

ABSTRACT

Title of Dissertation: USING ONLINE SEARCH DATA TO FORECAST NEW PRODUCT SALES

Gauri M. Kulkarni, Doctor of Philosophy, 2010

Dissertation directed by: Professor P.K. Kannan and
Professor Wendy W. Moe
Department of Marketing

This dissertation focuses on online search as a measure of consumer interest. Internet use is at an all-time high in the United States, and according to the Pew Internet & American Life Project, 91% of Internet users use search engines to find information. Consumers' choices of search terms are not well understood. However, we argue that people will focus their searches on terms that are of interest to them. As such, data on the search terms used can provide valuable measures and indicators of consumer interest in a market. This can be particularly valuable to managers in search of tools to gauge potential product interest in a new product launch. In this research, we develop a model of pre-launch search activity. We find search term usage to follow rather predictable patterns in the pre-launch and post-launch periods. As such, we extend our pre-launch search model to link pre-release search behavior to release-week sales – providing a very valuable forecasting tool. We illustrate this approach in the context of motion pictures. Our modeling framework links search activity to

sales and incorporates product characteristics. Our results indicate consistent patterns of search over time and systematic relationships between search volume, sales, and product attributes. We extend our model by studying the role of advertising. This allows us to better understand the relationship between advertising and online search activity and also allows us to compare the forecasting performances of each of the two approaches. We find that search data offers significant forecasting power in opening-weekend box-office revenues. We further find that advertising, combined with search data, offers improved forecasting ability.

USING ONLINE SEARCH DATA TO FORECAST NEW PRODUCT SALES

By

Gauri M. Kulkarni

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:

Professor P.K. Kannan, Co-Chair
Professor Wendy W. Moe, Co-Chair
Professor Roger Betancourt
Professor Siva Viswanathan
Professor Jie Zhang

© Copyright by
Gauri M. Kulkarni
2010

Dedication

*To my parents,
Mukund and Prabha Kulkarni*

Acknowledgements

I would like to acknowledge the contributions of several people who played important roles in the completion of this dissertation. First, I would like to thank my advisors, Dr. P.K. Kannan and Dr. Wendy Moe, for their help, patience, and direction through each step of the research process, and for setting excellent examples of exceptional researchers, as well as valued colleagues. Successful completion of this dissertation would not have been possible without their support.

I would also like to thank my committee members, Dr. Roger Betancourt, Dr. Siva Viswanathan, and Dr. Jie Zhang for their helpful suggestions at various stages of my dissertation. I am also grateful to Dr. Brian Ratchford and Dr. Joydeep Srivastava for their encouragement and guidance throughout my time in the doctoral program.

The support of my fellow students played an instrumental role during my years as a doctoral student. I would especially like to thank Nevena Koukova, Shweta Oza, Peggy Tseng, and Ted Matherly for their mentoring, friendship, and advice, in addition to assistance with research. I would also like to thank my colleagues at Loyola University Maryland for their support during the final stages of completion.

I would like to thank my sister, Manjiree Kulkarni, for her sense of humor and assistance with data entry. Lastly, I would like to thank my parents, Mukund and Prabha Kulkarni, for their unconditional love and unwavering support.

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Chapter 1: Introduction	1
Chapter 2: Literature Review	7
2.1 Overview	7
2.2 Information Search	7
2.2.1 General	7
2.2.2 Online Search	9
2.3 Word of Mouth	11
2.3.1 Overview	11
2.3.2 Online WOM	13
2.4 Advertising	15
2.5 Motion Picture Industry	17
2.5.1 Word of Mouth	17
2.5.2 Advertising	19
2.5.3 Forecasting	21
Chapter 3: Conceptual Framework	23
3.1 Overview	23
3.2 Internet and Search Engines	23
3.3 Motion Picture Context	25
3.4 Pre- and Post-Launch	27
3.5 Conceptual Development	28
3.5.1 Pre-release Search (Product Interest)	28
3.5.2 Post-release Search (Product Interest and Consumption Interest)	29
3.5.3 Purchase (Box-Office Sales)	29
3.5.4 Product Characteristics	30
3.5.5 Competition	30
Chapter 4: Search – Sales Model	32
4.1 Overview	32
4.2 Model Development	33
4.2.1 Product Interest Search	33
4.2.2 Consumption Interest Search	34
4.2.3 Search Volume	35
4.2.4 Box-Office Sales	36
4.2.5 Segment Membership	36
4.3 Data Description	37
4.4 Estimation	41
4.5 Model Fit	41
4.6 Results	45

4.6.1 Segment Structure.....	45
4.6.2 Search Pattern	45
4.6.3 Search Volume.....	46
4.6.4 Sales	47
4.6.5 Movie Covariates	47
4.7 Discussion	48
Chapter 5: Search-Sales Model with Advertising.....	50
5.1 Overview.....	50
5.2 Conceptual Development.....	50
5.3 Model Development.....	51
5.4 Data Description	53
5.5 Estimation Results	55
5.6 Discussion	56
Chapter 6: Forecasting.....	58
6.1 Literature.....	58
6.2 Calibration and Validation.....	58
6.3 Forecasting Results	61
6.4 Discussion	62
Chapter 7: Conclusion.....	64
7.1 Overview	64
7.2 Contribution	65
7.3 Limitations and Future Research	66
Appendix.....	69
References.....	70

List of Tables

Table 1 – Movie Descriptive Statistics	39
Table 2 – Movie Summary Statistics	39
Table 3 – Search Data Summary Statistics	40
Table 4 – Model Fit Based On BIC	42
Table 5 – Search – Sales Model: Estimation Results	43
Table 6 – Weeks of Pre-launch Advertising Expenditures	53
Table 7 – Advertising by Pre-Launch Week.....	54
Table 8 – Advertising Expenditure Summary Statistics	54
Table 9 – Search-Sales Model with Advertising: Estimation Results	55
Table 10 – Search – Sales Model: Calibration Estimation Results	59
Table 11 – Search-Sales Model with Advertising: Calibration Estimation Results	60
Table 12 – Forecasting Results	62

List of Figures

Figure 1 – Search - Sales Model: Conceptual Framework	28
Figure 2 – Search and Sales Pattern for <i>Casino Royale</i>	34
Figure 3 – Weibull Distribution Plot	44
Figure 4 – Exponential Distribution Plot.....	44
Figure 5 – Search - Sales Model with Advertising Framework.....	51

Chapter 1: Introduction

With the advent of the Internet, consumers are able to search for information about virtually anything with the click of a button, often in the comfort of their own homes. The Internet has dramatically lowered the cost of information search, and there has been much work done on this (e.g. Brynjolfsson and Smith, 2000; Ratchford, Lee, and Talukdar, 2003; Johnson et al, 2004). According to recent research, search engine use is a major activity of people who use the Internet, and data on terms that are searched for can be easily obtained. This data is often collected for search engine advertising purposes. It can also have several other useful applications that have not received much attention in the marketing literature. Of particular interest is its use as a measure of word-of-mouth, buzz, effect of advertising, etc – or overall consumer interest.

Both Internet use and search engine use are becoming increasingly common in the United States. Internet penetration in the United States has hit an all-time high with 73% of adults reporting Internet use and 65% of users reporting daily use (PEW/Internet, 2006). Of the Internet users, 91% report using a search engine to find information. This activity is second out of all Internet activities, with using the Internet to send or read e-mail as the most common online activity (PEW/Internet, 2006). Thus, it is clear that both Internet use and search engine use have become a major part of adult life in the United States. The percentage of American Internet users who say “the Internet has greatly improved the way they pursue hobbies and interests” has increased to 33% from 20% in 2001 (PEW/Internet, 2006). A detailed

report on search engine users finds that Internet users have positive online search experiences. In general, search engine users are confident and successful in their searching experiences and feel that search engines are an unbiased source of information (PEW/Internet, 2005). The same report finds that the most common search terms are related to “pop culture, news events, trends, and seasonal topics” (PEW/Internet, 2005). The entertainment or recreation category was sixth in terms of number of online queries in 2002 (PEW/Internet, 2005). Thus, it is reasonable to believe that many consumers search for new product information online using a search engine.

The search literature has differentiated between consumer search for product information when (1) they seek knowledge on specific attributes of products and (2) they seek knowledge on how a particular product compares relative to others in a given product category (e.g. Urbany, Dickson, and Wilkie, 1989; Moorthy, Ratchford, and Talukdar, 1997). Prior work suggests that consumers will still benefit from search if they have knowledge on the offerings of a product but are uncertain about how that product stands relative to others when making a choice. In the case of new products, it is very likely that consumers are uncertain about choice of a product relative to others, since they have no prior experience with the product. Thus, consumer information search for a new product is likely to take place.

In our context, the search we focus on is search term volume, or the number of times particular terms are submitted to online search engines. Given the characteristics of search terms and search engine use that are mentioned earlier, we argue that this measure can capture interest in current trends in pop culture, and new

products. In other words, the terms that consumers choose to submit to search engines indicate their interest or concern for specific topics or products (Ettredge, Gerdes, and Karuga, 2005), and the overall volume of the terms can indicate the general level of interest. Therefore, we propose the use of this measure to predict product sales.

The research objective of this dissertation is to introduce a new measure of consumer interest and evaluate the predictive power of this measure in new product sales forecasting. We aim to answer two questions: (1) Is search term volume a good measure of consumer interest? and (2) Does this measure offer good forecasting power? To the best of our knowledge, we are the first to consider search term volume as a marketing metric. While several recent research studies in marketing have examined various sources of online word-of-mouth, including online conversations, reviews, opinion platforms, etc, (e.g. Dellarocas, 2003; Godes and Mayzlin, 2004) we propose a measure that offers several advantages.

A major advantage of search term data is that one is able to search for a product prior to its launch. Thus, the measure can be obtained during a product's pre-launch phase, allowing for pre-launch forecasting, a task that managers have long struggled with. Search engine use is also a more prevalent online activity than participation in newsgroup conversations, writing online reviews, or blogging. Therefore, search engine data is more likely to be representative of the general online population. Search term volume also does not require analysis of content or great amounts of data cleaning or coding, making it an attractive measure for managers to work with. It can also be collected or obtained with ease and little cost. Changes in

search term volume over time also allow for examination of trends or patterns in consumer interest. Thus, we argue that search term volume offers several advantages to many online measures that have recently been used in new product sales forecasting.

We illustrate our framework by forecasting motion picture box office revenues using online search term activity. Motion pictures are experiential products that have high levels of pre-launch marketing and heavily publicized launch dates. They are common topics of word-of-mouth or everyday conversation. The Internet offers many sources of information on motion pictures during both the pre-launch and post-launch periods. Other product categories that often share similar characteristics are music, video games, and electronics. DVD launches of motion pictures also share many of these features.

We develop a forecasting model that distinguishes between pre-launch and post-launch search, and ultimately forecasts sales. We focus on opening weekend sales only, since these are most critical and also most difficult to predict. We posit that pre-launch search is largely driven by consumer interest in a product, while post-launch search is driven by interest in both the product, as well as consumption of the product. Product interest refers to interest in specific attributes of the product, while consumption interest refers to interest in consuming the product. Thus a consumer may search for information regarding actors/actresses, trailers, plot summaries, etc during the pre-launch phase of a motion picture. A consumer may search for this information during the post-launch as well, but he may also search for information about theater locations, show times, critical/consumer reviews, etc during the post-

launch phase. We incorporate product characteristics in our model, such as genre and MPAA rating, as these are likely to have effects on both search and sales. We also control for competition.

In our base framework, we have not accounted for any drivers of search. Therefore, a possible alternative explanation for the predictive power of search term volume is that it is capturing the response to advertising, and since advertising expenditure data is also available during the pre-launch period, search term volume does not offer a significant gain. In 2006, theater admissions, ticket prices, number of movies released, box-office revenues, and production costs were up from 2005 (MPAA 2006). The average production cost per film for MPAA member companies was \$65.8 million, while the average marketing cost was \$34.5 million (MPAA 2006). These figures suggest that motion pictures remain a growing industry in the entertainment category, and marketing expenditures play an important role in the success of motion pictures. Therefore, we extend our analysis and examine the role of advertising in our framework. Modeling advertising expenditures allows us to determine whether search term volume is capturing the effect of advertising or consumer interest generated from other sources, such as word-of-mouth, above and beyond advertising. It also allows us to compare forecasting performances for the two modeling frameworks.

Our research contributes to the existing literature on online search, online word-of-mouth, and new product sales forecasting. While we choose to illustrate our proposed framework in the motion picture industry, our approach can be generalized to other product categories such as books or movies. For example, search volume

may be used to predict release-week sales for new music albums or books. The pattern and level of search may also be useful in predicting sales in later weeks, since these products tend to have longer product life cycles than most motion pictures. We find that both search volume and search volume pattern over time are important in predicting box-office sales. In doing so, we have developed a forecasting approach that will aid managerial decision-making for new products.

A recent *Wall Street Journal* article (Delaney, 2007) reports that Google “could predict with 82% or higher accuracy based on consumer search activity as early as six weeks before the opening whether a film would top \$25 million in receipts its first weekend.” Thus, it appears that online search activity offers significant explanatory power for predicting box-office revenues for motion pictures, even prior to the movie’s release. We examine this idea in this dissertation.

Chapter 2: Literature Review

2.1 Overview

The focus of this dissertation is to illustrate the effectiveness of online search term volume as a measure of consumer interest in a new product, and then to use this measure to forecast new product sales. We illustrate this framework in the context of motion pictures. First, we briefly review the literature on information search and online search and how it relates to word-of-mouth (WOM) and the motion picture industry. Since we are focusing on online search, we begin with an overview of the literature on information search in general, as well as online search. We also review the literature on WOM, both offline and online, since we posit that online search may be a response to WOM and therefore a measure of WOM activity. We also discuss the literature on advertising, as we go on to examine the role of advertising in our framework. Lastly, because we are illustrating our framework in the context of motion pictures, we discuss prior work in the motion picture area and how it relates to WOM, advertising, and forecasting.

2.2 Information Search

2.2.1 General

Consumer information search is an area that has received much attention in both the economics and marketing literature. Meyer (1982) develops a formal model of consumer information search behavior where he finds that consumers are more likely to search when there is positive information about the product initially, when there is uncertainty about the product, and when the cost of searching for more

information is low. Ratchford (1982) and Hauser, Urban, and Weinberg (1993) take a cost-benefit approach at modeling information search behavior in a constrained utility maximization framework. Keil and Layton (1981) cluster information searchers into groups based on measures of various dimensions. They find a high search group, a low search group, and a selective search group in their analysis. This suggests consumers are heterogeneous in their search behavior.

The literature suggests that consumer uncertainty is an underlying cause of information search. Urbany, Dickson, and Wilkie (1989) differentiate between knowledge uncertainty and choice uncertainty. Knowledge uncertainty refers to uncertainty regarding information about products, while choice uncertainty refers to uncertainty about the best choice. Similarly, Moorthy, Ratchford, and Talukdar (1997) examine consumer information search and brand perceptions. They distinguish between relative brand uncertainty and individual brand uncertainty. “Relative uncertainty is the uncertainty about which brand is the best, whereas individual uncertainty is the uncertainty about what each brand offers.” They find that information search is only necessary when relative brand uncertainty is present for consumers with prior brand perceptions. In the case of new products, there is a likely to be both relative and individual brand uncertainty, particularly during the pre-launch phase, since consumers have yet to experience the product. Thus, the need for more information results in consumer search.

Klein and Ford (2003) study Internet use and pre-purchase search for automobiles. Traditionally, there have been two dimensions to information sources. These are independent vs. seller-dominated and impersonal vs. interpersonal. The

authors suggest inclusion of a third dimension – online vs. offline. Though they focus on automobiles specifically, they find that online search is taking the place of traditional search. This substitution results in higher levels of search overall. We discuss the literature related specifically to online search in the next section.

2.2.2 Online Search

Online search can be any type of information search that takes place on the Internet, and much research has been done on this type of search. Johnson et al (2004) study online search behavior in a retailing context. They find that consumers visit very few sites (on average, less than two) when shopping online for products such as CDs, books, and airline tickets. However, they did find that more active shoppers tend to visit more sites. Bucklin and Sismeiro (2003) model online browsing behavior and find that with repeat visits to an online site, browsers view fewer pages within the site, but the time they spend viewing each page is unchanged. Moe (2003) uses clickstream data to identify online shoppers as buyers, searchers, browsers, or knowledge-builders based on characteristics of their site visit sessions, while Moe (2006) uses clickstream data to develop a two-stage model of the consumer decision process. Wu and Rangaswamy (2003) study online navigation and the formation of consideration sets at an online grocery retailer's site. Brynjolfsson and Smith (2000) and Smith and Brynjolfsson (2001) study price dispersion on the Internet.

While these studies focus on online search that involves web-site visits, we address a different type of search that involves terms submitted to search engines but not any web-site visits. While some research has focused on search engine

performance (e.g. Bradlow and Schmittlein, 2000; Kumar and Lang, 2007), search engines and e-commerce (e.g. Spiteri, 2000; Jansen and Molina, 2006) or search engine visits over time (e.g. Telang, Boatwright, and Mukhopadhyay 2000), the use of search terms as a marketing measure has not received much attention in the literature.

However, a recent paper was able to link online search term data and official unemployment data (Ettredge, Gerdes, and Karuga, 2005). The authors were able to show that job-related search term (jobs, resume, employment, etc) volume is significantly related to official unemployment rates. The premise of their study is “that people reveal useful information about their needs, wants, interests, and concerns via their Internet behavior, and that terms submitted to search engines reveal this information.” Similarly, Deighton and Kornfeld (2008) discuss search engines and “thought tracing.” They suggest that “when search leaves a trail, it is as if curiosity itself is revealed. The search engine knows what is on the person’s mind.” This information can be useful to marketers. “Sometimes the person who searches has consumption on their mind.” (Deighton and Kornfeld, 2008) The same premise underlies our study, as we aim to link online searches and consumption.

We draw from the literature on information search in general, because the underlying reason for information search remains critical to the foundation of our study, and we focus on the online domain, since our study involves search that takes place on the Internet.

2.3 Word of Mouth

2.3.1 Overview

We review the literature on WOM, since this is one possible driver of consumer interest and therefore, online search. Thus, online search may serve as a measure of WOM activity. Also, as we discuss in a later section, it is very relevant to our context of motion pictures. The Word-of-mouth Marketing Associate defines word-of-mouth (WOM) as “the act of consumers providing information to other consumers.” By this definition, WOM is a phenomenon that has been occurring for quite some time and is very relevant to marketing. As Brown and Reingen (1987) point out, “WOM communication plays an important role in shaping consumers’ attitudes and behaviors.” These attitudes and behaviors include those towards new products. Several studies have been done to more closely examine WOM, including aspects such as mediators and moderators, motivations, and consequences.

Bone (1995) looks specifically at WOM and product judgments. She finds that WOM can impact both short-term and long-term product judgments. She further finds that the effect is stronger when a consumer is involved in a disconfirming experience, rather than a confirming experience. The effect is also enhanced when the source of the WOM message is perceived to be an expert. Herr, Kardes, and Kim (1991) consider the vividness of information. They find that WOM communication has a greater impact on judgment than the same information presented in printed format. They suggest the reason for this is that information received in-person is more accessible than other less vivid presentations. However, they go on to find that the effect of WOM can be reduced in the presence of more diagnostic information.

Several other studies also look at various dimensions of WOM effects. These include strong-tie versus weak-tie sources (e.g. Duhan, et al, 1997; Brown and Reingen, 1987) and consumer satisfaction and commitment (Brown et al, 2005). Banerjee and Fudenberg (2004) develop an analytical model of WOM learning, while Goldenberg, Libai, and Muller (2001) examine the WOM process using a complex systems approach. Dodson and Muller (1978) develop an analytical model of diffusion through advertising and WOM, and Arndt (1967) also looks at WOM's role in new product diffusion. He finds that exposure to positive WOM increases purchase probability, while exposure to negative WOM decreases purchase probability.

Holmes and Lett, Jr. (1977) study WOM specifically in the context of product sampling and find that consumers with positive attitudes towards a sampled brand are more likely to spread positive WOM. Richins (1983) focuses on dissatisfied consumers and negative WOM. The author finds that consumers are more likely to spread negative WOM the more serious the problem resulting in the dissatisfaction, the greater the blame on the marketing institution rather than the consumer, and the more negative the consumer's perceptions of responsiveness.

With customer satisfaction being an important component in the area of service products, several studies have also looked at the role of WOM specifically in the service sector. These include Mangold, Miller, and Brockway (1999), Bansal and Voyer (2000), and Harrison-Walker (2001). Haywood (1989) discusses the importance of managing WOM and puts forth guidelines for managers of service businesses.

While several papers have looked at the outcomes of WOM, the motivation of consumers to engage in WOM communications is equally important. Sundaram, Mitra, and Webster (1998) study the motivations of consumers in spreading WOM, both positive and negative. The authors find that the motivations behind positive WOM include altruism, product involvement, self enhancement, and helping the company. The motivations behind negative WOM include altruism, anxiety reduction, vengeance, and advice seeking. They further find that the majority of positive WOM is a result of “satisfying product performance or employee-consumer contact”, while the majority of negative WOM is a result of “inadequate responses to problems with the product and consumers’ poor value perceptions during post-purchase evaluations.”

Godes and Mayzlin (2004) discuss the challenge of measurement in doing WOM research. Surveys were the primary tool for WOM measurement until recently. While traditionally, WOM was generally thought of as consumers communicating with each other directly, the Internet has changed the dynamic of this phenomenon quite drastically, and online WOM has become an important area of research in marketing.

2.3.2 Online WOM

With the increase in Internet use, online WOM is becoming an important area of research in marketing. Some studies have examined the motivations behind consumer participation in online communications – both providing and receiving. Hennig-Thurau and Walsh (2003) focus on the *reading* motivation. They find that consumers choose to read online opinions of products to save time in making

purchase decisions and to make better purchase decisions. On the flip-side, Hennig-Thurau et al (2004) look at motivations behind *sharing* comments on online opinion forums. They find that “consumers’ desire for social interaction, desire for economic incentives, their concern for other consumers, and the potential to enhance their own self-worth are the primary factors.” Phelps et al (2004) look at motivations specific to viral marketing, or the forwarding of e-mails. They find that targeting recipients who find the information relevant and developing messages that evoke strong emotion are necessary for successful viral marketing campaigns. Berger and Milkman (2009) also study viral marketing. Specifically, they study the characteristics of things that lend themselves to be particularly viral.

Another stream of research focuses on online avenues that allow for measurement of WOM. Dellarocas (2003) looks specifically at online feedback systems, which allow consumers to review and/or rate a variety of goods and services. He emphasizes the Internet’s ability to gather information from large groups of people and create WOM communities with ease and little cost. Godes and Mayzlin (2004) look at online conversations as a way to measure WOM. They use this measure to predict TV show ratings. They point out that prior to the Internet, direct observation of private conversations, or traditional WOM, was very difficult. Chevalier and Mayzlin (2006) study the effect of online reviews on book sales. Among their findings is that a negative review has a greater effect on sales than a positive review.

These forums are just some examples of online WOM. While online reviews and conversations can definitely be thought of as WOM activity, we aim to look at a

slightly different measure – online search term volume. We discuss WOM because of the important role it plays in consumers' consumption choices, particularly in our product category – movies. In addition, WOM can be thought of as one possible trigger of online search. Lastly, although our goal in this study is not to measure WOM specifically, the data source that we use is similar to some recent online measures of WOM.

2.4 Advertising

We review the literature on advertising, since one of our research objectives is to examine the role of advertising expenditures in the search activity – sales relationship. Lavidge and Steiner (1961) define three functions of advertising. These are to create awareness and knowledge, to create favorable attitudes or feelings, and to produce action or purchase of the product. In their review paper, Vakratsas and Ambler (1999) develop a general framework of how advertising works. Advertising is considered an initial input for the consumer. The authors propose two levels of responses from the consumer. Firstly, an intermediate response in the form of a mental (conscious or unconscious) response to the advertising occurs. The major intermediate effects are cognition, affect, and experience. These intermediate mental effects then result in a behavioral effect, such as choice or consumption. They suggest that advertising must first have a mental effect before a behavioral effect. In their framework, they allow the consumer's behavior to feed back into the intermediate mental effect of experience. This is particularly relevant for packaged goods where repeat purchases frequently occur. This can also occur with experiential products, such as motion pictures, since past experiences with motion pictures are

likely to influence one's behavior towards future motion pictures. In these cases, the consumer's behavior will be affected by his previous experience with the product, in addition to any advertising that takes place. The authors also include factors, such as involvement and motivation, as mediators of individual responses to advertising. Their framework is general and can be applied to several product categories.

There have been several behavioral studies on various aspects of advertising, including moderators and mediators of its effect. For example, Hoch and Ha (1986) find that advertising has no effect on product quality judgments when consumers have unambiguous evidence on product quality. Other research has focused on attitude formation (e.g. Mitchell and Olson, 1981) and attitude towards the ad (e.g. MacKenzie, Lutz, and Belch, 1986; MacKenzie and Lutz, 1989). Studies have also examined the roles of emotions (e.g. Holbrook and Batra, 1987) and feelings (e.g. Edell and Burke, 1987) in advertising effects. Consumer involvement in both the product (Petty, Cacioppo, and Schumann, 1983) and the advertising (Greenwald and Leavitt, 1984) have also been found to be important moderators of advertising effectiveness.

There is also a stream of research within the advertising literature that focuses on modeling various aspects of advertising. Milgrom and Roberts (1986) develop a model that looks at advertising as a signal of product quality. They find that advertising may signal quality, but this usually occurs in the presence of price signaling as well. Other models of advertising examine areas such as advertising strategy (e.g. Mahajan and Muller, 1986; Sasieni, 1989), price sensitivity (e.g. Krishnamurthi and Raj, 1985; Kaul and Wittink, 1995; Mela, Gupta, and Lehmann,

1997), targeting (e.g. Iyer, Soberman, and Villas-Boas, 2005), and competitive response (e.g. Steenkamp et al, 2005). As of recent, studies related to online advertising have also emerged (e.g. Dreze and Hussherr, 2003; Chatterjee, Hoffman, and Novak, 2003).

An important aspect of advertising research is its *wearout*, or its reduction in effectiveness due to various factors (e.g. Calder and Sternthal, 1980). Naik, Mantrala, and Sawyer (1998) model two types of wearout: repetition wearout and copy wearout. Repetition wearout refers to wearout resulting from excessive advertising, and copy wearout refers to wearout resulting from the passage of time.

While we are not interested in advertising content or advertising strategy in our study, these papers suggest the importance of advertising in a consumer's choice to purchase a product. Thus, while we take advertising expenditures as given, we aim to look at the role of advertising in driving consumer to online search. Advertising creates awareness, so in order to obtain more information about something that consumers are aware of, they may use online search engines.

2.5 Motion Picture Industry

We review the literature on motion pictures and how WOM and advertising play a role in this industry specifically, since both of these can result in online search. We focus on these two drivers of consumer interest, and subsequently online search. We also discuss forecasting approaches that have been studied in the past.

2.5.1 Word of Mouth

There is a stream of literature that focuses on WOM and the motion picture industry specifically. Burzynski and Bayer (1977) study the effect of both favorable and unfavorable WOM on motion picture appreciation. The authors find that subjects exposed to negative WOM before viewing a movie rate film enjoyment significantly lower than subjects who are exposed to positive WOM prior to viewing the movie. The ratings of subjects who were not exposed to any WOM were not significantly different from the ratings of those who were exposed to either positive or negative WOM. The authors' results suggest that movie satisfaction is affected by WOM. In a similar domain, Moul (2007) takes an economic approach to examining WOM's impact in the motion picture industry. He develops and estimates a demand model for movie admissions. He finds that a significant amount of the variance in movie admissions can be explained by WOM, even while controlling for movies' fixed effects.

Liu (2006) looks at an online measure of WOM specifically for movies. The author uses messages from Yahoo Movies to study both volume and valence of WOM. Volume refers to the amount of WOM activity, while valence refers to how positive or negative the WOM activity is. He finds that WOM activities are most prevalent prior to the movie's release and in the opening week. He also finds that WOM data provide significant explanatory power for box office revenues. Further, he finds that volume offers most of this explanatory power, as opposed to valence. This finding leads directly to the motivation of our research, since search term data offers a volume measure, but does not have a valence dimension to it.

As Eliashberg, Elberse, and Leenders (2006) suggest, WOM can be an important factor of product performance in the entertainment industry. They point out two reasons for this. Firstly, these products are often consumed in groups. Secondly, these products are often topics of daily conversations. Their review paper on the motion picture industry gives several examples of research that looks at the relationship between advertising expenditures and box-office revenues (e.g. Zufryden, 1996). They go on to suggest that “the amount of advertising necessary to market a movie is inversely related to the amount of WOM that the movie is likely to generate.” Bayus (1985) suggests that WOM is an indirect effect of marketing activity and a key factor in a consumer’s purchase decision. These marketing efforts include advertising and promotion – two marketing strategies that are heavily employed in the motion picture industry.

Thus, while it is interesting to examine online search activity and the potential for its usefulness as a forecasting measure, it is also important to control for the effect of advertising. As Godes and Mayzlin (2004) point out, it is likely that at least some WOM results from advertising, therefore it is imperative to account for these effects.

2.5.2 Advertising

The relationship between advertising expenditures and motion picture revenues has been widely researched. Zufryden (1996) develops a marketing planning model that focuses on planned advertising expenditures. The author proposes a behavioral framework where advertising expenditures, combined with WOM, affect film awareness. This awareness affects intention to view a film, and ultimately, this intention affects the number of tickets purchased. The model also

incorporates movie characteristics and other marketing variables. Similar to our framework, their model allows for pre-launch forecasting of a film. However, their model differs from ours in that their main objective is to develop a planning model, thus it allows for simulation of various advertising levels. Particularly relevant to our research objective, their model results suggest that anticipated WOM and advertising expenditures have an inverse relationship. In other words, to reach a given performance level, a film with high anticipated WOM may be able to lower its advertising expenditures. Conversely, a film with low anticipated WOM may need higher levels of advertising to obtain a given performance level.

Elberse and Anand (2005) also study pre-launch advertising for motion pictures. They focus on market expectations of sales rather than actual sales. These expectations can be observed before a motion picture is released, and are shown to be accurate predictors of actual sales upon release. They also incorporate product quality in their approach. Their main findings suggest that advertising significantly affects pre-launch expectations, and this relationship is stronger for products of higher quality. The second finding suggests that advertising's effect is not only persuasive, but also informative.

While these studies are most relevant to our research in that they focus on the pre-launch phase of motion pictures and advertising, several other papers also study and find significant relationships between advertising and motion picture performance. These include Bruce and Foutz (2007) and Lehmann and Weinberg (2000) who focus on theater and home video releases, Basuroy, Desai, and Talukdar

(2006) who focus on sequels, Elberse and Eliashberg (2003) who focus on international markets, and Luan and Sudhir (2007) who focus on the DVD market.

2.5.3 Forecasting

As is evident, there has been much work done on the motion picture industry, and several methods have been used to predict the success, as measured by box office revenues, of a motion picture (e.g. Sawhney and Eliashberg, 1996). These include factors such as film critics, star power, budgets, (e.g. Basuroy, Chatterjee, and Ravid, 2003) and more recently, online movie reviews (e.g. Liu, 2006). Neelamegham and Chintagunta (1999) focus on forecasting film performance in domestic and international markets.

The role of critical acclaim has also been examined. Basuroy, Chatterjee, and Ravid (2003) and Eliashberg and Shugan (1997) study the role of critics. Eliashberg and Shugan (1997) find that critical reviews seem to serve as predictors rather than influencers, while results from Basuroy, Chatterjee, and Ravid's (2003) study seem to suggest that critics can serve as both predictors and influencers.

Dellarocas, Zhang, and Awad (2007) explore revenue forecasting of motion pictures in the online domain. They use the valence of online movie ratings and opening weekend revenues as predictors of a movie's future revenues. They find that their method performs better than other approaches that have been previously used, including the movie's marketing budget and critical reviews. Foutz and Jank (2007) focus on pre-release forecasting using another online metric - virtual stock markets. They find that these stock markets offer significant predictive power for a motion picture's opening weekend box-office sales. Pre-launch forecasting has also been

studied in the music industry. Moe and Fader (2002) use advance purchase orders for music to forecast post-launch album sales.

It is well established in the motion picture literature that movie revenues follow predictable patterns of decay in the weeks following their opening weekend (e.g. Krider and Weinberg, 1998). Sawhney and Eliashberg (1996) find that movie revenue patterns can be categorized into three groups, and total revenues can be accurately predicted using revenue data from the first few weeks. Since the opening weekend is most critical for a motion picture's overall performance and is also the most difficult to predict, we focus our analysis on opening weekend revenues in the context of our application.

We discuss the literature on the motion picture industry, since this is the product category we focus on in our study. It is important to emphasize that the framework and forecasting approach we develop can be generalized to other product categories as well. We discuss this idea in later chapters.

Chapter 3: Conceptual Framework

3.1 Overview

In our study, online search term volume refers to the number of times a movie title is searched for using an online search engine. We posit that online search term volume offers a main advantage to measures such as online reviews (e.g. Liu, 2006). Unlike posting a review, a consumer need not have viewed the movie to search for it online. Thus, this activity can be captured several weeks before the movie is released in theaters – the pre-release period. Additionally, since virtually anybody can search for a movie title online, effects of the distribution or availability of the movie do not come into play as much. In other words, anybody can search for a movie regardless of when, where, or how often it is showing. Although the search volume does not capture any content or valence as a review does, we think that the advantage of having more pre-release data outweighs this. We argue that searches indicate consumer interest in a product, and thus we aim to use this measure of interest to forecast sales. We discuss the components of our framework in detail in the next sections.

3.2 Internet and Search Engines

Many would agree that the Internet is at the core of the recent technological revolution. The age of information is associated with the ease of availability of information, largely due to the Internet. Consumers are able to search for information about virtually anything using the Internet. The most basic tool that enables this is an

online search engine (PEW/Internet, 2006). In addition to the Google explosion, search engines such as Yahoo!, MSN, AltaVista, etc. enable a person to enter a keyword or search term and access the most relevant web pages without knowing an exact web-site address. This is becoming so common that the term “Google” is often used as a verb for online search using a search engine, specifically Google. Often times if an issue is disputed or there is uncertainty about something in a casual conversation, one will want to “Google it” to obtain more information. Perhaps one overhears a conversation about a new restaurant, book, movie, TV show, news event, etc. and wants more information. One is likely to search for it online using a search engine of some sort. Since an online search requires an action on the part of the consumer, i.e. entering a search term, there must be a trigger or driver of any given search and the desire for more information. This driver could be WOM, advertising, promotion, news coverage, or any combination of these. Thus, one could suggest that online searches are a measure of interest or “buzz” for a topic or product.

An interesting next step would be to investigate the relationship between these searches and actual consumption of products. In other words, can the number of searches for a given product suggest the level of interest in that product and therefore help predict sales for that product? Given that searches tend to relate to pop culture and trends (PEW/Internet, 2005), this type of relationship is likely to be strongest for products that are new or “trendy” and have heavy pre-launch marketing campaigns. These would include high-technology products, music, and movies. These types of products tend to generate high levels of WOM and often have heavily publicized

launch dates. Thus, we aim to use searches for a new product to predict the sales of that product.

Search term data is easily collected and obtained. It also does not require as much data cleaning or coding as analyses of online messages or conversations, thus it can be used on a larger scale. Additionally, since search engine use is a very prevalent online activity as compared to blogging, participation in newsgroups, etc., the data is less likely to suffer from selection bias and is more representative of the general online population. Search term volume is able to measure consumer interest very early in the consumer decision process (before the consumer has made a purchase decision), as well as in early stages (pre-launch) of the product life-cycle. We explore this stage and its implications for forecasting in a digital context.

3.3 Motion Picture Context

We choose the motion picture industry to explore this idea because movies have relatively short life-cycles and reliable data is easily available. There are also several sources of movie-related information available online (Eliashberg, Elberse, and Leenders 2006). Additionally, motion pictures involve high levels of pre-launch marketing activity that generate consumer interest. Therefore, consumers are able and likely to search for motion picture information prior to release. Other product categories that are also characterized by high levels of pre-launch marketing are music, video games, and electronics. DVD launches of motion pictures and popular press books also exhibit pre-launch marketing. While many of these products have longer product life-cycles than box-office motion pictures, our framework of linking

pre-release search as a measure of consumer interest to post-release sales can still be applied.

Wierenga (2006) points out that consumers and the way they behave is an important area in the literature on motion pictures. The consumer movie decision process is described by “need recognition, search for information, evaluation of alternatives, purchase, consumption, and post-consumption evaluation” (Blackwell, Miniard, and Engel, 2001). The information search stage is most relevant to our work, since this is usually when online search activity will take place. The author points out that WOM is particularly important in this phase (Eliashberg et al, 2000). Eliashberg, Elberse, and Leenders (2006) give several examples of movie-related information sources that are available online. These include chat rooms, Web logs, portals, recommendation sites, customer and critic review sites, official movie sites, and databases. Additionally, consumers may search movie titles to find show times or locations of theaters screening a particular title. Thus, we argue that consumers who are interested in viewing a movie may search for it online using a search engine.

The first week of release is often the most crucial for motion picture revenues, and is also the most difficult to forecast. Therefore, pre-launch forecasting will provide several useful implications to managers. Given the characteristics of online search terms that are mentioned earlier and these characteristics of the motion picture industry, we think it is a good area to begin investigating the potential of search term activity as a forecasting measure. While we use motion pictures to illustrate our forecasting framework, we’d like to emphasize that our approach is generalizable to other new product launches, particularly those that involve high levels of pre-launch

marketing activity and heavily advertised launch dates. Specifically, search term volume and pattern over time can be used as a similar measure of consumer interest for products such as books or movies, and our framework can be applied to forecast release-week sales and extended to predict sales in later weeks.

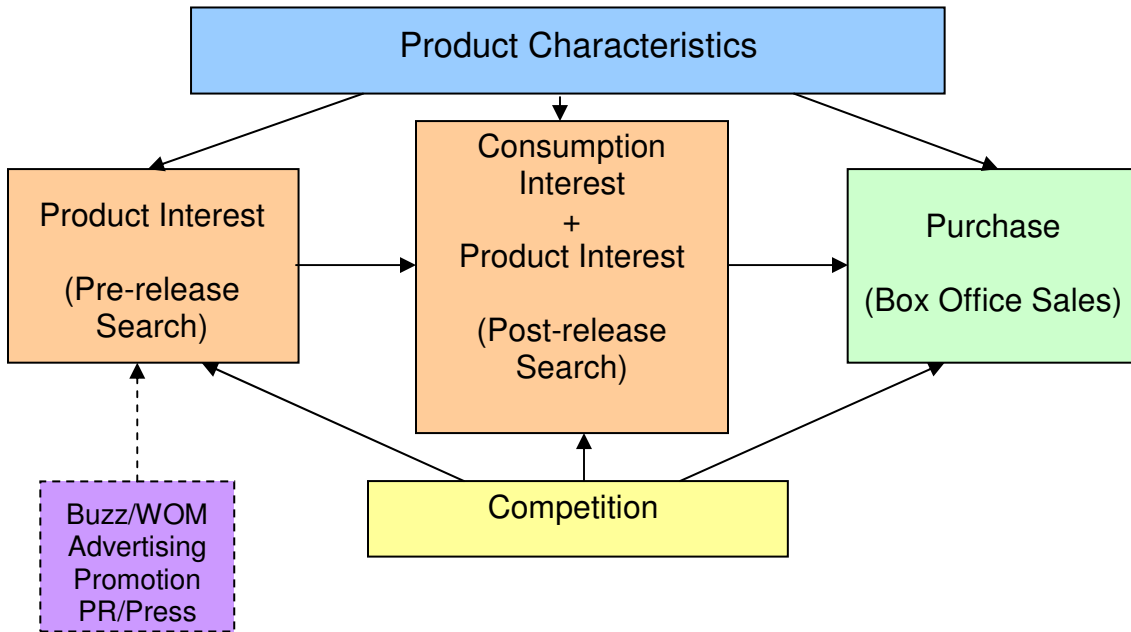
3.4 Pre- and Post-Launch

We consider search for a product in two phases – pre-launch and post-launch. We posit that pre-launch search is largely driven by consumer interest in the product. Consumers may search for general information about the product, features about the product, press releases, etc. Pre-launch information search is likely to be an indication of product interest. This could be a response to WOM, advertising, promotion, or other media coverage related to the product’s upcoming launch. Post-launch search would also include product interest, however it will also be affected by interest in consumption of the product. Thus, after a product is launched, consumers may search for a product with interest in finding information about availability, reading reviews, etc. This is in addition to the increased WOM or buzz that is likely to take place after a product is launched (Liu, 2006), since other consumers have now purchased the product and are able to talk about it. This increase in WOM is also likely to result in increased online search.

In the context of motion pictures, pre-launch search may be driven by consumers’ interest in learning more about the actors or actresses in the movie, viewing trailers, reading media coverage, etc. Post-launch search may be driven by all of these, in addition to need for information about theater locations, show times, consumer reviews, etc. with particular interest in actual consumption or viewing the

movie. Therefore, we posit that pre-launch search is a measure of product interest, while post-launch search is a measure of both product and consumption interest.

Figure 1: Search – Sales Model: Conceptual Framework



3.5 Conceptual Development

We illustrate our conceptual framework in Figure 1 and discuss it in detail in the next section.

3.5.1 Pre-release Search (Product Interest)

We begin with pre-release search. We argue that this indicates consumer interest in the product. The sources of this interest can include exposure to WOM, advertising, promotion or press coverage, although we do not explicitly measure any of these in the current framework. Given that the product is not yet available for consumption or purchase, a search for it suggests some level of interest in it. For

example, a consumer that searches for a movie before it is released is likely to be interested in obtaining information specific to the movie, such as details about actors/actresses, plot or story line, trailer, etc, although the movie is not yet available for purchase or consumption.

3.5.2 Post-release Search (Product Interest and Consumption Interest)

Once the product is launched or released, online search will indicate interest in both the product itself, as well as in consumption of the product. There is likely to be a strong relationship between pre-release search and post-release search, as high levels of product interest are likely to translate into high levels of consumption interest. Therefore, we model the link between pre-release search and post-release search. In our illustration, consumption interest could drive consumers to search for a movie to find information about show times, theater locations, critical/consumer reviews, etc. This information is not likely to be available during the pre-release period.

3.5.3 Purchase (Box-Office Sales)

We further link post-release search to box-office sales. Since we are using search activity as a measure of interest, this interest should translate into purchase of the product. Again, we argue that it is reasonable to believe that the levels of consumption interest and product interest will have a strong relationship with ultimate purchase. Since post-release search is a measure of both product interest and consumption interest, it is likely to be a better predictor of sales than pre-release search alone, which is a measure of only product interest.

3.5.4 Product Characteristics

Characteristics inherent to the product are the main attributes that potential consumers are interested in. We argue that product characteristics can have an impact on both search and sales. In the case of motion pictures, product characteristics include attributes such as genre, MPAA rating, and production budget. Particular genres or MPAA rating levels of movies are likely to generate varying levels of consumer interest. Big-budget films may result in search behavior and sales different from smaller, niche films. Thus we model heterogeneity across products to capture differences in search behavior in both the pre- and post-launch periods. These product characteristics will also play a role in the performance of a given movie once it is released (Sawhney and Eliashberg, 1996).

3.5.5 Competition

It is also important to control for competition or the number of alternatives available for purchase (Liu, 2006). Competition will also affect both search and sales. Consumers are not likely to search for information about every movie in the market. Thus, the more movies there are in the market, the less likely a consumer will search for any given alternative. This is true of both the pre- and post-release periods. The same rationale extends to sales – the more competition there is, the fewer the sales for any given movie.

Our framework aims to link product interest, consumption interest, and product sales while controlling for the effects of competition and product characteristics. We propose online searches as a measure of product and

consumption interests and examine their relationship with sales. We develop a model and discuss our estimation and results in the following chapters.

Chapter 4: Search – Sales Model

4.1 Overview

As discussed in the conceptual framework, we aim to develop a model that links pre-launch search, post-launch search, and sales. Since we differentiate between product interest and consumption interest, we use two separate processes to represent them. We first model searches resulting from product interest as a Weibull process. We choose the Weibull process because of its flexibility in capturing a variety of different patterns. Since search activity typically declines in weeks following the week of launch, we model consumption interest to follow an exponential distribution. Thus, post-launch search represents a combination of product and consumption interest and is therefore the sum of the Weibull and exponential processes. Moe and Fader (2002) use a similar approach in their paper on forecasting music sales.

While the Weibull and exponential processes capture pattern of search, the level or volume of search is also very important. Therefore, we incorporate a penetration rate to account for the volume of searches.

There is likely to be heterogeneity across the data sample as well as in the relationships between the components of our framework. Therefore, we use a latent-class segmentation approach (Kamakura and Russell, 1989) to segment the movies, where the probability of belonging to any given segment is determined by search, sales, and product characteristics/competition.

We model sales to follow a normal distribution. We discuss each of these components in detail in the next section.

4.2 Model Development

4.2.1 Product Interest Search

We begin by modeling the search pattern as a Weibull process. We choose the Weibull for its flexibility in capturing various shapes, such as the ones we observe in the data. The hazard function, $h(t | s)$, survival function, $S(t | s)$, and cumulative distribution function, $F(t | s)$, for each segment s are as follows:

$$(1) \quad h(t | s) = \lambda_s c_s t^{c_s - 1}$$

$$(2) \quad S(t | s) = e^{-\lambda_s t^{c_s}}$$

$$(3) \quad F(t | s) = 1 - S(t | s) = 1 - e^{-\lambda_s t^{c_s}}$$

where $t =$ week in movie's launch period ($t=1,2,\dots,T$)

$\lambda_s =$ slope parameter for search in segment s

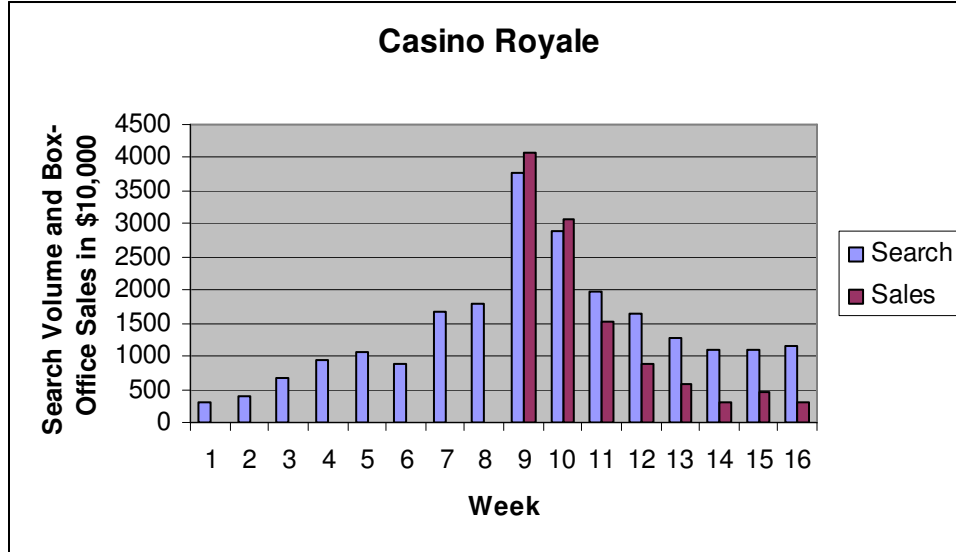
$c_s =$ shape parameter for search in segment s

Thus, the probability of search occurring in any given week t for a given segment s is:

$$(4) \quad \begin{aligned} P(t | s) &= F(t | s) - F(t - 1 | s) \\ &= e^{-\lambda_s (t-1)^{c_s}} - e^{-\lambda_s t^{c_s}} \end{aligned}$$

We posit that pre-launch search captures only consumer interest in the product, as consumption of the product is not yet feasible. In general, our search data shows a non-linear growth trend as the week of launch approaches, and the Weibull performs well in modeling this pattern over time. Figure 2 illustrates a typical pattern of search.

Figure 2: Search and Sales Pattern for *Casino Royale*



4.2.2 Consumption Interest Search

In our conceptual framework, we differentiate between pre-launch and post-launch search. We argue that pre-launch search indicates product interest, while post-launch search indicates both product interest and consumption interest. As seen in Figure 2, our data shows a surge in searches occurring in the week of a movie's launch, representing both product interest and consumption interest. We model post-launch search in subsequent weeks in order to capture the peak in searches that typically occurs in the week of release and the decline in searches that typically follows. We model consumption interest as an exponential process with probability density function and cumulative distribution function as follows:

$$(5) \quad f_e(\tau) = \lambda_{es} e^{-\lambda_{es}\tau}$$

$$(6) \quad F_e(\tau) = 1 - e^{-\lambda_e\tau}$$

where τ = week of post-release, beginning with the first week of release. It is important to note the differentiation between the λ parameters – from this point onwards we will use λ_w to refer to the Weibull component and λ_e to refer to the exponential component. It is also important to note the different time periods. $t = 1$ refers to the first week in the movie's launch period, including pre-launch, for the Weibull process, while $\tau = 1$ refers to the first week of launch of the movie (post-launch only) for the exponential process. Therefore $f(\tau) = 0$ for the exponential process during the pre-launch period.

We incorporate an inflation parameter to capture the consumption interest aspect of the search pattern. Thus, the probability of observing a search at time t for a given segment s can be represented as:

$$(7) \quad P(t | s) = (1 - \phi_s)[F_e(\tau | s) - F_e(\tau - 1 | s)]I_t + \phi_s[F_w(t | s) - F_w(t - 1 | s)]$$

where $\tau = t - 8$, since we have 8 weeks of pre-launch search. F_e represents the cdf of the exponential process (consumption interest) and F_w represents the cdf of the Weibull process (product interest). $I_t = 1$ for post-launch weeks ($I_t = 0$ otherwise), ϕ_s represents the proportion of search resulting from product interest, and $(1 - \phi_s)$ represents the proportion of search resulting from consumption interest, which is present only in the weeks that the product is available for consumption.

4.2.3 Search Volume

It is reasonable to believe that volume, not only pattern, of search activity is also important when trying to capture consumer interest in a product. Therefore, we

incorporate a penetration rate, α_s , to capture the level of search activity. Our probability of a search is now as follows:

$$(8) \quad P(t | s) = \alpha_s \{ (1 - \phi_s) [F_e(\tau | s) - F_e(\tau - 1 | s)] I_t + \phi_s [F_w(t | s) - F_w(t - 1 | s)] \}$$

where again $\tau = t - 8$. In order to measure penetration, it becomes necessary to incorporate the size of the potential market. Thus, we need to account for the non-searchers. In our case, we will define the non-searchers as:

$$(9) \quad non - search_i = M - \sum_{t=1}^T sv_{it}$$

where M is the market size and sv_i is the search volume for movie i at time t . The market size represents the total number of potential searches that could occur.

4.2.4 Box-Office Sales

Our objective is to examine the relationship between online searches and opening weekend box office sales. Therefore, we specify box-office sales to follow a normal distribution with probability density function as follows:

$$(10) \quad f_n(x_i | s) = \frac{1}{\sigma_s \sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu_s)^2}{2\sigma_s^2}\right)$$

where x_i is the natural log of opening weekend box-office sales for movie i , μ is average opening weekend box-office sales, and σ is standard deviation of opening weekend box-office sales.

4.2.5 Segment Membership

In order to account for heterogeneity, we segment the movies using a latent-class segmentation approach. We specify membership probability in segment s , π_s , to be a function of product characteristics (Gupta and Chintagunta, 1994). In our

context, product characteristics refer to movie characteristics. We incorporate these covariates as follows:

$$(11) \quad \pi_{is} = \frac{e^{\theta_{is}}}{1 + \sum_s e^{\theta_{is}}} \quad \text{where } \theta_{is} = \beta_{is} Z_{is}$$

where Z_{is} is a vector of covariates that include an intercept and the following:

$PG13_i$ = dummy variable for MPAA rating “PG13”

R_i = dummy variable for MPAA rating “R”

$Comedy_i$ = dummy variable for genre comedy and romantic comedy

$Drama_i$ = dummy variable for genre drama

$Comp_i$ = number of other movies in the dataset released in the same week

The dummy variables take the value of “1” if the movie belongs to that category and “0” otherwise. Our omitted category for MPAA rating is “PG,” and our omitted category for genre is “ActAdvHor,” action, adventure, and horror. We do not include production budget as a covariate in our estimation because production budget and advertising budget, which we include later, are very highly correlated. However, we do provide summary statistics on it in the next section for the sake of information. Lastly, we link the search and sales components using the segment membership probability. We discuss our estimation procedure and empirical analysis in the next sections.

4.3 Data Description

We create our dataset by merging two types of data – one consisting of search data and the other consisting of movie-related variables. We obtain the search data from a search term research service that collects and compiles search term data. The service maintains a database of search terms collected from all of the major search

engines, including Google, Yahoo!, MSN, AltaVista, Ask.com, Lycos, AOL, HotBot, Information.com, and Dogpile. Search term volume is available at both the weekly and monthly levels and refers to the number of times the term is searched for within the given time period. For example, for the term “ipod,” one could obtain the number of searches on that term for a given week or month. The data is available for a period of twelve months, making it possible to see trends or patterns of search over time. The data comes from a database based on user panel data and is free from skew caused by automated agents. It contains data on over 4.3 billion searches. We obtain the data at the aggregate level and do not have demographic information on the user panel. Specifically, we obtain the weekly search volume for each movie title in our analyses during both the pre-launch and post-launch phases. In other words, we have the number of times each movie title is searched for on a weekly basis. Our analysis focuses on 16 weeks of search data – 8 weeks of pre-launch and 8 weeks of post-launch.

Our dataset includes movies released from July to November of 2006 in the United States. There were 599 new feature films released in the US in 2006, and 63 grossed more than \$50 million in box office revenues (MPAA 2006). We limit our set of titles to those that were in the top fifty in box office revenues during their opening week. After removing titles for which we were not able to obtain all variables or had very sparse search data, we conduct our analyses on 63 movie titles. We obtain motion picture data from two popular movie sites, The Numbers (<http://www.the-numbers.com>) and Yahoo Movies (<http://movies.yahoo.com/>). We collect box office release date, weekend box office revenues, MPAA rating, and

genre for each movie title in our analyses. We use the box-office release date to determine the competition level. We define competition as the number of other movies in our dataset that are released in the same week as a given movie. We collect production budget data from The Internet Movie Database Pro (<http://www.imdb.com/>). Summary and descriptive statistics on the movie sample can be found in Tables 1 and 2. Summary statistics on search activity for the movie sample can be found in Table 3. “SVPreSum” indicates the sum of searches in all pre-release weeks, while “SVRel” refers to the number of searches in the week of release.

Table 1: Movie Descriptive Statistics

Categorical Variables	Category	Proportion	Description
Genre	Comedy	0.270	Comedy or Romantic Comedy
	ActAdvHor	0.397	Action, Adventure, or Horror
	Drama	0.333	Drama
MPAA Rating	PG	0.127	Rated 'PG'
	PG13	0.429	Rated 'PG13'
	R	0.444	Rated 'R'

Table 2: Movie Summary Statistics

Variable	Min	Max	Mean	SD
Production Budget	\$400,000	\$225,000,000	\$36,839,683	\$42,329,493
Opening Weekend Sales	\$33,316	\$135,634,554	\$13,283,776	\$19,339,359
Competition	1	5	2.86	1.19

Table 3: Search Data Summary Statistics

Average				
Categorical Variables	Category	SVPreSum	SVRel	Description
Genre	Comedy	2346	1704	Comedy or Romantic Comedy
	ActAdvHor	3019	2044	Action, Adventure, or Horror
	Drama	1922	1023	Drama
MPAA Rating	PG	1944	1435	Rated 'PG'
	PG13	2127	1293	Rated 'PG13'
	R	2955	1970	Rated 'R'
Overall		2472	1612	

We have three categories of genre. *Comedy* refers to comedy or romantic comedy movies. *ActAdvHor* refers to movies that are action, adventure, or horror. *Drama* represents drama. Each of the genre categories is fairly equally represented, with action, adventure, or horror being the largest. We also have three categories of MPAA ratings, PG, PG13, and R. Our dataset does not contain any movies rated G or NC17. Movies rated PG are fewest in number in our dataset.

We can see that there is a great deal of variance in both production budget and opening weekend sales. The number of competing movies ranges from one to five, with the average being about three.

In our analyses, we focus only on the opening weekend box office sales. We do so because box office sales tend to follow fairly predictable patterns in subsequent weeks, and these can often be determined from the first week's sales.

As an example, Figure 2 shows the search and sales patterns for one movie in our sample, *Casino Royale*. This figure illustrates the usual pattern of search volume over time, and the relationship between search volume and opening weekend sales. Search tends to follow a non-linear growth pattern in the week up until the release

week, then surges in the week of release. Box-office revenues tend to peak in the opening week and then exponentially decline in subsequent weeks.

4.4 Estimation

We estimate our model using Maximum Likelihood Estimation on a sample of 63 movies. We first separate the search and sales component of our probability statement as follows:

$$(12) P_{i,s,search}(t|s) = [\alpha_s\{(1 - \phi_s)[F_e(\tau|s) - F_e(\tau - 1|s)]I_t + \phi_s[F_w(t|s) - F_w(t - 1|s)]\}]$$

$$(13) P_{i,s,sales}(x|s) = [f_n(x_i|s)]$$

where F_e is given in equation (6), F_w is given in equation (3), and π_{is} is given in equation (11). $\tau = t-8$ as before.

Our likelihood function is then given by:

$$(14)$$

$$L = \sum_s \pi_{is} [\prod_i \prod_t P_{i,search}(t|s)^{svit}] [\prod_i [1 - \sum_t P_{i,search}(t|s)]^{non-search_i}] [\prod_i P_{i,sales}(x|s)^{sales}]$$

Note that we account for the non-searchers in L_{search} . Our search data comes from a user panel, and we do not have information on the size of the panel. Therefore, we specify $M=100,000$, since none of the movies in our sample has a sum of search volume greater than 100,000. The highest number of searches we have is 75,419 summed over 16 weeks for the movie *Borat*. (See Appendix for estimation results for various values of M .) We note that one limitation of our framework is the independence of the search and sales components in our likelihood function.

We discuss the estimation results in the next sections.

4.5 Model Fit

We begin by determining the best model fit. We first estimate the model without any movie characteristics using a latent class segmentation approach that minimizes the Bayesian Information Criterion (BIC). The BIC is given by $BIC = -2LL + k\ln(N)$, where LL is the log-likelihood, k is the number of parameters, and N is the sample size. The BIC measure penalizes overparameterization (Schwarz, 1978) and is commonly used to compare latent class segment models. We then estimate the model, incorporating movie covariates, also using the latent class segmentation approach. We begin with a no-covariate model to examine the importance of incorporating heterogeneity in product characteristics. The results can be found in Table 4. (We will discuss the last set of results, “With Advertising,” in a later chapter.)

Table 4: Model Fit Based on BIC

	No Covariates		With Covariates		With Advertising	
	BIC	LL	BIC	LL	BIC	LL
1 segment	6.31496E+06	-3.15745E+06	-	-	-	-
2 segments	6.29550E+06	-3.14770E+06	6.20380E+06	-3.10183E+06	6.10333E+06	-3.05158E+06
3 segments	6.29522E+06	-3.14753E+06	6.20389E+06	-3.10183E+06	6.10345E+06	-3.05158E+06
4 segments	6.29528E+06	-3.14753E+06	-	-	-	-

We find that a two-segment model with movie covariates performs the best. We can see from the results that movie covariates contribute substantially, as the optimal number of segments in the no-covariate model is three. The improvement in both the BIC and log-likelihood is also substantial after controlling for movie effects. These results highlight the importance of incorporating product characteristics.

The estimation results of the two-segment model can be found in Table 5.

The plots for each of the distributions (Weibull and exponential) for search can be found in Figures 3 and 4.

Table 5: Search – Sales Model: Estimation Results

Parameter		Segment 1		Segment 2	
ϕ	<i>Inflation Parameter</i>	0.798	(0.007)	0.769	(0.002)
λ_w	<i>Weibull Slope Parameter</i>	0.007	(0.000)	0.006	(0.000)
c	<i>Weibull Shape Parameter</i>	1.798	(0.018)	1.960	(0.010)
μ	<i>Normal Mean</i>	15.073	(0.412)	15.389	(0.288)
σ	<i>Normal Standard Deviation</i>	2.128	(0.093)	1.888	(0.095)
λ_e	<i>Exponential Parameter</i>	0.494	(0.010)	0.710	(0.004)
α	<i>Penetration Rate</i>	0.038	(0.001)	0.126	(0.001)
β_0		-212.570	(0.961)	-	-
β_{PG13}		-2.448	(0.031)	-	-
β_R		-43.223	(0.718)	-	-
β_{Comedy}		164.691	(0.732)	-	-
β_{Drama}		-173.948	(0.896)	-	-
β_{Comp}		61.188	(0.335)	-	-
Standard errors are in parentheses					
All estimates are significant at the .05 level					

Figure 3: Weibull Distribution Plot

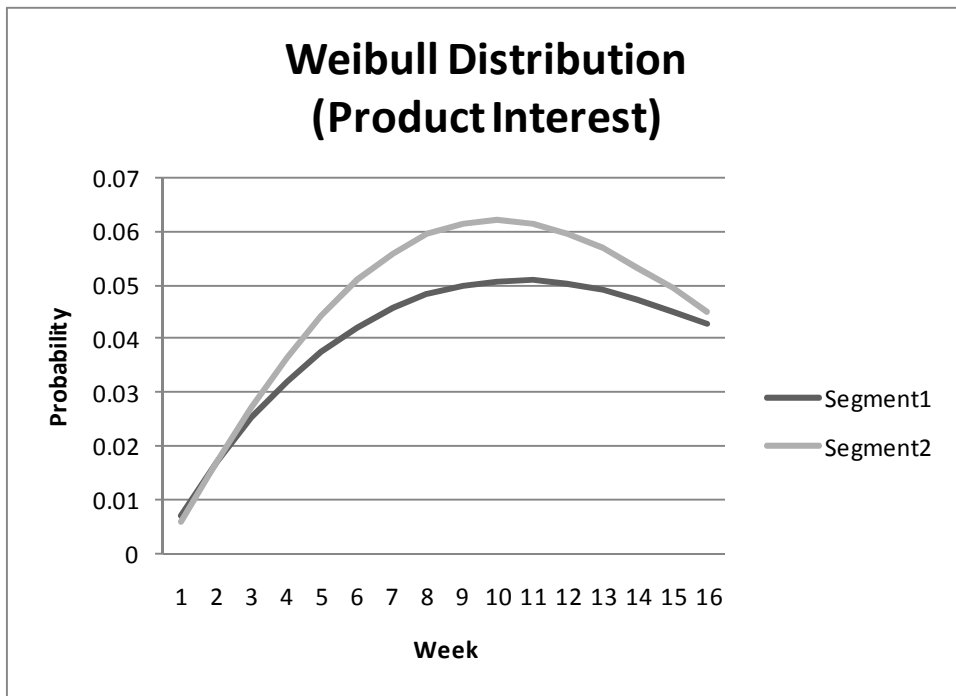
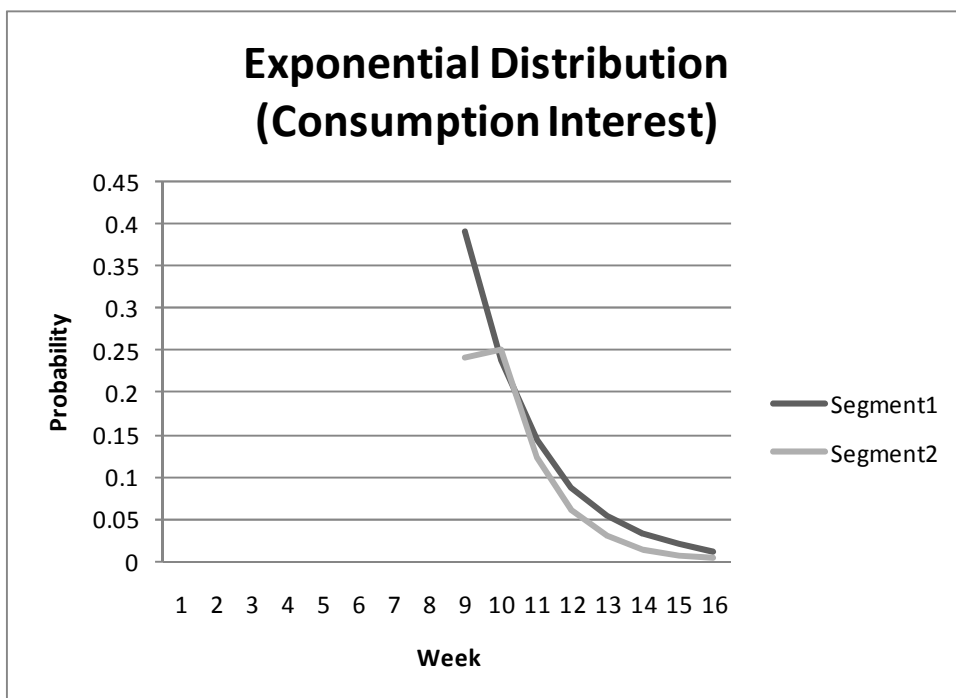


Figure 4: Exponential Distribution Plot



4.6 Results

We discuss the estimation results for each of the parameters in the two segments in detail.

4.6.1 Segment Structure

Our estimation results indicate a well-defined segment structure. We find that search pattern, overall, does not differ much across the two segments. However, the two segments differ in terms of search volume, sales, and movie attributes. Segment 1 consists mainly of comedy films. These movies also tend to exhibit lower levels of search activity (penetration) and lower opening-weekend box-office revenues. The movies in segment 1 also exhibit higher levels of variance in sales. Segment 2 consists mainly of films with MPAA rating of “R” and dramas. These movies are perhaps of more popular genres. These movies also exhibit higher levels of search activity and opening-weekend box-office revenues, and smaller variance in opening-weekend box-office revenues. We discuss the estimation of each of the parameters in detail in the next section.

4.6.2 Search Pattern

The parameter estimates provide some insight into the characteristics of the two segments. First, we begin with (ϕ) , which represents the proportion of search attributed to the Weibull process or product interest. The estimates for (ϕ) , .798 for segment 1 and .769 for segment 2, suggest that most of the search can be attributed to the Weibull process, which we specify as capturing consumer interest in the product.

This suggests that even in the post-release period, most of the search is driven by an interest in the product.

Next we focus on the slope (λ_w) and shape (c) parameters from the Weibull process, which represents product interest. These parameters capture the pattern of search over time. Since the level of interest may differ across time, it is important to model this aspect of search. The slope parameter for segment 1 is larger than that of segment 2, while the shape parameter for segments 2 is larger than that of segment 1. From a practical perspective, these parameters are not substantially different across the segments, suggesting that the pattern of product interest is fairly similar across the movies in our sample. Figure 3 illustrates the pattern captured by the Weibull process for both segments.

Lastly, we look at the parameter for the exponential process (λ_e). This parameter is capturing the consumption interest that occurs during the post-launch period. This parameter is fairly different across the two segments, suggesting that the pattern of consumption interest may differ across the movies in our sample. Figure 4 illustrates the pattern of post-launch search captured by the exponential process for segments 1 and 2. We can see that the pattern differs initially, but overlaps for some of the later weeks for the two segments. Thus, we can conclude that search *patterns*, in both the pre- and post-launch periods, are rather predictable and consistent for motion pictures.

4.6.3 Search Volume

We incorporate search volume using a penetration rate (α). Since the level of search can vary drastically across movies, it is important to incorporate volume of

search. The parameter estimates for the penetration rates (α) are very different across the two segments. The results indicate that movies with higher levels of search volume tend to belong to segments 2, while movies with lower levels of volume tend to belong to segment 1. Higher levels of search are associated with movies with higher levels of sales. These results suggest the importance of modeling not only search pattern, but also search *volume*.

4.6.4 Sales

The sales component of the model consists of two parameters, (μ) and (σ), where (μ) is the mean and (σ) is the standard deviation of opening weekend box-office sales. These two parameters allow us to capture both average sales, as well as the dispersion in sales. The parameter estimates for these parameters (μ and σ) also differ across the two segments. Segment 1 is characterized by movies with lower average sales and higher variance, while segment 2 is characterized by movies with higher average sales but lower variance. The results suggest that we have identified a relationship between search (particularly search volume) and sales. Movies that have higher levels of search also tend to have higher levels of opening-weekend box-office revenues, which is fairly intuitive. Higher levels of consumer interest will generate higher sales.

4.6.5 Movie Covariates

Lastly, we discuss the results for the movie covariates. These covariates are used to segment the movies. The results for the coefficients of the movie parameters are very interesting. Firstly, the results of the segmentation indicate that the movie

covariates contribute substantially to model fit. They also help to characterize segment membership using product attributes.

Movies rated R are likely to belong to segment 2, which is also characterized by high sales and high search. Movies rated PG13 are also likely to be in segment 2. Comedies and romantic comedies (*Comedy*) likely belong to segment 1, while dramas (*Drama*) likely belong to segment 2. It is not surprising to find that segment 2 consists mainly of both dramas and R-rated films, as these types of movies tend to be popular with mainstream consumers.

The coefficient for the competition (*Comp*) measure is positive. This suggests that movies released in weeks of higher competition are more likely to belong to segment 1.

4.7 Discussion

These results provide some insights into the performance of our modeling framework. The significant differences in parameters for search volume, sales, and movie covariates illustrate the importance of segmenting the market to account for heterogeneity. The result that search pattern does not differ dramatically across the two segments suggests that search trend is fairly predictable in both the pre- and post-launch periods. Thus, if the pattern deviates from the norm, perhaps a manager can respond to improve both consumer interest in the product, as well as sales.

The characteristics of the two segments are also interesting. Segment 1 seems to consist of films that are associated with lower levels of search and sales, while segment 2 seems to consist of more popular genres that tend to be associated with higher levels of search and sales. For segment 1, lower sales and higher variance in

sales suggests that smaller, niche films are likely to belong to this segment. They cover a wider range, in terms of box office sales, and also do not generate as much in sales. Also, the releases of larger, mainstream films are often timed to not compete directly with other big films, thus the positive coefficient on competitions for segment 1 is logical. Smaller films are more likely to be released in weeks of higher competition. Segment 2 films are what are typically considered popular blockbusters. High levels of consumer interest (as measured by search penetration rate), high sales, and less variability in sales characterize these films. They are also of popular, well-performing MPAA ratings (PG13 and R) and genres (dramas).

The results imply that different types of movies generate different levels of interest. Thus, perhaps managers can use search volume as a marketing metric and adjust marketing mix variables to enhance consumer interest and sales. For example, if consumer interest is higher or lower than expected for a product of given characteristics, managers can increase, decrease, or otherwise change their marketing strategy accordingly.

Managers can also use our framework to forecast sales. We will discuss the forecasting procedure and performance of our model in a later chapter.

Chapter 5: Search-Sales Model with Advertising

5.1 Overview

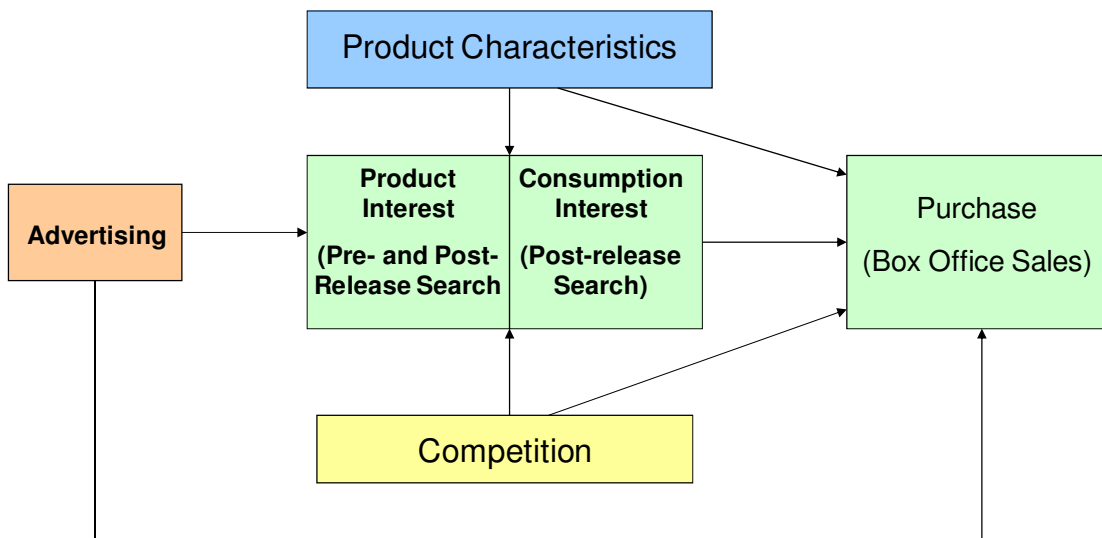
Thus far, we have developed a model to examine the effectiveness of online search activity as a measure of consumer interest in a product, particularly during the pre-launch phase. We will use this measure of consumer interest to predict post-launch sales of the product. However, we have not explicitly accounted for any drivers of this consumer interest. While we argue that online search can serve as an all-encompassing measure of interest stemming from a variety of sources, including WOM and marketing campaigns, an interesting extension would be to study the role of a product's advertising, since it is likely that at least some search activity is triggered by advertising. Since we are illustrating our framework in the context of motion pictures, where advertising budgets can be substantial, it is particularly important to examine the effect of advertising.

5.2 Conceptual Development

Our conceptual framework for the advertising expenditures parallels that of the search – sales model discussed earlier. However, we do not distinguish between pre- and post- launch advertising. We are interested in the relationship between advertising expenditures, online search, and sales. We argue that advertising can have a direct effect on sales and also an indirect effect on sales through online search. In other words, one of the triggers of online search may be exposure to advertising. Also, as is evidenced in the literature, advertising has also been shown to have a

relationship with box-office sales. There are likely to be systematic differences in advertising expenditures of movies of various genres and MPAA ratings, so we again include those as covariates. We do not include production budget as a covariate in the model, due to the high degree of correlation between production budget and advertising expenditures. Competition, in the form of other movies released at the same time, is also likely to play a role in advertising strategy of a given movie, so we also include this measure as a covariate. Our conceptual framework can be found in Figure 5.

Figure 5: Search-Sales Model with Advertising Framework



5.3 Model Development

Our modeling framework extends the model developed in the previous chapter to include advertising expenditures. We incorporate advertising in two ways – (1) an indirect effect on sales through search and (2) a direct effect on sales. Our existing model is defined as follows (equations 12 and 13):

$$(12) P_{i,s,search}(t|s) = [\alpha_s \{ (1 - \phi_s) [F_e(\tau|s) - F_e(\tau - 1|s)] I_t + \phi_s [F_\omega(t|s) - F_\omega(t - 1|s)] \}]$$

$$(13) P_{i,s,sales}(x|s) = [f_n(x_i|s)]$$

where α_s is the penetration rate. The penetration rate is used to capture search volume. Thus, we introduce an effect of advertising on search by specifying the penetration rate to be a function of advertising.

$$(15) \alpha_{is} = \frac{e^{a_{0s} + a_{1s} Adv_i}}{1 + e^{a_{0s} + a_{1s} Adv_i}}$$

This formulation specifies penetration rate (α_{is}) as a logit transform of an intercept-term capturing the baseline penetration rate (a_{0s}) and a coefficient (a_{1s}) of advertising (Adv_i) to capture the effect of advertising on search penetration. Secondly, we incorporate the direct and indirect effect of advertising on sales. As previously discussed, we specify sales to follow a normal distribution with μ (mean) and σ (standard deviation). We define μ (the mean of the normal distribution) to be a function of the baseline penetration rate (a_{0s}) and advertising (Adv_i) as follows.

$$(16) \mu_{is} = m_{0s} + m_{1s} * a_{0s} + m_{2s} * Adv_i$$

This formulation allows for both of the effects, direct and indirect (through the penetration rate) that we previously discussed. In other words, the direct effect of advertising on sales is captured through the coefficient (m_{2s}) of advertising (Adv_i) in equation 16. The indirect effect of advertising on sales through search is captured through the coefficient (m_{1s}) on baseline penetration (a_{0s}), where baseline penetration is specified in equation 15. We use baseline penetration, rather than the actual penetration rate, to avoid multicollinearity between the penetration rate and the advertising expenditures.

We describe our advertising data and estimation results in the next sections.

5.4 Data Description

Our data consists of advertising expenditures for each of the movies in our dataset for both the pre- and post- launch periods on a weekly basis. The expenditures reflect the week that the advertising occurred and not the week that the payment was made. The data includes advertising expenditures across television, radio, magazines, newspapers, Internet, and outdoors. Similar to our approach in the search volume framework, we consider advertising expenditures up to eight weeks prior to the motion picture’s release. We find that for many of the movies in our dataset, advertising expenditures do not occur until much closer to the movie’s release date. A summary of the pre-launch weeks of advertising expenditures can be found in Tables 6 and 7. Summary statistics on advertising expenditures can be found in Table 8. The advertising data is in \$1000, and “PreSumAd” indicates the sum of advertising expenditures during the pre-release period, while “AdRel” indicates the advertising expenditures in the week of release. The advertising measure that we use in our estimation is the sum of all advertising expenditures in the eight weeks of pre-launch and the week of release (total of nine weeks).

Table 6: Weeks of Pre-Launch Advertising Expenditures

Number of Weeks of Pre-launch Advertising	Number of Movies
0-1	5
2-3	16
4-5	22
6-7	13
8+	7

Table 7: Advertising by Pre-Launch Week

Pre-launch Week	Number of Movies
-8	8
-7	13
-6	21
-5	32
-4	41
-3	50
-2	58
-1	62

Table 8: Advertising Expenditure Summary Statistics

	Min	Max	Mean	SD
PreSumAd	\$0	\$22,049,700	\$7,943,473	\$6,339,462
AdRel	\$11,200	\$9,394,900	\$4,503,687	\$2,789,035

We can see that most of the movies in our dataset have advertising expenditures for two to five weeks of the pre-launch period. Only seven movies engage in pre-launch advertising for eight or more weeks. Since we observe search activity for nearly all of the movies (61 out of 63) in our dataset for the entire pre-launch period of eight weeks, we can infer that search activity is capturing consumer interest that is not simply a response to advertising. Additionally, we have five movies with zero or one week of pre-launch advertising and one movie with no pre-launch advertising. For these movies, search activity may provide a valuable opportunity for gauging consumer interest during the pre-launch phase.

5.5 Estimation Results

We estimate our extended model using Maximum Likelihood Estimation. Since we are still using a latent class segmentation approach, we begin by determining the optimal number of segments. The last set of results in Table 4 reflects the advertising framework. We again find that a two-segment model is optimal, based on minimization of the BIC. The estimation results of the two-segment model can be found in Table 9.

Table 9: Search-Sales Model with Advertising: Estimation Results

Parameter		Segment 1		Segment 2	
ϕ	<i>Inflation Parameter</i>	0.838	(0.005)	0.773	(0.019)
λ_w	<i>Weibull Slope Parameter</i>	0.008	(0.003)	0.007	(0.002)
c	<i>Weibull Shape Parameter</i>	1.652	(0.098)	1.958	(0.105)
σ	<i>Normal Standard Deviation</i>	1.010	(0.117)	0.727	(0.153)
λ_e	<i>Exponential Parameter</i>	0.545	(0.015)	0.726	(0.018)
m_0		260.807	(6.987)	6.775	(2.218)
m_1		51.214	(1.233)	1.336	(0.100)
m_2		0.986	(0.114)	1.347	(0.114)
a_0		-5.097	(0.073)	-9.537	(0.334)
a_1		0.137	(0.009)	0.473	(0.019)
β_0		-169.990	(5.992)	-	-
β_{PG13}		-131.522	(8.719)	-	-
β_R		-203.593	(10.486)	-	-
β_{Comedy}		225.409	(10.163)	-	-
β_{Drama}		-95.126	(3.171)	-	-
β_{Comp}		88.716	(2.697)	-	-
Standard errors are in parentheses					
All estimates are significant at the .05 level					

There are several interesting results. First, we find that even after incorporating the role of advertising, the parameter estimates for search are still significant. This finding lends support for our premise that online search activity is measuring consumer interest that is driven by sources other than only advertising. Second, we find that the coefficient on baseline penetration (m_1) and the coefficient

on advertising in the sales component (m_2) are positive, suggesting a direct relationship between search and sales, as well as advertising and sales. This implies that advertising is indeed an important factor in the success of a motion picture. The large difference in magnitude for m_1 suggests a much stronger effect in segment 1. We also find that the coefficient on advertising in the penetration component (a_1) is positive and significant, implying a direct relationship between advertising expenditures and search volume.

We discuss our results in the next section.

5.6 Discussion

We have already proposed the use of online search as a measure of consumer interest. We further differentiate between pre-launch and post-launch search. We posit that pre-launch search represents product interest, while post-launch search represents consumption interest, along with product interest. We have not explicitly measured any of the drivers of this interest in a new product but have linked online search for a motion picture to opening-weekend box-office revenues. In this chapter, we link advertising expenditures to search, as well as opening-weekend box-office revenues. Therefore, we incorporate advertising in both the search and sales components of our proposed framework. Our results indicate that although advertising expenditures are important in modeling the performance of motion pictures, search is also significant. Thus, search appears to indicate consumer interest in a movie that results from sources other than advertising. These sources could be WOM, press releases, etc. We do not explicitly measure any other drivers of online search in our study. We do focus on advertising, since these expenditures are

especially important in the motion picture industry. Our results also suggest the usefulness in incorporating both search and advertising in a forecasting model.

Lastly, we seek to more closely examine the relationship between online search and advertising and their respective abilities to forecast opening weekend sales. Forecasting product sales during the pre-launch period for a new product is a problem that is often faced by managers. Our approach addresses this issue. We discuss our forecasting procedure using our proposed framework in the next section.

Chapter 6: Forecasting

6.1 Literature

Some recent work in the forecasting literature has focused on the pre-release or pre-launch period. Urban, Weinberg, and Hauser (1996) look at pre-launch forecasting for really-new products, specifically, electric vehicles. They use a qualitative framework that is based on a “virtual-buying environment” where consumers are engaged in simulated product-related experiences and able to search for product information. Moe and Fader (2002) look at advance purchase orders for CDs to forecast sales. Advance purchase orders refer to customer orders that are placed for an item prior to the item being available for purchase. The authors use the pattern of advance orders to forecast new album sales using data from an online retailer of music albums. Lee, Boatwright, and Kamakura (2003) also look at pre-launch sales forecasting of music. They use sales of previous albums and pre-launch information about the album to forecast weekly sales. Our research also follows a similar stream in that we look at the pre-launch or pre-release period for motion pictures. We discuss our forecasting procedure and results in the next sections.

6.2 Calibration and Validation

We test the forecasting performance of our modeling approach using calibration and validation samples. Our calibration sample consists of 48 movies, and our validation sample consists of the remaining 15 movies. After sorting the movie titles in alphabetical order, every fourth title is included in the validation sample. We

estimate both the Search-Sales Model and Search-Sales Model with Advertising on the calibration sample of 48 movies. The models are estimated on all 16 weeks of search data. The estimation results can be found in Tables 10 and 11.

Table 10: Search – Sales Model: Calibration Estimation Results

Parameter		Segment 1		Segment 2	
ϕ	<i>Inflation Parameter</i>	0.741	(0.015)	0.765	(0.008)
λ_w	<i>Weibull Slope Parameter</i>	0.011	(0.001)	0.006	(0.000)
c	<i>Weibull Shape Parameter</i>	1.951	(0.043)	2.032	(0.027)
μ	<i>Normal Mean</i>	14.944	(0.457)	15.466	(0.430)
σ	<i>Normal Standard Deviation</i>	2.328	(0.367)	1.911	(0.141)
λ_e	<i>Exponential Parameter</i>	0.714	(0.016)	0.774	(0.006)
α	<i>Penetration Rate</i>	0.024	(0.001)	0.128	(0.000)
β_0		-212.521	(4.288)	-	-
β_{PG13}		-2.202	(0.092)	-	-
β_R		-38.358	(1.376)	-	-
β_{Comedy}		166.542	(3.200)	-	-
β_{Drama}		-173.948	(1.000)	-	-
β_{Comp}		59.683	(1.149)	-	-
Standard errors are in parentheses					
All estimates are significant at the .05 level					

Table 11: Search-Sales Model with Advertising: Calibration**Estimation Results**

Parameter		Segment 1		Segment 2	
ϕ	<i>Inflation Parameter</i>	0.076	(0.001)	0.126	(0.001)
λ_w	<i>Weibull Slope Parameter</i>	0.006	(0.000)	0.002	(0.000)
c	<i>Weibull Shape Parameter</i>	2.369	(0.006)	2.860	(0.005)
σ	<i>Normal Standard Deviation</i>	2.794	(0.218)	0.897	(0.078)
λ_e	<i>Exponential Parameter</i>	0.006	(0.000)	0.009	(0.000)
m_0		88.264	(1.944)	4.869	(0.934)
m_1		4.427	(0.707)	0.408	(0.077)
m_2		1.182	(0.642)	1.195	(0.080)
a_0		-20.702	(0.071)	-20.713	(0.103)
a_1		1.323	(0.005)	1.325	(0.007)
β_0		-45.213	(0.500)	-	-
β_{PG13}		-43.781	(0.464)	-	-
β_R		-76.643	(0.857)	-	-
β_{Comedy}		73.217	(0.703)	-	-
β_{Drama}		-67.567	(0.677)	-	-
β_{Comp}		28.397	(0.255)	-	-
Standard errors are in parentheses					
All estimates are significant at the .05 level					

We use these estimation results to determine segment membership probabilities for each of the movies in the validation sample. Segment membership for the validation sample is determined using only movie characteristics (i.e., no search data). For the Search-Sales model, the segment membership probabilities and the estimated values for μ for each segment give us predicted sales. For the Search-Sales Model with Advertising model, segment membership probabilities are determined similarly. The predicted sales (μ) are calculated using the calibration sample estimates for m_0 , m_1 , and m_2 as specified in the model. It is important to note that we are forecasting the natural log of sales. The results of the forecasting are discussed in the next section.

6.3 Forecasting Results

As discussed earlier, we forecast sales for a validation sample of 15 movies. The results of the forecasting are found in Table 12. The forecasting performance of our Search-Sales model is quite good. The MAPE (mean absolute percentage error) is 9.46%. We can see from the results that there is a fairly large range in the APEs (absolute percentage error) for the movies in the validation sample. The movies with the highest APEs are *Babel*, *Little Miss Sunshine*, and *Quinceanera*. These movies gained popularity in later weeks of release, and therefore did not follow the typical pattern of sales. Overall, this model performs well. The results offer support for our premise that online search offers a useful measure of consumer interest in a new product, and therefore offers predictive power in forecasting sales for a new product.

The forecasting performance of our Search-Sales Model with Advertising is better than the Search-Sales model. We can see from the results that the range for the APEs is narrower, and the MAPE (3.20%) is lower than the Search-Sales model as well. This is not surprising, as advertising expenditures in the movie industry are substantial, and as our results appear to support, can play a large role in the success of a motion picture. However, it is important to highlight that the search parameters are still significant, even after the inclusion of advertising in our model. We discuss our results and their implications in more detail in the next section.

Table 12: Forecasting Results

	SearchSales	Advertising
Title	APE	APE
Babel	19.55	5.82
Clerks II	6.52	1.98
Déjà Vu	8.61	0.41
Flicka	2.95	0.45
Happy Feet	12.27	4.20
Jackass: Number Two	12.28	9.13
Little Miss Sunshine	17.53	0.29
One Night with the King	1.06	11.31
Quinceanera	34.22	2.28
Scoop	0.96	0.50
The Ant Bully	5.49	3.49
The Descent	3.83	4.11
The Illusionist	12.00	0.75
The Protector	0.28	1.70
The Wicker Man	4.29	1.54
MAPE	9.46	3.20

6.4 Discussion

The primary focus of this dissertation is to investigate the predictive power of online search data in forecasting new product sales, movies, in our specific case. We can see from our earlier results, that search patterns for movies are fairly predictable. The forecasting results of our model indicate that search data does indeed offer predictive power in forecasting sales. The MAPE is under 10%, suggesting that search data may offer a useful measure to managers interested in forecasting sales of a new product, particularly in the pre-launch period. We emphasize that in our validation sample, the only data that is used to forecast is movie characteristics (including competition), which is available well in advance of a motion picture's release. Thus, this framework can be used to forecast before any advertising data is available, the only necessary data is on product characteristics. This again highlights

the usefulness of our modeling framework and forecasting procedure for pre-launch forecasting of new product sales.

In our specific context, the motion picture industry, advertising is known to be an important factor in the success of a new movie. Our results also illustrate the importance of incorporating advertising expenditures into a forecasting model. The forecasts are greatly improved with the inclusion of advertising data, suggesting that the combination of search data and advertising data may offer a powerful tool to managers in industries where advertising expenditures are substantial. Managers may be able to predict, with a fair amount of accuracy, new product sales during the pre-launch period. It is important to note that this framework can only be applied once advertising data for a motion picture is available. As discussed earlier, the forecasting of new product sales early in their life-cycle is a problem that managers have long struggled with. Our framework offers one possible approach to address this issue.

The overall results and managerial implications of our study, as well as concluding remarks and areas for future research, are discussed in the next section.

Chapter 7: Conclusion

7.1 Overview

Our main objective is to illustrate the effectiveness of online search volume as a new product sales forecasting measure. We are particularly interested in the *pre-launch* aspect of forecasting. We illustrate this in the context of motion picture revenues. We distinguish between online search that takes place before launch and search that takes place after launch (post-launch). We use pre-launch search volume as a measure of product interest and post-launch search as a measure of both consumption interest and product interest. Post-release search can be driven by product interest, as well as interest driven by other characteristics such as product availability.

We develop a modeling framework that links pre-launch search, post-launch search, and box-office revenues. We also incorporate product characteristics, including competition. We extend our framework and model the effect of advertising. Doing so allows us to account for at least one driver of online search and compare the forecasting performances of our two modeling approaches – with and without advertising data. We find that online search is a significant predictor of opening-weekend box-office revenues. We also find that our framework performs well as a forecasting tool. Further, our results indicate improved forecasting with the inclusion of advertising. Although advertising is also a significant predictor of sales, we find that search is also significant. Thus, the combination of search activity and

advertising expenditures can provide managers a valuable tool in managing consumer interest in new products.

7.2 Contribution

In this dissertation, we aim to address two questions. (1) Is online search term volume a good measure of consumer interest? and (2) Does this measure offer sales forecasting power? Our first contribution lies in the proposal of a new measure of consumer interest that is available prior to a new product's launch. WOM is one indication of consumer interest, and measurement of WOM is an issue that has been raised in the literature on WOM. Therefore, we introduce an easily-obtained, cost-effective measure for consumer interest in a new product. Data on search activity is easy to collect and clean. It does not require much coding or analysis of content. Search engine activity is also a very prevalent online activity as compared to blogging or writing consumer product reviews. The data is available early in the product's life cycle, and can capture consumer interest in a new product early in the consumer decision process. In other words, search data is available before a new product is launched, often several months prior to launch. Therefore, this measure can provide useful measures of consumer interest in a new product well in advance of the product's launch. This gives managers the opportunity to adjust their marketing strategy as necessary, depending on the levels of consumer interest in the new product before the product is even available for consumption.

Our second contribution is in the development of a model that uses this measure to forecast new product sales. We first develop a model that forecasts sales using product characteristics and search term data. We then extend our modeling

framework to include one possible driver of search activity - advertising. We focus on advertising because we are illustrating our framework in the context of motion pictures, where advertising expenditures can play a large role in the success of a movie. We find that search data offers significant predictive power in forecasting opening-weekend box-office revenues, and our modeling framework and forecasting procedure perform quite well. We further find that search data combined with advertising data improves the forecasting ability of our framework. It is important to note that the significance of the search data is not eliminated with the inclusion of advertising. Thus, we can infer that search activity is capturing consumer interest in a product stemming from other drivers as well. These results have useful implications for managers of new products. Managers can monitor the search pattern and volume for terms related to their products during the pre-launch period to gauge interest in their new products. They can use this information to alter their marketing strategy, if necessary.

Our study offers several opportunities for further research in this area. We discuss limitations of our research and areas for future work in the next section.

7.3 Limitations and Future Research

One problem that is faced when using search data is the lack of very clean data. For example, some movie titles that are also related to other products (e.g. water, firewall, cars) may be searched for using search terms that involve more than just the title. Thus, the data for these terms may be contaminated. Also, sometimes a motion picture title is also a book title. If a movie title is very long, perhaps online searchers only enter the first few words or the main words as search terms.

Consumers may also search using the names of actors or actresses with roles in the motion picture. Thus, in some cases, it may be difficult to tell exactly what a consumer would enter as a search term when searching for information on a motion picture. Some experimental work in this area would help to better understand consumer choice of a search term.

Although this work represents an example of the usefulness of online search volume as a predictor of motion picture success, there exist many opportunities for future research. A similar problem could be examined for DVD release dates. Also, with the historical data on search volume, other times of high motion picture interest could be determined. For example, DVDs are often given as gifts during the holiday season. Thus, there is likely to be a surge in searches for particular titles during this time. As mentioned earlier, the most recent work in this area has focused on online reviews and ratings posted by consumers. An obvious extension would be to combine reviews/ratings and search volume into a consumer interest measure in a forecasting model.

Also, the timing problem could be examined. There has been research done on the optimal timing of a motion picture release on DVD based on the success of the motion picture in theaters (Lehmann and Weinberg, 2000). Perhaps WOM data captured by search volume could help optimize this solution as well. Elberse and Eliashberg (2003) look at the sequential release of motion pictures in international markets. They find that “the longer is the time lag between releases, the weaker is the relationship between domestic and foreign performance.” Though they focus on screen allocations, the authors suggest that this is “consistent with the idea that the

‘buzz’ for a movie is perishable.” If this is the case with international markets, it is likely to be the case for the box-office – DVD performance relationship. Again, perhaps the strength of this “buzz” could be captured by search volume to help optimize this solution as also.

Our approach could also be extended to product categories beyond motion pictures. Search data on terms related to other products are just as easily available. Thus, our framework could be applied to products such as books, music, or technology products. Our framework is flexible in the sense that the probability distributions that are used can be adapted to fit the data better. Other distributions may be used for products with longer life-cycles or different patterns of search, in general.

Lastly, there are some demographic biases when looking only at Internet data. For example, it has been reported that males tend to use the Internet more than females. Also, Internet use tends to increase with household income and education level. People of Hispanic ethnicity are least likely to use the Internet (PEW/Internet, 2006). Additionally, younger Internet users are more likely to use search engines and use them often (PEW/Internet, 2005). We do not address these issues in our study.

While there are many limitations and opportunities for further study in our area of research, we take the first step of measuring consumer interest in a new product using online search term data and use this data to successfully predict new product sales.

Appendix

Search-Sales Model: Estimation Results for Various Market Sizes

Parameter		Segment 1		Segment 2	
ϕ	<i>Inflation Parameter</i>	0.7969	(0.0027)	0.7525	(0.0017)
λ_w	<i>Weibull Slope Parameter</i>	0.0090	(0.0001)	0.0046	(0.0001)
c	<i>Weibull Shape Parameter</i>	1.8301	(0.0094)	2.1313	(0.0086)
μ	<i>Normal Mean</i>	15.0043	(0.2382)	16.1473	(0.1935)
σ	<i>Normal Standard Deviation</i>	2.0248	(0.0891)	1.4711	(0.0332)
λ_e	<i>Exponential Parameter</i>	0.6543	(0.0060)	0.7733	(0.0084)
α	<i>Penetration Rate</i>	0.0683	(0.0004)	0.2091	(0.0006)
β_0		-137.4962	(1.2277)	-	-
β_{PG13}		8.6298	(0.0719)	-	-
β_R		-235.4280	(1.3861)	-	-
β_{Comedy}		276.6742	(1.8006)	-	-
β_{Drama}		318.2247	(1.8862)	-	-
β_{Comp}		72.4208	(0.6206)	-	-
Standard errors are in parentheses					
All estimates are significant at the .05 level					
M=95,000					

Parameter		Segment 1		Segment 2	
ϕ	<i>Inflation Parameter</i>	0.8312	(0.0108)	0.7678	(0.0034)
λ_w	<i>Weibull Slope Parameter</i>	0.0062	(0.0004)	0.0064	(0.0001)
c	<i>Weibull Shape Parameter</i>	1.7543	(0.0372)	1.9627	(0.0352)
μ	<i>Normal Mean</i>	15.0784	(0.7276)	15.3873	(0.1019)
σ	<i>Normal Standard Deviation</i>	2.1276	(0.2084)	1.8885	(0.1723)
λ_e	<i>Exponential Parameter</i>	0.4855	(0.0144)	0.7058	(0.0047)
α	<i>Penetration Rate</i>	0.0372	(0.0005)	0.1008	(0.0029)
β_0		-228.4689	(0.9218)	-	-
β_{PG13}		-168.2976	(0.7054)	-	-
β_R		-224.0473	(0.6523)	-	-
β_{Comedy}		298.4611	(0.9048)	-	-
β_{Drama}		-339.2037	(1.0737)	-	-
β_{Comp}		109.1394	(0.3108)	-	-
Standard errors are in parentheses					
All estimates are significant at the .05 level					
M=125,000					

References

- Arndt, Johan (1967), "Role of Product-Related Conversations in the Diffusion of a New Product," *Journal of Marketing Research*, 4 (August), 291–295.
- Banerjee, Abhijit and Drew Fudenberg (2004), "Word-of-mouth Learning," *Games and Economic Behavior*, vol. 46, 1–22.
- Bansal, Harvir S. and Peter A. Voyer (2000), "Word-of-Mouth Processes Within a Services Purchase Decision Context," *Journal of Service Research*, 3(2), 166–177.
- Basuroy, Suman, Subimal Chatterjee, and S. Abraham Ravid (2003), "How Critical are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets," *Journal of Marketing*, 67 (October), 103–117.
- 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), 287–295.
- Bayus, Barry L. (1985), "Word of Mouth: The Indirect Effects of Marketing Efforts," *Journal of Advertising Research*, 25(3), 31–39.
- Berger, Jonah and Katy Milkman (2009), "Social Transmission and Viral Culture," The Wharton School, University of Pennsylvania working paper.
- Blackwell, Roger D, Paul W. Miniard, and James F. Engel (2001), *Consumer Behavior*, 9th edition, Harcourt, Orlando, FL.
- Bone, Paula Fitzgerald (1995), "Word-of-Mouth Effects on Short-term and Long-term Product Judgments," *Journal of Business Research*, vol. 32, 213–223.
- Bradlow, Eric T. and David C. Schmittlein (2000), "The Little Engine That Could: Modeling the Performance of World Wide Web Search Engines," *Marketing Science*, 19(1), 43–62.
- Brown, Jacqueline Johnson and Peter H. Reingen (1987), "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research*, 14(3), 350–362.
- Brown, Tom J., Thomas E. Barry, Peter A. Dacin, and Richard F. Gunst (2005), "Spreading the Word: Investigating Antecedents of Consumers' Positive Word-of-Mouth Intentions and Behaviors in a Retailing Context," *Journal of the Academy of Marketing Science*, 33(2), 123–138.

- Bruce, Norris I. and Natasha Zhang Foutz (2007), "Dynamic Effectiveness of Advertising and Word-of-mouth in the Sequential Distribution of Short Lifecycle Products," working paper.
- Brynjolfsson, Erik and Michael D. Smith (2000), "Frictionless Commerce: A Comparison of Internet and Conventional Retailers," *Management Science*, 46(4), 563–585.
- Bucklin, Randolph E. and Catarina Sismeiro (2003), "A Model of Web Site Browsing Behavior Estimated on Clickstream Data," *Journal of Marketing Research*, 40(3), 249–267.
- Burzynski, Michael H. and Dewey J. Bayer (1977), "The Effect of Positive and Negative Prior Information on Motion Picture Appreciation," *Journal of Social Psychology*, vol. 101, 215–218.
- Calder, Bobby J. and Brian Sternthal (1980), "Television Commercial Wearout: An Information Processing View," *Journal of Marketing Research*, 17 (May), 173–186.
- Chatterjee, Patrali, Donna L. Hoffman, and Thomas P. Novak (2003), "Modeling the Clickstream: Implications for Web-Based Advertising Efforts," *Marketing Science*, 22(4), 520–541.
- Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43(3), 345–354.
- Deighton, John and Leora Kornfeld (2008), "Digital Interactivity: Marketing Without Power," Harvard Business School working paper.
- Delaney, Kevin J. (2007), "The New Benefits of Web-Search Queries," *The Wall Street Journal*, February 6, 2007, B3.
- Dellarocas, Chrysanthos (2003), "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science*, 49(10), 1407–1424.
- Dellarocas, Chrysanthos, Xiaoquan (Michael) Zhang, and Neveen Farag Awad (2007), "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21(4), 23–45.
- Dodson, Joe A. and Eitan Muller (1978), "Models of New Product Diffusion Through Advertising and Word-of-Mouth," *Management Science*, 24(15), 1568–1578.
- Dreze, Xavier and Francois-Xavier Hussherr (2003), "Internet Advertising: Is Anybody Watching?," *Journal of Interactive Marketing*, 17(4), 8–23.

- Duhan, Dale F., Scott D. Johnson, James B. Wilcox, and Gilberg D. Harrell (1997), "Influences on Consumer Use of Word-of-Mouth Recommendation Sources," *Journal of the Academy of Marketing Science*, 25(4), 283–295.
- Edell, Julie A. and Marian Chapman Burke (1987), "The Power of Feelings in Understanding Advertising Effects," *Journal of Consumer Research*, 14(3), 421–433.
- Elberse, Anita and Bharat Anand (2005), "Advertising and Expectations: The Effectiveness of Pre-Release Advertising for Motion Pictures," Harvard Business School working paper.
- 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(3), 329–354.
- 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), 638–661.
- Eliashberg, Jehoshua, Jedid-Jah Jonker, Mohanbir S. Sawhney, and Berend Wierenga (2000), "MOVIEMOD: An Implementable Decision-Support System for Prerelease Market Evaluation of Motion Pictures," *Marketing Science*, 19(3), 226–243.
- Eliashberg, Jehoshua, and Steven M. Shugan (1997), "Film Critics: Influencers or Predictors?," *Journal of Marketing*, 61(2), 68–78.
- Ettredge, Michael, John Gerdes, and Gilbert Karuga (2005), "Using Web-based Search Data to Predict Macroeconomic Statistics," *Communications of the ACM*, 48(11), 87–92.
- Fallows, Deborah (2005), "Search Engine Users," Pew Internet and American Life Project, 1–29.
- Foutz, Natasha and Wolfgang Jank (2007), "The Wisdom of Crowds: Pre-release Forecasting for New Products via Functional Data Analysis of Online Virtual Stock Market," University of Maryland working paper.
- Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23(4), 545–560.
- Goldenberg, Jacob, Barak Libai, and Eitan Muller (2001), "Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth," *Marketing Letters*, 12(3), 211–233.
- Greenwald, Anthony G. and Clark Leavitt (1984), "Audience Involvement in Advertising: Four Levels," *Journal of Consumer Research*, 11(1), 581–592.

- Gupta, Sachin and Pradeep K. Chintagunta (1994), "On Using Demographic Variables to Determine Segment Membership in Logit Mixture Models," *Journal of Marketing Research*, 31(1), 128–136.
- Harrison-Walker, L. Jean (2001), "The Measurement of Word-of-Mouth Communication and an Investigation of Service Quality and Customer Commitment as Potential Antecedents," *Journal of Service Research*, 4(1), 60–75.
- Hauser, John R., Glen L. Urban, and Bruce D. Weinberg (1993), "How Consumers Allocate Their Time When Searching for Information," *Journal of Marketing Research*, vol. 30 (November), 452–466.
- Haywood, K. Michael (1989), "Managing Word of Mouth Communications," *Journal of Services Marketing*, 3(2), 55–67.
- Hennig-Thurau, Thorsten, Kevin P. Gwinner, Gianfranco Walsh, and Dwayne D. Gremler (2004), "Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?," *Journal of Interactive Marketing*, 18(1), 38–52.
- Hennig-Thurau, Thorsten and Gianfranco Walsh (2003), "Electronic Word-of-Mouth: Motives for and Consequences of Reading Customer Articulations on the Internet," *International Journal of Electronic Commerce*, 8(2), 51–74.
- Herr, Paul M., Frank R. Kardes, and John Kim (1991), "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective," *Journal of Consumer Research*, 17(4), 454–462.
- Hoch, Stephen J. and Young-Won Ha (1986), "Consumer Learning: Advertising and the Ambiguity of Product Experience," *Journal of Consumer Research*, 13(2), 221–233.
- Holbrook, Morris B. and Rajeev Batra (1987), "Assessing the Role of Emotions as Mediators of Consumer Responses to Advertising," *Journal of Consumer Research*, 14(3), 404–420.
- Holmes, John H. and John D. Lett, Jr. (1977), "Product Sampling and Word of Mouth," *Journal of Advertising Research*, 17(5), 35–40.
- Iyer, Ganesh, David Soberman, J. Miguel Villas-Boas (2005), "The Targeting of Advertising," *Marketing Science*, 24(3), 461–476.
- Jansen, Bernard J. and Paulo R. Molina (2006), "The Effectiveness of Web Search Engines for Retrieving Relevant Ecommerce Links," *Information Processing and Management*, 42(4), 1075–1098.

- Johnson, Eric J., Wendy W. Moe, Peter S. Fader, Steven Bellman, and Gerald L. Lohse (2004), "On the Depth and Dynamics of Online Search Behavior," *Management Science*, 50(3), 299–308.
- Kamakura, Wagner A. and Gary J. Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26(4), 379–390.
- Kaul, Anil and Dick R. Wittink (1995), "Empirical Generalizations About the Impact of Advertising on Price Sensitivity and Price," *Marketing Science*, 14(3), G151–G160.
- Kiel, Geoffrey C. and Roger A. Layton (1981), "Dimensions of Consumer Information Seeking Behavior," *Journal of Marketing Research*, vol. 18 (May), 233–239.
- Klein, Lisa R. and Gary T. Ford (2003), "Consumer Search for Information in the Digital Age: An Empirical Study of Prepurchase Search for Automobiles," *Journal of Interactive Marketing*, 17(3), 29–49.
- Krider, Robert E. (2006), "Research Opportunities at the Movies," *Marketing Science*, 25(6), 662–664.
- Krider, Robert E. and Charles B. Weinberg (1998), "Competitive Dynamics and the Introduction of New Products: The Motion Picture Timing Game," *Journal of Marketing Research*, 35(1), 1–15.
- Krishnamurthi, Lakshman and S. P. Raj (1985), "The Effect of Advertising on Consumer Price Sensitivity," *Journal of Marketing Research*, 22(2), 119–129.
- Kumar, Nanda and Karl R. Lang (2007), "Do Search Terms Matter for Online Consumers? The Interplay Between Search Engine Query Specification and Topical Organization," *Decision Support Systems*, vol. 44, 159–174.
- Lavidge, Robert J. and Gary A. Steiner (1961), "A Model for Predictive Measurements of Advertising Effectiveness," *Journal of Marketing*, 25(6), 59–62.
- Lee, Jonathan, Peter Boatwright, and Wagner A. Kamakura (2003), "A Bayesian Model for Prelaunch Sales Forecasting of Recorded Music," *Management Science*, 49(2), 179–196.
- Lehmann, Donald R. and Charles B. Weinberg (2000), "Sales Through Sequential Distribution Channels: An Application to Movies and Videos," *Journal of Marketing*, 64 (July), 18–33.
- Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (July), 74–89.

- Luan, Y. Jackie and K. Sudhir (2007), "Forecasting Advertising Responsiveness for Short Lifecycle Products," working paper.
- MacKenzie, Scott B. and Richard J. Lutz (1989), "An Empirical Examination of the Structural Antecedents of Attitude Toward the Ad in an Advertising Pretesting Context," *Journal of Marketing*, vol. 53 (April), 48–65.
- MacKenzie, Scott B., Richard J. Lutz, and George E. Belch (1986), "The Role of Attitude Toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations," *Journal of Marketing Research*, vol. 23 (May), 130–143.
- Madden, Mary (2006), "Internet Penetration and Impact," Pew Internet and American Life Project, 1–5.
- Mahajan, Vijay and Eitan Muller (1986), "Advertising Pulsing Policies for Generating Awareness for New Products," *Marketing Science*, 5(2), 89–106.
- Mangold, W. Glynn, Fred Miller, and Gary R. Brockway (1999), "Word-of-mouth communication in the service marketplace," *Journal of Services Marketing*, 13(1), 73–89.
- Mela, Carl F., Sunil Gupta, and Donald R. Lehmann (1997), "The Long-Term Impact of Promotion and Advertising on Consumer Brand Choice," *Journal of Marketing Research*, 34(2), 248–261.
- Meyer, Robert J. (1982), "A Descriptive Model of Consumer Information Search Behavior," *Marketing Science*, 1(1), 93–121.
- Milgrom, Paul and John Roberts (1986), "Price and Advertising Signals of Product Quality," *Journal of Political Economy*, 94(4), 796–821.
- Mitchell, Andrew A. and Jerry C. Olson (1981), "Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude?," *Journal of Marketing Research*, 18(3), 318–332.
- Moe, Wendy W. (2003), "Buying, Searching, or Browsing: Differentiating Between Online Shoppers Using In-Store Navigational Clickstream," *Journal of Consumer Psychology*, 13(1,2), 29–39.
- . (2006), "An Empirical Two-Stage Choice Model with Varying Decision Rules Applied to Internet Clickstream Data," *Journal of Marketing Research*, 43(4), 680–692.
- Moe, Wendy W. and Peter S. Fader (2002), "Using Advance Purchase Orders to Forecast New Product Sales," *Marketing Science*, 21(3), 347–364.

- Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, vol. 23 (March), 263–277.
- Moul, Charles C. (2007), "Measuring Word of Mouth's Impact on Theatrical Movie Admissions," *Journal of Economics and Management Strategy*, 16(4), 859–892.
- MPAA (2006), *U.S. Theatrical Market Statistics*, 1–24.
- 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), 214–345.
- Neelamegham, Ramya and Pradeep Chintagunta (1999), "A Bayesian Model to Forecast New Product Performance in Domestic and International Markets," *Marketing Science*, 18(2), 115–136.
- Petty, Richard E., John T. Cacioppo, and David Schumann (1983), "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement," *Journal of Consumer Research*, vol. 10 (January), 135–146.
- PEW/Internet American Life Project (2005).
- PEW/Internet American Life Project (2006).
- Phelps, Joseph E., Regina Lewis, Lynne Mobilio, David Perry, and Niranjan Raman (2004), "Viral Marketing or Electronic Word-of-Mouth Advertising: Examining Consumer Responses and Motivations to Pass Along Email," *Journal of Advertising Research*, 44(4), 333–348.
- Ratchford, Brian T. (1982), "Cost-Benefit Models for Explaining Consumer Choice and Information Seeking Behavior," *Management Science*, 28(2), 197–212.
- Ratchford, Brian T., Myung-Soo Lee, and Debabrata Talukdar (2003), "The Impact of the Internet on Information Search for Automobiles," *Journal of Marketing Research*, 40(2), 193–209.
- Richins, Marsha L. (1983), "Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study," *Journal of Marketing*, 47(Winter), 68–78.
- Sasieni, Maurice W. (1989), "Optimal Advertising Strategies," *Marketing Science*, 8(4), 358–370.
- Sawhney, Mohanbir S. and Jehoshua Eliashberg (1996), "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures," *Marketing Science*, 15(2), 113–131.

- Schwarz, Gideon (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6(2), 461–464.
- Smith, Michael D. and Erik Brynjolfsson (2001), "Consumer Decision-Making at an Internet Shopbot: Brand Still Matters," *Journal of Industrial Economics*, 49(4), 541–558.
- Spiteri, Louise E. (2000), "Access to Electronic Commerce Sites on the World Wide Web: An Analysis of the Effectiveness of Six Internet Search Engines," *Journal of Information Science*, 26(3), 173–183.
- Steenkamp, Jan-Benedict E. M., Vincent R. Nijs, Dominique M. Hanssens, and Marnik G. Dekimpe (2005), "Competitive Reactions to Advertising and Promotion Attacks," *Marketing Science*, 24(1), 35–54.
- Sundaram, D.S., Kaushik Mitra, and Cynthia Webster (1998), "Word-of-Mouth Communications: A Motivational Analysis," *Advances in Consumer Research*, vol. 25, 527–531.
- Telang, Rahul, Peter Boatwright, and Tridas Mukhopadhyay (2004), "A Mixture Model for Internet Search-Engine Visits," *Journal of Marketing Research*, 41(2), 206–214.
- "Traditional Media Fuels Online Searches," *Chain Store Age*, April 2007, 16.
- Urban, Glen L., Bruce D. Weinberg, and John R. Hauser (1996), "Premarket Forecasting of Really-New Products," *Journal of Marketing*, 60 (January), 47–60.
- Urbany, Joel E., Peter R. Dickson, and William L. Wilkie (1989), "Buyer Uncertainty and Information Search," *Journal of Consumer Research*, 16(2), 208–215.
- Vakratsas, Demetrios and Tim Ambler (1999), "How Advertising Works: What Do We Really Know?," *Journal of Marketing*, 63 (January), 26–43.
- Wierenga, Berend (2006), Motion Pictures: Consumers, Channels, and Intuition," *Marketing Science*, 25(6), 674–677.
- Word of Mouth Marketing Association, www.womma.org.
- Wu, Jianan and Arvind Rangaswamy (2003), "A Fuzzy Set Model of Search and Consideration with an Application to an Online Market," *Marketing Science*, 22(3), 411–434.
- Zufryden, Fred S. (1996), "Linking Advertising to Box Office Performance of New Film Releases: A Marketing Planning Model," *Journal of Advertising Research*, 36(4), 29–41.