie directed by Steven Spielberg is of higher quality than a movie directed by a less successful director. To extend this notion, when no review information is available, consumers may consider the past performance and filmography of the director of a movie to infer the quality and quality uncertainty of the movie. Therefore, we use the past performance of the director of a focal movie to operationalize consumers' pre-launch perception on quality and quality uncertainty about the movie. Specifically to operationalize the pre-launch quality of a movie, we use the average of user ratings of past movies with which the director of the focal movie have been involved as a director, writer, or producer (termed "average director rating from the past"). To operationalize the pre-launch quality uncertainty about a movie, we use the standard deviation of ratings of past movies with which the director of the focal movie have been involved as a director, writer, or producer (termed "standard deviation of director ratings from the past"). The average director

[7] Critic reviews are generally available immediately before movie release.

rating and standard deviation of director ratings from the past are used to examine the effect of perceived quality and quality uncertainty of a movie on its pre-launch search volume.

Other observable quantities in the pre-launch period are critic reviews and ratings. Usually critic reviews on a new movie are available immediately before the movie's release. When available, they can be used by consumers to update their perception on quality and quality uncertainty of the new movie. As such, to examine the effect of quality perception on the conversion of pre-launch search volume into opening-week demand of a movie, we use the average critic rating and standard deviation of critic ratings of the movie, in addition to the average director rating and standard deviation of director rating.

Post-Launch Period. To a movie consumer, user reviews are considered an important source for making the movie-going decision (Chintagunta, Gopinath, and Venkataraman 2010; Holbrook 1999). Furthermore, when both critic reviews and user reviews are available, consumers tend to rely more heavily on user reviews than critic reviews. Therefore, in the post-launch period, we turn to user reviews to operationalize the quality of movies. Specifically, we use average user rating (i.e., word-of-mouth valence) from the previous week to operationalize the perceived quality of a movie in the current week; we use the standard deviation of ratings from the previous week to operationalize the perceived uncertainty (or disagreement among consumers) about the quality of a movie in the current week.

4.2.4. Hypotheses

Summarizing the above theory, we propose the following hypotheses that relate the measurable variables of quality constructs with search volumes and revenues of movies. Hypotheses H1 through H4 deal with the pre-launch and opening-week relationship of movie quality, search volume, and revenue. H5 through H8 hypothesize the post-launch relationships of weekly movie quality, weekly search volume, and weekly revenue.

H1: The pre-launch search volume of a movie is positively associated with the average director rating from the past—i.e., the average user rating of the past movies with which the focal movie's director was involved as a director, writer, or producer.

H2: The pre-launch search volume of a new movie is positively associated with the standard deviation of director ratings from the past—i.e., the standard deviation of user ratings of the past movies with which the focal movie's director was involved as a director, writer, or producer.

H3: The conversion of pre-launch search volume into opening-week revenue of a movie is positively moderated by the perceived quality of the movie, namely the average director rating from the past and/or the average critic rating of the movie.

H4: The conversion of pre-launch search volume into opening-week revenue of a movie is either negatively influenced or not influenced by the perceived quality uncertainty of the movie, namely the standard deviation of director ratings from the past and/or the standard deviation of critic rating of the movie.

H5: After opening week, the weekly search volume of a movie is positively associated with the previous week's average user rating of the movie.

H6: After opening week, the weekly search volume of a movie is positively associated with the standard deviation of user ratings of the movie from the previous week.

H7: After opening week, the conversion of weekly search volume into revenue of a movie is positively moderated by the previous week's average user rating of the movie.

H8: After opening week, the conversion of weekly search volume into weekly revenue of a movie is either negatively influenced or not influenced by the standard deviation of user ratings of the movie from the previous week.

4.3. The Data

We construct a movie data set to test the hypotheses. Our data set consists of 174 movies, most of which were widely released in 2009 in the U.S. For each movie, we collect various movie characteristics, online search volume, revenue, marketing activities (advertising spending and number of screens), and online word-of-mouth (WOM) activities. Online search volume, revenue, marketing, and online WOM activities are collected weekly from 60 weeks before release to 10 weeks after release. Table 4.1 summarizes our variables and their sources.

Table 4.1: Variables and Data Sources

Category	Variable	Source of Data
Marketing activities	Weekly advertising spending	Nielsen
	Weekly number of screens	The numbers
Outcome variables	Weekly revenue	The numbers
Search activity	Weekly search index of movie keywords	Google
Perception on movie quality in pre-launch period and opening week	Average director rating from the past—i.e., average user rating of the past movies with which the focal movie's director was involved as a director, writer, or producer [range: 1 – 10]	IMDb
	Standard deviation of director ratings from the past—i.e., standard deviation of user ratings of the past movies with which the director of the focal movie was involved as a director, writer, or producer [range: 1 – 10]	
	Average critic rating [range: 1 – 100]	Metacritic
	Standard deviation of critic ratings [range: 1 – 100]	
Perception of movie quality in post-launch period	Weekly average user rating [range: 1 – 10]	IMDb, Yahoo
	Weekly standard deviation of user ratings [range: 1 – 10]	IMDb, Yahoo
Movie characteristics	Genre, MPAA rating, Sequel, Holiday, Production budget	IMDb, Wikipedia

Let us explain key variables. For the director of each movie in our data set, we collect performance of past movies with which the director was involved as a director, writer, or producer since 1990 until one calendar year before the focal movie' release. The total number of such past movies is 2,559. Among them, we include only the movies whose U.S. gross revenues are reported to IMDb.com. This equals 1,344 movies.[8] The average user rating of the 1,344 movies is 6.74 (10-point scale). The average of the standard deviations of user ratings of the 1,344 movies is 1.98. The average of total U.S. gross revenues by the directors since 1990 was $814 million. On average, each director was involved with 8.45 movies since 1990 as director, writer, or producer.

The expert reviews are collected from Metacritic.com. Metacritic.com collects professional reviews from up to 40 leading U.S. film critics and scores them between 1 and 100 by analyzing the review texts. The average number of reviews per movie is 27.4. The mean score is 57.1 with the maximum of 92.7 and the minimum of 14.3. The average of the standard deviations of critic rating scores is 15.7.

For the post-launch measure for quality and quality uncertainty of a movie, we collect user ratings from IMDb.com and the Yahoo Movies website. For each movie and week, we calculate the average rating and standard deviation of ratings. Therefore, in the post-launch period, a movie's quality varies on a weekly basis. While IMDb.com uses a 10-point scale for user ratings, the Yahoo Movies website uses a 5-point scale. To merge the two sites' information on the same scale, we multiply the user ratings from the Yahoo Movies website by two.

For marketing activities, we collect the weekly advertising spending and weekly number of screens of movies. The advertising data is provided by Nielsen and covers all major media

[8] The movies that were not reported to IMDb.com are small movies in terms of revenue. They appeal to a very small number of consumers who are particularly interested in the content of the movies. As such, excluding those movies from our analysis will not significantly affect the results.

outlets such as TV, radio, print, and Internet. The weekly number of screens is collected from The Numbers.com. Table 4.2 summarizes descriptive statistics of our focal variables.

Table 4.2: Descriptive Statistics

	Mean	Median	Std. Dev.	Min	Max
No. of user reviews per movie (N=174, up to 10 week after release)	4,835.9	2,203.5	8,272.2	113.0	71,764.0
Average user rating (N=174, up to 10 week after release)	7.0	7.1	1.2	3.8	9.5
Standard deviation of user ratings (N = 174, up to 10 week after release)	2.6	2.7	0.4	1.4	3.3
No of critic reviews per movie (N = 172)	27.8	29.5	7.8	1.0	39.0
Average critic rating (N=172, 100 point scale)	57.2	58.1	15.3	14.3	92.7
Standard deviation of critic ratings (N=171, 100 point scale)	15.7	15.6	3.2	5.0	23.1
No. of past movies of the focal movie's director (N=169)	8.1	5.0	8.2	0.0	49.0
Total U.S. gross box-office revenue of past movies of the focal movie's director (N=169)	\$774 M	\$337 M	\$1,052 M	\$ 0	\$6,518 M
Average director rating from the past (N=161)	6.7	6.8	0.6	4.8	8.7
Standard deviation of director ratings from the past (N=161)	2.0	2.0	0.2	1.5	3.4
Advertising spending (N=174)	\$20 M	\$20 M	\$12 M	\$6.5 K	\$51 M
Production budget (N=156)	\$ 55 M	\$38 M	\$53 M	\$11 K	\$250 M

Constructing Cross-Sectionally Comparable Search Volume. Google Trends provides weekly search indices of keyword queries entered into the Google search engine. Because the Google search index is normalized to conceal the actual search volume of the keyword, researchers cannot compare the search volumes across different keywords if the raw search index is used. We avoid this problem with a method that transforms the raw search indices from Google into cross-sectionally comparable search volume measures. The detailed methodology of collecting weekly Google search indices of movie keywords and transforming them into cross-sectionally comparable measures can be found in chapter 5, the appendix to the dissertation.

4.3.1. Correlation Analyses

Partial correlation analyses on our data suggest that online search volume is less predictive for certain types of movies. We examine the partial correlation coefficients between opening-week revenue, opening-week average user rating, average critic rating, and the cumulative online search volume up to one week before the opening (t = -1), after partialling out the effects of the opening-week screens and pre-launch advertising. The partial correlation analysis is conducted for three samples: the whole sample and two median-split subsamples by U.S. gross revenue. The results are in Table 4.3.

Table 4.3: Partial Correlation Analyses

(a) The Total Sample (N = 160)

Corr. Coef. (P-val)	Avg. critic rating	Avg. user rating at t = 0	Cumulative search at t = -1
Avg. user rating at t = 0	.42 (.00)		
Cumulative search at t = -1	.06 (.43)	.06 (.43)	
Opening revenue	.19 (.02)	.23 (.00)	.68 (.00)

(b) Median Split Subsample by U.S. Gross Revenue – Upper Half (N=80)

Corr. Coef. (P-val)	Avg. critic rating	Avg. user rating at t = 0	Cumulative search at t = -1
Avg. user rating at t = 0	.38 (.00)		
Cumulative search at t = -1	.00 (.97)	-.02 (.83)	
Opening revenue	.07 (.56)	.05 (.67)	.71 (.00)

(c) Median Split Subsample by U.S. Gross Revenue – Lower Half (N=80)

Corr. Coef. (P-val)	Avg. critic rating	Avg. user rating at t = 0	Cumulative search at t = -1
Avg. user rating at t = 0	.29 (.00)		
Cumulative search at t = -1	.03 (.81)	.07 (.54)	
Opening revenue	-.00 (.99)	.05 (.68)	.25 (.03)

First note that, across the three analyses, pre-launch search volume (i.e., cumulative search volume at t = -1) is much more highly correlated with opening-week revenue than average critic rating and average user rating are, implying that pre-launch online search volume can be a superior predictor for movie success to the rating variables.[9] However, comparing the partial correlation analyses of the two median-split subsamples reveals a stark difference: the partial correlation between pre-launch search volume and opening-week revenue is much stronger in the upper-half subsample than in the lower-half subsample. This shows a snippet that the predictive

[9] Also note that the partial correlation coefficients between opening-week revenue and average rating (both user ratings and critic ratings) become statistically insignificant in the two subsamples, whereas the partial correlation coefficient between the pre-launch search volume and opening-week revenue remains significant.

performance of online search volume can be better for successful movies than for less successful movies.

4.4. The Models

To test our hypotheses, we develop a system of dynamic simultaneous equations that closely follows the work of Elberse and Eliashberg (2003). We use a multiplicative formulation and distinguish between a movie's opening week and its run in subsequent weeks. In the model, index j represents movie j and index t represents week t. The release week is denoted by t = 0; k weeks before release is denoted by t = -k; and k weeks after release is denoted by t = k.

4.4.1. The Opening-Week Model

In the opening-week model, we construct a system of four equations: one equation with pre-launch search volume (the cumulative search volume at t = -1) as the dependent variable (the "search" equation), one with opening-week revenue as the dependent variable (the "revenue" equation), one with opening-week screen as the dependent variable (the "screen" equation), and one with average critic rating as the dependent variable (the "average critic rating" equation).

Search Equation. The model specification for pre-launch search volume is given by (4-1).

$$\text{(4-1)} \quad \begin{aligned} CSearch_{j,-1} = {} & e^{\alpha_0} CAd_{j,-1}^{\alpha_1} Avg\_DirectorRating_j^{\alpha_2} Sd\_DirectorRating_j^{\alpha_3} \\ & \times X_{CSearch,j}^{\alpha_4} e^{\alpha_5 D_{CSearch,j}} e^{\varepsilon_{j0}}. \end{aligned}$$

$CSearch_{j,-1}$ is the cumulative search volume of movie j at one week before its release (t = -1), $CAd_{j,-1}$ is the cumulative advertising of movie j at one week before its release.

$Avg\_DirectorRating_j$ is the average director rating from the past, and $Sd\_DirectorRating_j$ is the standard deviation of director ratings from the past. $Avg\_DirectorRating_j$ and $Sd\_DirectorRating_j$ operationalize consumers' pre-launch perception on movie j's quality and quality uncertainty, respectively. The parameter α_2 tests H1 and the parameters α_3 tests H2. $X_{Csearch,j}$ is the vector of numerical movie characteristics. $D_{CSearch,j}$ is the vector of categorical movie characteristics such as genre, MPAA rating, and seasonality dummy variables.

Revenue Equation. Equation (4-2) specifies opening-week revenue.

(4-2) $$Revenue_{j0} = e^{\beta_0} Ad_{j0}^{\beta_1} Scrns_{j0}^{\beta_2} CSearch_{j,-1}^{\beta_{3j}} X_{Revenue,j}^{\beta_4} e^{\beta_5 D_{Revenue,j}} e^{u_{j0}},$$

where

$$\begin{aligned}\beta_{3j} = {} & \beta_{30} + \beta_{31} \cdot Avg\_DirectorRating_j + \beta_{32} \cdot Avg\_CriticRating_j \\ & + \beta_{33} \cdot Sd\_DirectorRating_j + \beta_{34} \cdot Sd\_CriticRating_j.\end{aligned}$$

$Revenue_{j0}$ is the opening-week revenue of movie j, Ad_{j0} and $Scrns_{j0}$ are the advertising and the number of screens of movie j in the opening week. The parameter β_{3j} measures the conversion elasticity from pre-launch search into opening-week revenue. To test H3 and H4, we model β_{3j} as a function of quality and quality uncertainty variables that are available to consumers in the launch week, namely average director rating from the past ($Avg\_DirectorRating_j$), average critic rating ($Avg\_CriticRating_j$), standard deviation of director ratings from the past ($Sd\_DirectorRating_j$), and standard deviation of critic ratings ($Sd\_CriticRating_j$). $Avg\_DirectorRating_j$ and $Avg\_CriticRating_j$ operationalize the quality of movie j, and $Sd\_DirectorRating_j$ and $Sd\_CriticRating_j$ operationalize the uncertainty about quality of movie j. H3 is tested by parameters β_{31}, and β_{32}; H4 is tested by parameters β_{33}, and β_{34}. $D_{Revenue,j}$ and $D_{Revenue,j}$ are the set of numerical and categorical characteristics of movie j.

Screen Equation. The model for the number of screens in the opening week of movie j is specified by (4-3).

$$\text{(4-3)} \qquad Scrns_{j0} = e^{\gamma_0} CAd_{j,-1}^{\gamma_1} Avg\_CriticRating^{\gamma_2} X_{Scrns,j}^{\gamma_3} e^{\gamma_4 D_{Scrns,j}} e^{v_{j0}} .$$

$X_{Scrns,j}$ is the vector of numerical movie characteristics of movie j, and $D_{Scrns,j}$ is the vector of categorical movie characteristics. We do not include the expected revenue as a covariate as Elberse and Eliashberg (2003) do for two reasons. First, the objective of this study is not to examine the role of expected revenue in determining the number of screens. Second, because calculating the expected revenue requires double exponential smoothing, the revenues of the first two weeks (the opening week and one week after the opening week) cannot be calculated. Using the actual revenues for the first two weeks can cause identification problems—especially for the opening-week model.[10]

Average Critic Rating Equation. Equation (4-4) specifies the model for the average critic rating of movie j.

$$\text{(4-4)} \qquad Avg\_CriticRating_j = e^{\delta_0} CAd_{j,-1}^{\delta_1} Avg\_DirectorRating_j^{\delta_2} X_{Quality,j}^{\delta_3} e^{\delta_4 D_{Quality,j}} e^{\omega_{j0}} .$$

$X_{Quality,j}$ is the vector of numerical movie characteristic, and $D_{Quality,j}$ is the vector of categorical movie characteristics. To examine the effect of commercial communication on professional reviews, we include the cumulative advertising at one week before release. Also, we include the average director rating from the past ($Avg\_DirectorRating_j$) to examine the predictive ability of the focal director's past performance in predicting professional opinions on his/her current movie.

[10] Elberse and Eliashberg (2003) avoid this problem by using the movie stock price at the Hollywood Stock Exchange for the opening week's expected revenue.

4.4.2. The Post-Launch Period Model

The post-launch period model is similarly constructed by replacing the variables in the opening-week model by their corresponding counterparts available in the post-launch periods. Specifically, we use the average *user* rating and standard deviation of *user* ratings from the previous week to operationalize the weekly varying perception on quality and quality uncertainty of a movie. Using the weekly changing variables for the quality constructs enables us to test whether our theory holds even within the same movie as the quality-related information changes over time. The post-launch period model consists of five equations: search equation, revenue equation, screens equations, equation of weekly average user rating, and equation of weekly standard deviation of user ratings. In each equation, we include week dummy variables to control for the unobserved time effect.

Search Equation. The weekly search volume in the post-launch period is modeled in (4-5).

$$\text{(4-5)} \qquad \text{Search}_{jt} = e^{\alpha_0} \text{Ad}_{jt}^{\alpha_1} \text{Avg\_UserRating}_{j,t-1}^{\alpha_2} \text{Sd\_UserRating}_{j,t-1}^{\alpha_3} e^{\alpha_4 \text{Weeklydummy}} e^{\varepsilon_{jt}},$$

$$\text{for } t \geq 1.$$

Search_{jt} and Ad_{jt} are the search volume and advertising of movie j in week t. $\text{Avg\_UserRating}_{j,t-1}$ and $\text{Sd\_UserRating}_{j,t-1}$ are the average user rating and the standard deviation of user ratings from the previous week. $\text{AvgUserRating}_{j,t-1}$ and $\text{SdUserRating}_{j,t-1}$ operationalize the perceived quality and quality uncertainty about movie j in week t. They test H5 and H6, respectively.

Revenue equation. Equation (4-6) models the weekly revenue of movie j in post-launch week t.

$$\text{(4-6)} \qquad \text{Revenue}_{jt} = e^{\beta_0} \text{Ad}_{jt}^{\beta_1} \text{Scrn}_{jt}^{\beta_2} \text{Search}_{jt}^{\beta_{3jt}} \text{Holiday}^{\beta_4} e^{\beta_5 \text{Weeklydummy}} e^{u_{jt}}, \ \text{for } t \geq 1,$$

$$\text{where } \beta_{3j} = \beta_{30} + \beta_{31} \cdot \text{Avg\_UserRating}_{j,t-1} + \beta_{32} \cdot \text{Sd\_UserRating}_{j,t-1}.$$

Revenue_{jt}, Ad_{jt}, Scrn_{jt}, and Search_{jt} are the revenue, advertising spending, number of screens, and search volume of movie j in post-launch week t. To test H7 and H8, the conversion of search volume into revenue (β_{3jt}) is modeled as a function of the average user rating and standard deviation of user ratings from the previous week. The variable Holiday is holiday dummy variable.

Screen Equation. The model for the number of screens of movie j in week t is specified by (4-7).

$$\text{(4-7)} \qquad \begin{aligned} \text{Scrn}_{jt} = {} & e^{\gamma_0} \text{Scrn}_{j,t-1}^{\gamma_1} \text{Ad}_{j,t-1}^{\gamma_2} \text{Search}_{j,t-1}^{\gamma_3} \text{Avg\_UserRating}_{j,t-1}^{\gamma_4} \text{Sd\_UserRating}_{j,t-1}^{\gamma_5} \\ & \times \text{Holiday}^{\gamma_6} e^{\gamma_7 \text{Weeklydummy}} e^{v_{jt}} \end{aligned},$$

$$\text{for } t \geq 1.$$

The previous week's number of screens ($\text{Scrn}_{j,t-1}$) is included to incorporate the decision inertia in movie theaters' screen allocation decision. The previous week's search volume ($\text{Search}_{j,t-1}$) is included because movie theaters may consider the previous week's popularity of the movie to allocate their screens. The coefficient of $\text{Search}_{j,t-1}$ will partially incorporate the advertising effect and revenue effect from the previous week. The variables, $\text{Avg\_UserRating}_{j,t-1}$ and $\text{Sd\_UserRating}_{j,t-1}$, are included to examine whether the quality and quality uncertainty of a movie influence the screen decision for that movie.

Equations for Weekly User Rating. The weekly average user rating and the weekly standard deviation of user ratings are specified by equations (4-8) and (4-9).

$$\text{(4-8)} \qquad \text{Avg\_UserRating}_{jt} = e^{\delta_0}\text{Ad}_{jt}^{\delta_1}\text{Avg\_UserRating}_{j,t-1}^{\delta_2}\text{Revenue}_{jt}^{\delta_3}e^{\delta_4 \text{Weeklydummy}}e^{\omega_{jt}}\text{, for } t\geq 1.$$

$$\text{(4-9)} \qquad \text{Sd\_UserRating}_{jt} = e^{\kappa_0}\text{Ad}_{jt}^{\kappa_1}\text{Sd\_UserRating}_{j,t-1}^{\kappa_2}\text{Revenue}_{jt}^{\kappa_3}e^{\kappa_4 \text{Weeklydummy}}e^{\xi_{jt}}\text{, for } t\geq 1.$$

We include advertising of movie j to control for the effect of advertising on the average user opinion and the disagreement in user opinions about the movie. The previous week's dependent variable is included to account for the high inertia of consumer opinions about movies over their life cycle. The potential association between the current week's revenue and the two dependent variables are controlled for by including Revenue_{jt} in the equations.

4.5. Empirical Analyses

We use a three-stage least-squares (3SLS) method to estimate the system of equations (4-1) through (4-9). A ordinary least squares (OLS) method is inconsistent because some covariates are endogenous. Also, a two-stage least-squares (2SLS) method is less efficient than a 3SLS method as the errors across equations may be correlated (Zeller and Theil 1962). For the opening-week model, equations (4-1) through (4-4) are estimated as a system and the variables $\text{CSearch}_{j,-1}$, $\text{Revenue}_{j,0}$, $\text{Scrns}_{j,0}$, and $\text{Avg\_CriticRating}_{j}$ are treated as endogenous variables. For the post-launch period model, equations (4-5) through (4-9) are estimated as a system where the variables Search_{jt}, Revenue_{jt}, Scrns_{jt}, $\text{Avg\_UserRating}_{jt}$, and $\text{Sd\_UserRating}_{jt}$ are treated as endogenous variables. When estimating the post-launch period model, we exclude the first-lag endogenous variables from the set of instrumental variables to reduce potential estimation problems related to serial correlation (Elberse and Eliashberg 2003). Instead, as in Elberse and Eliashberg (2003), we employ the set of time-invariant exogenous variables that are used to estimate the opening-week model. We also use the second-lag endogenous variables of the weekly average rating and the standard deviation of weekly average ratings, as they cause less serious problems for identification. The time-specific fixed effects are accounted for by week dummy variables in the post-launch period model.

4.5.1. Estimation of the Opening-Week Model

Table 4.4 shows the estimation results for the opening-week model.

Table 4.4: Estimation Results of the Opening-Week Model

(a) Search Equation (DV: pre-launch search volume)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	-4.19	4.23	0.32	-3.26	4.24	0.44	-3.34	4.01	0.40
Pre-launch advertising	0.44	0.07	0.00	0.35	0.08	0.00	0.35	0.07	0.00
Average director rating from the past	4.87	1.74	0.01	5.10	1.76	0.00	5.13	1.66	0.00
SD of director ratings from the past	4.75	1.35	0.00	5.06	1.36	0.00	5.08	1.29	0.00
Genre: Action	0.92	0.34	0.01	0.68	0.34	0.05	0.69	0.32	0.03
Genre: Comedy	-0.33	0.32	0.30	-0.57	0.32	0.08	-0.57	0.31	0.06
Genre: Drama	0.28	0.38	0.46	0.22	0.40	0.58	0.23	0.38	0.54
MPAA: R	-0.23	0.76	0.76	-0.98	0.84	0.24	-1.00	0.79	0.21
MPAA: PG	-0.36	0.76	0.63	-1.18	0.84	0.16	-1.20	0.79	0.13
MPAA: PG13	-0.60	0.75	0.43	-1.12	0.83	0.18	-1.14	0.79	0.15
January-April	-0.34	0.35	0.33	-0.17	0.36	0.63	-0.16	0.34	0.63
May-August	-0.32	0.34	0.35	-0.10	0.34	0.77	-0.09	0.32	0.79
September-October	-1.00	0.39	0.01	-0.96	0.38	0.01	-0.95	0.36	0.01
Holiday	-0.06	0.33	0.85	-0.04	0.32	0.89	-0.05	0.30	0.87
Sequel	2.18	0.41	0.00	2.14	0.40	0.00	2.14	0.38	0.00
R^2			0.49			N.A.			N.A.
Adj. R^2			0.43			N.A.			N.A.
SSR			1.40			1.35			1.35

(b) Revenue Equation (DV: opening-week revenue)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	7.96	0.51	0.00	7.67	0.65	0.00	7.66	0.60	0.00
Advertising in the opening week	0.12	0.04	0.00	0.09	0.05	0.09	0.08	0.05	0.07
Opening screens	0.69	0.03	0.00	0.78	0.04	0.00	0.78	0.04	0.00
Pre-launch search volume	-0.02	0.11	0.87	-0.03	0.12	0.82	-0.02	0.11	0.84
Pre-launch search volume × Avg. critic rating	1e-3	0.00	0.00	1e-3	0.00	0.00	1e-3	0.00	0.00
Pre-launch search volume × Avg. director rating from the past	0.01	0.01	0.58	0.01	0.01	0.44	0.01	0.01	0.43
Pre-launch search volume × SD of critic ratings	0.00	0.00	0.47	0.00	0.00	0.99	0.00	0.00	0.97
Pre-launch search volume × SD of director ratings from the past	0.03	0.02	0.11	0.03	0.02	0.14	0.03	0.02	0.11
Genre: Action	0.07	0.15	0.64	0.15	0.15	0.35	0.14	0.14	0.31
Genre: Comedy	0.01	0.13	0.92	0.07	0.15	0.65	0.07	0.14	0.63
Genre: Drama	0.07	0.16	0.67	0.27	0.18	0.14	0.26	0.16	0.11
MPAA: R	-0.57	0.31	0.07	-0.49	0.37	0.18	-0.48	0.34	0.15
MPAA: PG	-0.43	0.31	0.17	-0.52	0.38	0.17	-0.52	0.35	0.14
MPAA: PG13	-0.47	0.31	0.13	-0.52	0.37	0.16	-0.52	0.34	0.13
January-April	-0.20	0.15	0.18	-0.19	0.16	0.24	-0.20	0.15	0.19
May-August	-0.10	0.14	0.48	-0.19	0.15	0.20	-0.20	0.14	0.16
September-October	-0.31	0.17	0.07	-0.37	0.18	0.04	-0.37	0.17	0.03
Holiday	0.21	0.14	0.15	0.21	0.15	0.17	0.21	0.14	0.14
Sequel	0.46	0.19	0.02	0.40	0.21	0.05	0.40	0.19	0.04
R^2			0.93			N.A.			N.A.
Adj. R^2			0.93			N.A.			N.A.
SSR			0.57			0.58			0.58

(c) Screen Equation (DV: no. of opening-week screens)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	2.88	1.99	0.15	2.58	2.22	0.25	2.99	2.11	0.16
Pre-launch Advertising	0.94	0.06	0.00	0.94	0.07	0.00	0.94	0.07	0.00
Average Critic Rating	-0.97	0.41	0.02	-0.96	0.45	0.03	-1.05	0.43	0.01
Major Distributor	0.31	0.38	0.42	0.29	0.39	0.45	0.07	0.36	0.85
Genre: Action	0.15	0.31	0.62	0.16	0.33	0.63	0.17	0.31	0.58
Genre: Comedy	0.34	0.29	0.24	0.50	0.31	0.11	0.51	0.30	0.09
Genre: Drama	-0.68	0.34	0.04	-0.70	0.37	0.05	-0.71	0.35	0.04
MPAA: R	-0.18	0.65	0.78	0.02	0.72	0.98	0.01	0.69	0.98
MPAA: PG	0.74	0.65	0.26	0.95	0.73	0.20	0.98	0.70	0.16
MPAA: PG13	0.22	0.65	0.73	0.55	0.73	0.45	0.56	0.70	0.43
January-April	0.73	0.29	0.01	0.67	0.32	0.03	0.67	0.30	0.03
May-August	0.41	0.28	0.14	0.41	0.29	0.17	0.40	0.28	0.15
September-October	0.72	0.33	0.03	0.66	0.34	0.05	0.65	0.33	0.05
Sequel	0.32	0.40	0.42	0.34	0.41	0.40	0.37	0.39	0.34
R^2			0.73			N.A.			N.A.
Adj. R^2			0.71			N.A.			N.A.
SSR			1.34			1.35			1.35

(d) Quality Equation (DV: average critic rating)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	2.12	0.55	0.00	2.39	0.59	0.00	2.39	0.57	0.00
Pre-launch Advertising	-0.02	0.01	0.04	-0.02	0.01	0.11	-0.02	0.01	0.09
Average director rating from the past	1.05	0.24	0.00	0.94	0.26	0.00	0.94	0.25	0.00
Genre: Action	-0.04	0.06	0.51	-0.03	0.06	0.58	-0.03	0.06	0.57
Genre: Comedy	-0.11	0.05	0.04	-0.10	0.06	0.07	-0.10	0.05	0.06
Genre: Drama	0.01	0.06	0.92	0.03	0.06	0.64	0.03	0.06	0.63
MPAA: R	0.04	0.13	0.78	-0.01	0.15	0.96	-0.01	0.14	0.96
MPAA: PG	0.09	0.13	0.51	0.04	0.15	0.79	0.04	0.14	0.78
MPAA: PG13	-0.05	0.13	0.71	-0.11	0.15	0.47	-0.11	0.14	0.45
Major Distributor	-0.24	0.07	0.00	-0.18	0.07	0.01	-0.19	0.07	0.00
Sequel	-0.12	0.07	0.10	-0.14	0.07	0.05	-0.14	0.07	0.04
R^2			0.34			N.A.			N.A.
Adj. R^2			0.29			N.A.			N.A.
SSR			0.25			0.24			0.24

Let us discuss the effect of our focal variables in the 3SLS results. First, not only the average director rating from the past (our operationalization of perceived movie quality in the pre-launch period) but also the standard deviation of director ratings from the past (our operationalization of perceived quality uncertainty of movies in the pre-launch period) are positively associated with the pre-launch search volumes of movies. Therefore, H1 and H2 are supported. Second, the conversion from search into demand of a movie is positively moderated only by the average critic rating of the movie. This positive interaction effect of search volume and average critic rating on opening-week revenue indicates that the information-seeking behavior of consumers, i.e., online search, is better translated into demand when consumers are ensured about the quality of the movie through professional reviews. Note that the variables related to quality uncertainty do not moderate the conversion from search into demand. These findings support H3 and H4.

Our study finds the determinants of the average critic rating of movies (Table 4.4(d)). The average critic rating of a movie is positively associated with the average user rating of the past movies with which the focal movie's director was involved. On the other hand, the average critic rating of a movie is negatively associated with the pre-launch advertising amount, whether the movie was distributed by a major distributor, and whether the movie is a sequel. As the determinants of critic ratings are observable long before a movie's release time, this finding provides managers a tool for early prediction of a movie's average critic rating. More importantly, as average critic rating moderate the conversation rate from pre-launch search to opening-week revenue, this finding implies that the conversation rate from pre-launch search volume to opening-week revenue of a movie can be predicted in its early pre-launch period.

Other noteworthy findings are as follows. Pre-launch advertising increases pre-launch consumer search activity, and sequel movies tend to receive more attentions (i.e., searches) from consumers than original movies. A studio's decision to make a sequel movie signals that the sequel movie is of high quality in the eye of ordinary consumers (Basuroy, Desai, and Talukdar 2006), making more consumers interested in the sequel movie. The number of screens, pre-launch interests (as measured by pre-launch search volume), and opening-week advertising are pivotal determinants of opening-week revenue.

4.5.2. Estimation of the Post-Launch Period Model

Focusing on the 3SLS results, we discuss our findings. Table 4.5 shows that the similar findings are maintained in the post-launch period: 1) both the perceived quality (i.e., the previous week's average user rating) and the perceived uncertainty about quality (i.e., the standard deviation of user ratings from the previous week) increase the current week's search volume; 2) online search activity is more effectively converted to actual consumption if the previous week's average user rating is higher. These findings provide empirical evidences for our hypotheses H5 through H8.

Table 4.5: Estimation Results of the Post-launch Period Model

(a) Search Equation (DV: weekly search volume)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	6.75	0.47	0.00	0.13	1.18	0.91	0.38	1.17	0.75
Advertising	0.22	0.01	0.00	0.20	0.02	0.00	0.20	0.02	0.00
Avg. user rating from the previous week	0.57	0.19	0.00	2.27	0.38	0.00	2.19	0.37	0.00
SD of user ratings from the previous week	1.00	0.13	0.00	3.71	0.46	0.00	3.64	0.46	0.00
Holiday	0.20	0.08	0.02	0.16	0.10	0.09	0.17	0.09	0.07
R^2			0.30			N.A.			N.A.
Adj. R^2			0.30			N.A.			N.A.
SSR			1.15			1.28			1.27

(b) Revenue Equation (DV: weekly revenue)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	5.66	0.14	0.00	6.00	0.26	0.00	5.67	0.26	0.00
Advertising	0.19	0.01	0.00	0.17	0.01	0.00	0.16	0.01	0.00
Screen	0.74	0.01	0.00	0.62	0.03	0.00	0.62	0.02	0.00
Search volume	0.13	0.02	0.00	0.17	0.05	0.00	0.23	0.05	0.00
Search volume × Avg. user rating from the previous week	0.01	0.00	0.00	0.02	0.00	0.00	0.02	0.00	0.00
Search volume × SD of user rating from the previous week	0.00	0.00	0.90	0.00	0.01	0.98	-0.01	0.01	0.49
Holiday	0.10	0.04	0.01	0.07	0.04	0.10	0.06	0.04	0.14
R^2			0.91			N.A.			N.A.
Adj. R^2			0.91			N.A.			N.A.
SSR			0.54			0.56			0.57

(c) Screen Equation (DV: weekly no. of screens)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	1.20	0.16	0.00	0.83	0.35	0.02	0.93	0.35	0.01
Previous week's screens	0.76	0.01	0.00	0.78	0.02	0.00	0.79	0.02	0.00
Previous week's advertising	0.10	0.01	0.00	0.10	0.01	0.00	0.11	0.01	0.00
Previous week's search volume	0.06	0.01	0.00	0.03	0.02	0.19	0.00	0.02	0.98
Avg. user rating from the previous week	0.17	0.06	0.00	0.32	0.11	0.00	0.35	0.11	0.00
SD of user ratings of the previous week	-0.04	0.05	0.48	0.07	0.19	0.72	0.13	0.19	0.49
Holiday	-0.04	0.03	0.21	-0.03	0.03	0.34	-0.03	0.03	0.36
R^2			0.92			N.A.			N.A.
Adj. R^2			0.92			N.A.			N.A.
SSR			0.42			0.43			0.43

(d) Quality Equation (DV: average of weekly user ratings)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	0.61	0.08	0.00	0.18	0.13	0.16	0.23	0.13	0.08
Advertising	0.01	0.00	0.00	0.01	0.00	0.15	0.01	0.00	0.07
Avg. user rating from the previous week	0.51	0.03	0.00	0.88	0.05	0.00	0.89	0.05	0.00
Revenue of this week	0.02	0.00	0.00	0.00	0.01	0.85	-0.01	0.01	0.50
R^2			0.30			N.A.			N.A.
Adj. R^2			0.30			N.A.			N.A.
SSR			0.23			0.23			0.23

(e) Quality Uncertainty Equation (DV: SD of weekly user ratings)

	OLS			2SLS			3SLS		
Variable	Coef	SE	P-val	Coef	SE	P-val	Coef	SE	P-val
Constant	0.65	0.09	0.00	0.21	0.15	0.16	0.39	0.15	0.01
Advertising	-0.01	0.00	0.08	0.00	0.01	0.60	0.00	0.01	0.49
SD of user ratings from the previous week	0.31	0.03	0.00	0.70	0.08	0.00	0.74	0.08	0.00
Revenue of this week	0.02	0.01	0.00	0.01	0.01	0.26	0.00	0.01	0.74
R^2			0.11			N.A.			N.A.
Adj. R^2			0.10			N.A.			N.A.
SSR			0.25			0.25			0.25

Other important findings are as follows. Not only is the current week’s distribution highly dependent on the previous level of distribution, but it also is influenced by the previous week’s advertising level and the previous week’s average user rating. Average user rating in a week is influenced by that week’s advertising, but standard deviation of user ratings is not influenced by the week’s advertising spending.

4.5.3. Robustness Verification

Some may argue that the findings may vary according to the keywords that are chosen to collect the Google search index. To test whether the substantive findings change according to keyword selection rules, we collected the second set of search indices with a new keyword rule. The new rule is stricter than the original rule described in the appendix, chapter 5. Namely, the new keywords include the word "movie" in the search query. For example, the search index of the movie *12 Rounds* is gathered using the keyword "12 Rounds Movie". The new keyword rule is supposed to count only a subset of the search queries that are counted by the original rule, due to Google's search index calculation mechanism. Descriptive statistics and correlation analyses are presented in Table 4.6.

Table 4.6: Comparison of Search Volumes by the Two Rules

(a) Descriptive Statistics

	Mean	Median	Std. Dev.	Min	Max
Search volume by the original keyword rule	533,680	120,479	1,618,584	580	16,236,978
Search volume by the new keyword rule	2,832	796	7,912	0	87,564

(b) Partial Correlation Controlling For Pre-launch Advertising and Opening Screens

Corr. Coef. (p-val.)	Critic rating	User rating at the opening week	Pre-launch search volume by the original rule	Pre-launch search volume by the new rule
User rating at the opening week	.42 (.00)			
Pre-launch search volume by the original rule	.06 (.50)	.06 (.44)		
Pre-launch search volume by the new rule	.10 (.21)	.01 (.87)	.55 (.00)	
Opening revenue	.19 (.03)	.21 (.01)	.68 (.00)	.37 (.00)

The average search volume by the original rule is about 188 times greater than that by the new rule. This implies that consumers do not include the term “movie” when they search for movies. The partial correlation analyses show that the search volume by the original rule is much more highly correlated with other metrics than the search volume by the new rule. Especially, the correlation coefficients between opening-week revenue and pre-launch search volume indicate that the search index collected by the original rule reflects the aggregate purchase intent of consumers substantially better than the one collected by the new rule. This finding again supports the implication that most consumers generally do not include the term “movie” when they search information for a movie.

Table 4.7 shows the estimation results of our key equations by the 3SLS procedure. Note that the same substantive findings hold for both the search and revenue equations. Therefore, our findings are consistent with the two keyword selection rules.

Table 4.7: Estimation Results Using the New Search Volume: 3SLS Results

(a) Search Equation

	Opening-Week Model			Post-Launch Period Model		
Variable	Coef.	SE	P val	Coef.	SE	P val
Constant	-10.76	4.03	0.01	1.05	0.93	0.26
Pre-launch Advertising	0.32	0.08	0.00			
Average director rating from the past	6.39	1.66	0.00			
SD of director ratings from the past	5.57	1.32	0.00			
Advertising				0.15	0.01	0.00
Average user rating from the previous week				1.48	0.29	0.00
SD of user ratings from the previous week				0.78	0.36	0.03
R^2			N.A.			N.A.
Adj. R^2			N.A.			N.A.
SSR			1.34			0.98

(b) Revenue Equation

	Opening-Week Model			Post-Launch Period Model		
Variable	Coef.	SE	P val	Coef.	SE	P val
Constant	8.92	0.55	0.00	5.71	0.31	0.00
Advertising in the opening week	0.09	0.05	0.06			
Opening screens	0.75	0.04	0.00			
Pre-launch search volume	-0.13	0.19	0.48			
Pre-launch search volume × Avg. critic rating	0.00	0.00	0.00			
Pre-launch search volume × Average director rating from the past	-0.01	0.02	0.58			
Pre-launch search volume × SD of critic ratings	0.00	0.00	0.33			
Pre-launch search volume × SD of director ratings from the past	0.06	0.04	0.14			
Advertising				0.16	0.01	0.00
Screen				0.79	0.02	0.00
Search volume				0.19	0.09	0.04
Search volume × Avg. user rating from the previous week				0.03	0.01	0.00
Search volume × SD of user rating from the previous week				-0.02	0.02	0.32
R^2			N.A.			N.A.
Adj. R^2			N.A.			N.A.
SSR			0.60			0.55

4.6. Managerial Implications

The findings provide important implications for managers. First, systematic over-/under-prediction of box-office revenue is likely if we rely on online search volume without accounting for the quality uncertainty. Over-prediction is likely if a focal movie's director has experienced high standard deviation of user ratings for his or her past movies, and the focal movie's average critic rating is low. As an early sign for the professional reviews a new movie will receive, managers can use the determinants of critic ratings that are found in this study—namely, the past performance of the focal director, major distributor dummy, and whether the new movie is a sequel. On the other hand, systematic under-prediction of box-office revenue can occur for movies with the opposite characteristic—i.e., the directors of the new movies have shown stable performance in terms of the standard deviation of user ratings, and at the same time, the movies' average critic ratings are high.

Second, the same level of advertising leads to more revenue for movies with a higher level of user disagreement about quality, all else being equal. This is because, for movies of higher quality uncertainty, the same level of advertising creates more search activities (equations (4-1) and (4-2)), which is, in turn, associated with more revenue (equations (4-3) and (4-4)). In light of consumer behavior theory, this is related to the notion of awareness effect and mere exposure effect that are created by consumers' search activities (Duan, Gu, and Whinston 2008; Janiszewski 1993; Liu 2006; Zajonc 1968).

Third, the conversion rate from pre-launch search volume into opening-week revenue can be predicted in the early pre-launch period as the determinants of conversion are observable long before a new movie's release.

Fourth, as a practical way of measuring quality uncertainty, we propose to use the standard deviation of user ratings of past movies of the focal director and the standard deviation of user ratings of the previous week for the opening week and the subsequent weeks, respectively.

Lastly, keyword selection is crucial to accurately measuring consumers' online search activities. Too strict a rule for keyword selection can underestimate actual consumer interest while too loose a rule can overestimate marketing effects. Indeed, finding appropriate keywords is a realm of art and it may need industry experts who are knowledgeable about actual keyword selection by consumers. For this reason, if the objective of the study is to read market trends, it is recommended to collect multiple time-series and extract underlying latent factors that co-move the multiple time-series (Du and Kamakura 2012).

4.7. Conclusions

While collective consumer activities on the Internet have been considered a good predictor of demand, it is also reported to have limited predictive ability for certain products. This study shows that such limited predictive performance also happens with the Google search index and provides an explanation on why aggregate online search volume is not a good predictor for certain products. At the core of our theory lies quality-related constructs such as perceived quality and quality uncertainty. Specifically, we hypothesize that both perceived quality and quality uncertainty of a movie increase consumers' search activity for the movie, while only perceived quality positively influences the conversion from search activity to actual demand. We find empirical support for our theory in the U.S. movie industry.

We also make methodological contributions. We propose a practical way of measuring quality-related constructs in the movie industry, especially in the early pre-launch period where no reviews are available. Also, we devise a novel method of constructing cross-sectionally comparable search volume measures from readily available Google search indices.

Our study provides valuable managerial implications. First, it diagnoses which movies are more prone to over-/under-prediction when managers use online search volume for prediction purposes. Second, our study suggests that movie studios take a look at the quality and quality uncertainty variables to allocate their advertising budget across different movies. Third, the conversion rate from search into sales can be predicted early by use of available public information. Lastly, the study suggests that finding appropriate construction rules for keywords is critical.

This study also has limitations. First of all, our findings are specific to the movie industry. Extending to other industries will be necessary to generalize our theory. Second, as with any study using online data is, our study is subject to measurement error, especially the online search volume measure. While we show that our search volume measure is reasonable by showing its superior correlation with revenue, the search volume is subject to the keywords that we choose. In fact, any study that uses search volume—whether the data is at the aggregate level or the individual consumer level—has this measurement problem. In this sense, finding an appropriate rule of keyword selection can be an intriguing research topic.

There are many interesting future research topics that use readily available search volume data. First, we can consider combining online search volume with other data sources to better measure consumers' purchase intent. For example, to better forecast the demand of a new product, we can combine search volume with primary data sources that directly measure purchase intention, consumption interest, and WOM valence. Second, online search volume can be used to measure the effectiveness of marketing activities that have been impossible to measure in the past due to a lack of the appropriate response variables. For example, weekly search volume in the pre-launch period of new products can be used to measure the effectiveness of pre-launch advertising. Another example is to estimate latent demands of certain products across different geographic regions (e.g., DMAs) if the online search volume can be measured at the geographic market level. Third, studying the joint effect of online search and online opinion posting can be an appealing topic. Online search is related to reading information while online opinion posting is related to writing information. While there is ample research on the effect of online reviews on market outcomes, there is little research so far that jointly examines the effect of online reading and writing.

4.8. References

Basuroy, Suman, Kalpesh Kaushik Desai, and Debabrata Talukdar (2006), "An Empirical Investigation of Signaling in the Motion Picture Industry," Journal of Marketing Research, 43 (May), 287-95.

Beatty, Sharon E. and Scott M. Smith (1987), "External Search Effort: An Investigation across Several Product Categories," Journal of Consumer Research, 14 (1), 83-95.

Bolton, Gary E., Elena Katok, and Axel Ockenfels (2004), "How Effective Are Electronic Reputation Mechanisms? An Experimental Investigation," Management Science, 50 (11), 1587-1602.

Chatterjee, Patrali (2001), "Online Reviews: Do Consumers Use Them?" in Advances in Consumer Research, Mary C. Gilly and Joan Myers-Levy, eds. Provo, UT: Association for Consumer Research, 129-34.

Chintagunta, Pradeep K., Shyam Gopinath, and Sriram Venkataraman (2010), "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," Marketing Science, 29 (5), 944-57.

Clemons, Eric K., Guodong Gao, and Lorin M. Hitt (2006), "When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry," Journal of Management Information Systems, 23 (2), 149-71.

Du, Rex Yuxing and Wagner A. Kamakura (2012), "Quantitative Trendspotting," Journal of Marketing Research, forthcoming.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do Online Reviews Matter? – An Empirical Investigation of Panel Data," Decision Support Systems, 45 (November), 1007-16.

Elberse, Anita and Jehoshua Eliashberg (2003), "Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures," Marketing Science, 22 (May), 329-54.

Goel, Sharad, Jake M. Hofman, Sebastien Lahaie, David M. Pennock, and Duncan J. Watts (2010), "Predicting Consumer Behavior with Web Search," Proceedings of the National Academy of Sciences of the United States of America (PNAS), 107 (41), 17486-90.

Golder, Peter N., Debanjan Mitra, and Christine Moorman (2012), "What is Quality? An Integrative Framework of Processes and States," Journal of Marketing, 76 (July), 1-23.

Google Trends (2012), http://support.google.com/trends/answer/87284?hl=en&ref_topic=13975.

Hennig-Thurau, Thorsten, Mark B. Houston, and Shrihari Sridhar (2006), "Can Good Marketing Carry a Bad Product? Evidence from the Motion Picture Industry," Marketing Letters, 17(3), 205-19.

Hess, James (1982), "Risk and the Gain from Information," Journal of Economic Theory, 27 (1), 231-38.

Holbrook, Morris B. (1999), "Popular Appeal versus Expert Judgments of Motion Pictures," Journal of Consumer Research, 26 (September), 144-55.

Janiszewski, Chris (1993), “Preattentive Mere Exposure Effects,” Journal of Consumer Research, 20 (3), 376-92.

Kulkarni, Gauri, P. K. Kannan, and Wendy Moe (2012), “Using Online Search Data to Forecast New Product Sales,” Decision Support Systems, 52 (3), 604-11.

Liu, Yong (2006), “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” Journal of Marketing, 70 (July), 74-89.

Mao, Huina, Scott Counts, and Johan Bollen (2011), “Predicting Financial Markets: Comparing Survey, News, Twitter, and Search Engine Data,” arXiv:1112.1051.

Mestyan, Marton, Taha Yasseri, and Janos Kertesz (2012), “Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data,” arXiv:1211.0970v1.

Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), “Consumer Information Search Revisited: Theory and Empirical Analysis,” Journal of Consumer Research, 23 (March), 263-77.

Pavlou, Paul A. and David Gefen (2004), “Building Effective Online Marketplaces with Institution-Based Trust,” Information Systems Research, 15 (1), 37-59.

Punj, Girish N. and Richard Staelin (1983), “A Model of Consumer Information Search Behavior for New Automobiles,” Journal of Consumer Research, 19 (March), 366-80.

Srinivasan, Narasimhan and Brian T. Ratchford (1991), “An Empirical Test of a Model of External Search for Automobiles,” Journal of Consumer Research, 18 (September), 233-42.

Wong, Felix Ming Fai, Soumya Sen, and Mung Chiang (2012), “Why Watching Movie Tweets will not Tell The Whole Story?,” arXiv:1203.4642.

Wu, Lynn and Erik Brynjolfsson (2009), "The Future of the Prediction: How Google Searches Foreshadow Housing Prices and Sales," SSRN Working paper.

Zajonc, Robert B. (1968), "Attitudinal Effects of Mere Exposure," Journal of Personality and Social Psychology, 9 (2), 1-27.

Zeller, Arnold and H. Theil (1962), "Three-State Least Squares: Simultaneous Estimation of Simultaneous Equations," Econometrica, 30 (1), 54-78.

Zhu, Feng and Xiaoquan (Michael) Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," Journal of Marketing, 74 (March), 133-48.

5. Appendix: Constructing Cross-Sectionally Comparable Search Volume Measure from the Google Search Index

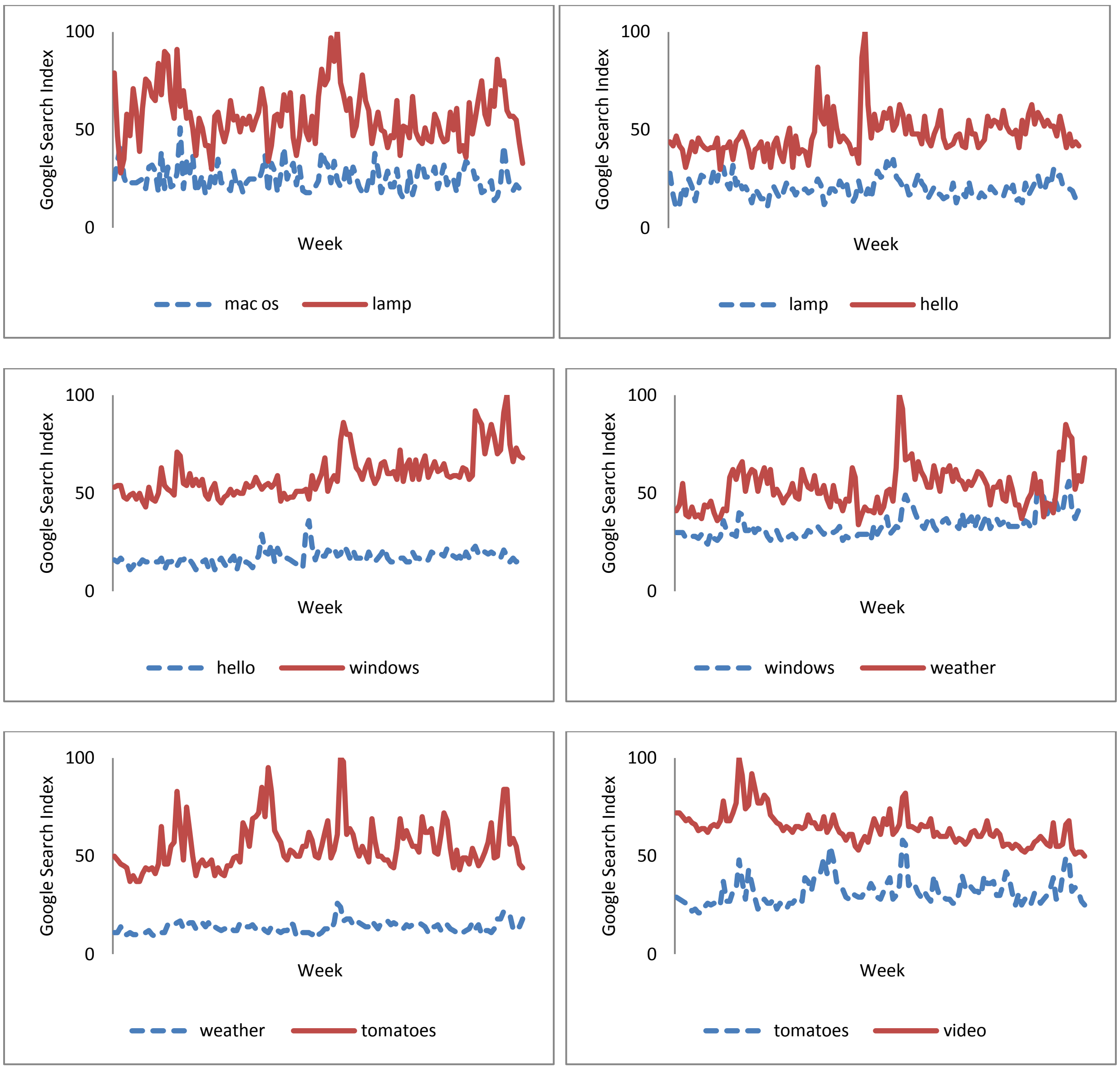
Google Trends provides weekly search indices of keyword queries entered into the Google search engine. Because the index is normalized to conceal the actual search volume of the keyword, researchers cannot compare the search volumes across different keywords if the raw search index is used as provided by Google Trends. In this chapter, we introduce a method to transform the weekly search indices from Google into cross-sectionally comparable search volume metrics. The cross-sectionally comparable search volume metrics in the previous chapters are acquired by applying the following method to the weekly Google search index of the focal movies.

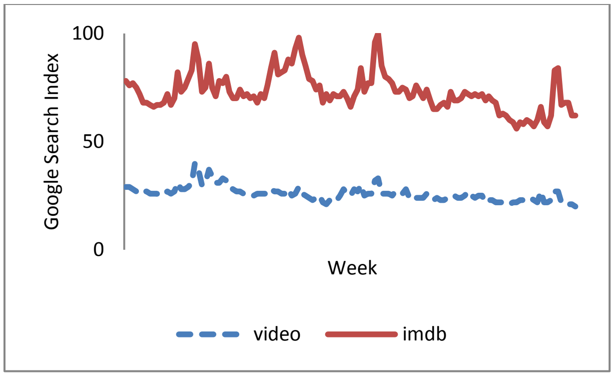
The method consists of three steps. The first is the keyword selection step, where basis keywords and movie keywords are selected. Any set of words can be selected for the basis keywords. The only requirement is that the search volume is neither too high nor too low when compared with the search volume of the focal movies. For our analysis, we select the following seven basis keywords: "mac os," lamp, hello, windows, weather, tomatoes, video, and imdb. They are listed in the order of search amount in the U.S. movie industry, according to Google Trends. That is, among the eight keywords, "mac os" is the least searched keyword and "imdb" is the most searched keyword in the U.S. movie industry. Then, for each movie, we select a set of keywords that are considered to be used by consumers to search the movie. For example, for the movie *12 Rounds*, we choose "12 Rounds" as the keyword for the movie. For the movie *Paul Blart: Mall Cop*, we choose "blart + mall cop," which means either blart or "mall cop."[11] Figure

[11] The selection of movie keywords is guided by the "Related terms" section of Google Trends. The chosen keywords for each movie can be acquired upon request.

5.1 shows the pairwise comparisons of the weekly search volume indices of a basis keyword with its adjacent one from August 2007 to January 2010.

Figure 5.1 Pairwise Comparison of Adjacent Basis Keywords





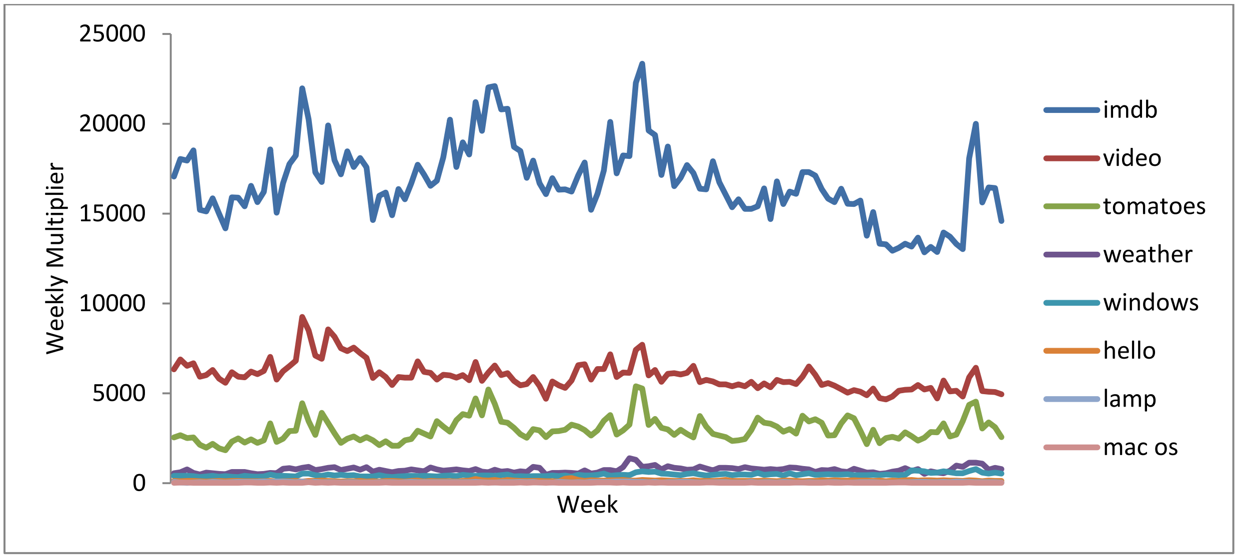
The second step is the keyword matching step. To each movie, we assign an appropriate basis keyword and collect the Google search index of the movie keyword along with that of the assigned basis keyword to the movie. Any basis keyword can be assigned to any movie as long as the search index of the movie keyword is comparable to that of the chosen basis keyword for the movie. That is, if the search volume of a certain basis keyword is too large compared to the search volume of a movie keyword, that basis keyword should not be used for that movie because the movie's search index so collected will be shrunk to zero for many or all of the weeks. Google Trends provides diverse filters to minimize the measurement error in collecting intended search indices. We limit our search so that the search volume is measured only from the U.S. movie industry.

The last step is the transformation step. We transform each movie's search index into our cross-sectionally comparable search volume measure. The mathematics behind this step can be explained as follows. Let k_j be the basis keyword at the j'th position (i.e., k_1 = "mac os", k_2 = lamp, …, k_8 = imdb), and let $I_t^{k_j}$ represent the search index of the j'th basis keyword at week t. We calculate the ratio of the Google search index of two adjacent basis keywords, $r_t^{j,j-1} = I_t^{k_j} / I_t^{k_{j-1}}$, for each t and for all seven pairs of adjacent basis keywords. Let I_t^m be the search index of movie m at week t. Suppose that, in the second step, the basis keyword of position j was assigned to movie m. Then, for movie m at week t, our cross-sectionally comparable search volume measure, denoted by S_t^m, is calculated as in (5-1).

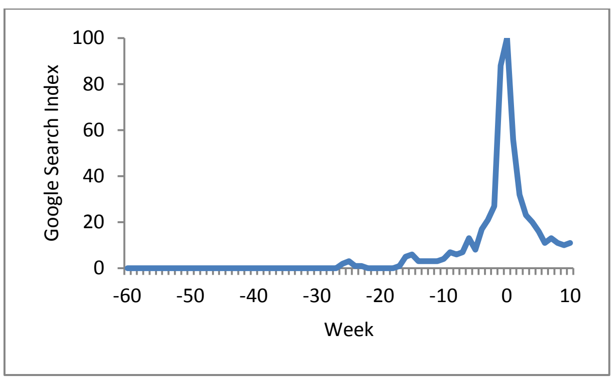
(5-1) $$S_t^m = I_t^m (r_t^{j,j-1} \times \cdots \times r_t^{2,1} \cdot r_t^{1,0})$$

, where $r_t^{1,0}$ is the weekly search index of the basis keyword “mac os” collected together with the keyword “lamp” (the first graph in Figure 5.1). For example, if movie m is compared with the basis keyword of the eighth position (i.e., “imdb”), then $S_t^m = I_t^m (r_t^{8,7} \times \cdots \times r_t^{2,1} \cdot r_t^{1,0})$ for that movie. If movie m is compared with the basis keyword of the first position (i.e., “mac os”), then $S_t^m = I_t^m \cdot r_r^{1,0}$ for movie m. Figure 5.2(a) shows the weekly multiplier associated with each basis keyword, i.e., $(r_t^{j,j-1} \times \cdots \times r_t^{2,1} \cdot r_t^{1,0})$ if the keyword is at the j’th position. For movies *Zombieland* and *X-Men Origins: Wolverine*, Figure 5.2(b) exemplifies the raw search indices of Google Trends and their transformed cross-sectionally comparable search volume measures from 60 weeks before the movies’ releases to 10 weeks after their releases. Note that our transformed search volume measures show a substantial difference in consumer search activities between the two movies.

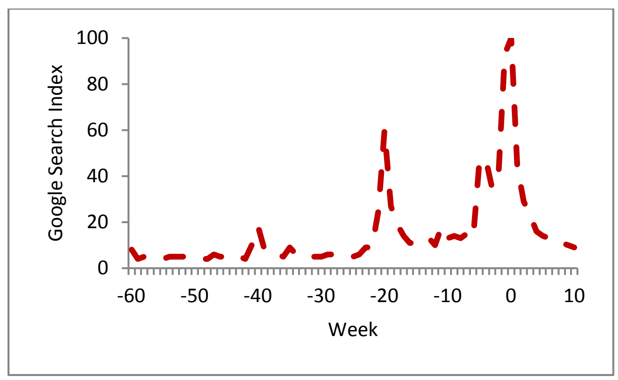
Figure 5.2 Constructing Cross-Sectionally Comparable Search Volume Measure



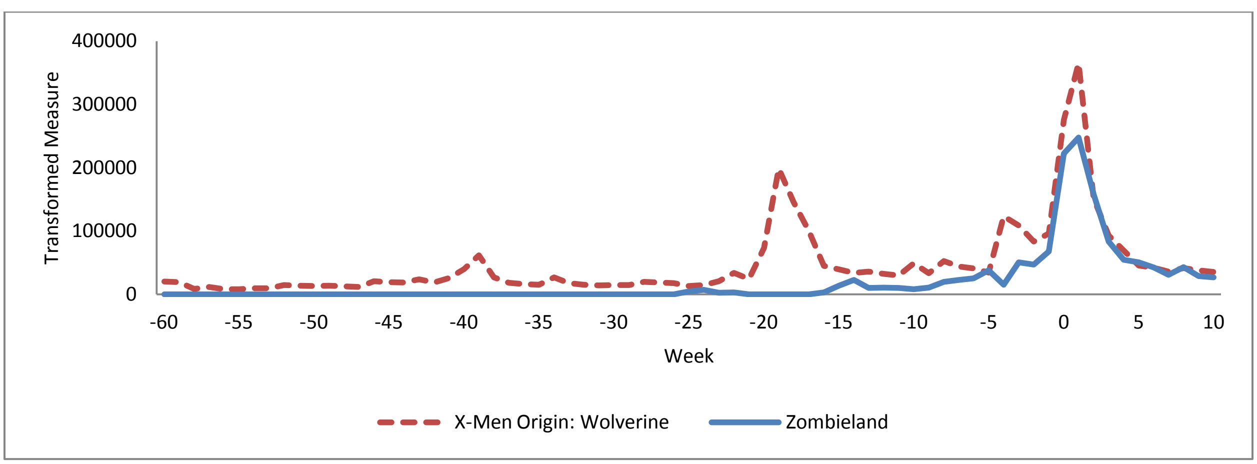
(a) Weekly Multiplier Associated With Basis Keywords



Zombieland



X-Men Origins: Wolverine



(b) Raw Search Indexes and Transformed Search Volume Measures