A temporal study of the effects of online opinions : information sources matter

Huang, Jianxiong; Boh, Wai Fong; Goh, Kim Huat

2018

Huang, J., Boh, W. F., & Goh, K. H. (2017). A temporal study of the effects of online opinions : information sources matter. Journal of Management Information Systems, 34(4), 1169-1202. doi:10.1080/07421222.2017.1394079

https://hdl.handle.net/10356/142489

https://doi.org/10.1080/07421222.2017.1394079

A TEMPORAL STUDY OF THE EFFECTS OF ONLINE OPINIONS: INFORMATION SOURCES MATTER

Jianxiong Huang
Assistant Professor
Business School
Nanjing University
jxhuang@nju.edu.cn

Wai Fong Boh
Professor
Nanyang Business School
Nanyang Technological University
awfboh@ntu.edu.sg

Goh Kim Huat
Associate Professor
Nanyang Business School
Nanyang Technological University
akhgoh@ntu.edu.sg

ABSTRACT

This study examines when and why online comments from different sources and platforms influence a movie's box office receipts over time. Premised on the theory of information search, we hypothesize that consumers are more likely to engage in active search in the early stages of a movie's release due to greater choice uncertainty, and passive attention is more likely to kick in for later stages of a movie's release, as uncertainty decreases. To test the proposed hypotheses, we tracked over 1,500 sources of online expert and consumer reviews for cinematic movies released for an entire year and continuously monitored major social media sites (e.g. Twitter) for comments. We text-mined the comments to elucidate the sentiments and analyzed the data. Confirming our hypotheses, the results showed that expert reviews and pull-based peer comments have a significant influence in early stages of a movie's release, and the effects decrease over time. In contrast, the volume of comments from push-based microblog platforms have a significant influence on later box office receipts. Our research demonstrates that online opinions are not always persuasive and useful, and our findings provide insights into when consumers are likely to pay attention to which types of online opinions.

Keywords: Cinematic movies, Online comments, Information Processing

A TEMPORAL STUDY OF THE EFFECTS OF ONLINE OPINIONS:

INFORMATION SOURCES MATTER

INTRODUCTION

In recent years, there is burgeoning interest in online word-of-mouth (WOM), which has become an important source of information that affects consumers' decision making about purchases of products and services. One key question that prior research has sought to understand is the impact that online WOM have on the actual sales of products and services [62, 63, 65, 94]. Many studies, in particular, have examined the impact of online WOM in the context of cinematic movies [e.g., 43, 75, 78]. This is likely because empirically, movies provide the ideal context to examine the effects of online opinions on sales. First, word-of-mouth effects are particularly important for entertainment goods such as movies, which tend to receive significant public interest and attention [7, 41]. Second, various information aggregators constantly monitor the daily box office receipts, making it possible to obtain daily sales data for movies. Third, it is feasible to design a sampling schema that is representative of the industry, yet practical for data collection efforts as the number of movies released per year is a manageable number to track. Moreover, it is feasible for researchers to comprehensively retrieve online comments about movies, as the dispersion of comments on each movie tends to be concentrated in the period around its release.

Despite the valuable insights gained from prior research endeavors [e.g., 18, 43], we are not well equipped to answer the question of how online comments from different sources (e.g. written by experts vs peers), or via different types of platform (e.g. pushed based vs pull-based platforms) affect box office sales over time. Hence, unlike prior research that tends to focus on one or limited online platforms, we go beyond a single type of online platform to examine the research question of how product comments from different sources and platforms of online media affect consumer purchase.

This research question deserves our attention because it not only answers fundamental research questions of interest to IS researchers – the differential impacts of different technology types, but also provides greater insights into how online comments are factored into consumer decision making. Answering this research question also provides greater practitioner insights into how consumers may react to information retrieved from different sources and via different platforms. This allows organizations to more effectively target the use of the appropriate marketing channel, or leverage the appropriate information source to achieve their marketing objectives. This is especially relevant since online opinions can originate from and be retrieved from multiple information sources and online platforms [8, 30], and it is common for consumers to consult multiple information sources to inform their purchase decisions [45, 55]. Practitioners thus need to design a well-integrated set of marketing communications [40], which requires careful selection of appropriate channels for their marketing messages, taking into account the role that each channel plays in influencing consumers. Hence, our research will help practitioners to formulate a more effective marketing communication strategy using a portfolio of online communication platforms.

With the large amount of information available via various online sources, and given that consumers' state of knowledge is dynamic, and will change as time passes [57], it is important to disentangle the temporal influence of online comments across different sources and platforms [e.g., 40]. Such insights will allow practitioners to strategically adjust the use of different marketing channels across time, leveraging the most effective channels at the right time.

In addition to the practical relevance, our research also contributes to the literature in three ways. First, previous inquiries tend to focus on one particular online media, or focus on examining how the product reviews from one site (e.g. Amazon.com and Barnesandnoble.com) affect product sales in the same site [e.g., 2, 22, 68]. As highlighted by Gu et al. [45], however, consumers often conduct searches through various external websites that are different from where they will make their purchases. Furthermore, consumer comments on external websites can have a more significant effect than those on internal retailer-hosted websites [45]. This highlights that consumers often search broadly for information on the Internet, and given the multiple channels of information propagation available for any product, it is insufficient to examine the reviews from a single source. Prior research, which focuses on comments from only one source has not accurately captured the breadth of online opinions from various sources and sites available to consumers. In addition, prior work has not considered how information arising from distinct information sources and online platforms would influence product sales differently. By comparing how different sources of information influence product sales, we seek insights into the ways in which online media characteristics influence individuals' decision making. To provide adequate comparison, our empirical data collection is more comprehensive than published studies to date, as we collected different types of online comments, and we used multiple sites to collect each source of information.

Second, prior research recognizes that online comments about movies are not static, but follows a dynamic phenomenon that spans the entire run of the movie [69]. New comments about movies are continuously posted on various platforms. Recognizing this, prior research that has examined the impact of online comments on movie sales has often broken down the analysis by weeks, with the recognition that the impact is not consistent across time [32, 35, 69]. The explanations provided about the differential influence of online comments across time, however, are usually exploratory in nature and covered only in the discussion sections of prior research. Furthermore, different studies have found inconsistent findings relating to whether online comments impact movie sales earlier or later in the product lifecycle, making it difficult to synthesize the findings and arrive at consistent conclusions about how the dynamic nature of online comments affect consumers' movie-going decisions (see below for more details).

This study looks at social media comments from a comprehensive perspective in terms of both breadth (i.e. comments from different platform sources) and depth (i.e. a longitudinal panel of comments). This holistic approach allows us to consider the issues of platform source differences as well as the dynamic nature of comments, to propose a coherent theoretical framework that can help to bridge the inconsistent findings in the extant literature. This approach is adopted as it allows us to empirically examine both different sources of online comments and across time by validating the framework with data collected about the same focal set of movies.

Third, premised upon how uncertainty drives consumers' information seeking and processing behaviors differently over time, we present a set of coherent theoretical arguments to explain whether, when and why consumers pay attention to online comments from different sources and platforms. Past research has suggested that uncertainty surround the early weeks of a movie release [18], and moviegoers find it difficult to evaluate the quality of movies and choose a movie consistent with their taste [74]. As time goes by, information such as box office rankings becomes available [18, 33], reducing uncertainty about the quality and popularity of a movie. We theorize that different levels of uncertainty have an impact on consumers' information seeking and processing behaviors, presenting empirical evidence that such adaptive response is characterized by changes in the preference of the type of online information over time.

LITERATURE REVIEW

Information Sources and Platforms

The information processing literature has shown that independent of message content, messages from different sources influence a person's attitude differently [e.g., 16, 38, 79]. Hence, one should expect that online comments from different sources would differ in their impact on consumers' attitude. In particular, we examine the difference between information arising from experts versus peers. Similarly, we expect comments from forums and discussion boards (which are mainly searched and browsed), versus comments from microblogs (which are mainly pushed to consumers) should also influence a person's attitude differently given differences in how these messages are propagated. In this section, we first discuss the differentiation between comments from experts versus peers. Then we discuss the differentiation between peer comments pulled from forums versus those pushed via microblogs.

Comments from Experts vs Peers. Source characteristics affect the way people process messages [38, 41]. In particular, expertise is a key aspect of a message source that influences the persuasiveness of a message [91]. It is, however, not always the case that a message from an expert is more persuasive than that from a non-expert. Prior research found that unsolicited expert recommendations can be counter-productive, resulting in consumers deliberately ignoring or even contradicting them [37]. In our study, we distinguish between online expert reviews versus peer reviews. Expert reviews are written by critics, defined as persons employed to "screen newly released movies and provide their subjective views and comments on the movie for the public's information" [23, p. 120]. Several studies show that expert reviews have a significant impact on explaining box office receipts [e.g., 5, 82], but findings differ as to *when* expert reviews influence movie sales. Building on prior research, we explain when and why expert reviews influence movie sales and how they differ from peer comments platforms.

Comments from Different Peer Comments Platforms. While prior research has compared the effects of experts versus peer opinions [69, 74, 97], prior studies have not differentiated between peer opinions originating from different platforms. The increasing proliferation of technologies provides more avenues through which online information reaches consumers [71]. Marketers are also using a broad array of social media tools (ranging from Twitter to user forums) to listen to and interact with customers. A logical approach to classify platforms providing peer opinions is to differentiate platforms based on how consumers obtain information from the platform: pull- versus push-based [54].

As consumers increasingly look to peers for information, there are now many websites available that serve as a forum for consumers to interact and share their experiences [13]. The success of Internet sites such as Epinions.com and Tripadvisor.com show that consumers value a place for them to share product reviews and consumption experiences. Given that consumers need to perform a search before they can locate information, such platforms require consumers to pull information from these sites [71]. Hence, we classify online discussion forums and user ratings websites as pull-based platforms, as peer comments from such channels generally reach other consumers only if the consumers search for and browse the content in the online discussion forums. [1]

In contrast, microblogs represent a push-based approach of distributing information. Microblogging is a form of communication in which users send short messages (usually less than 200

[1] Major movie websites (e.g. IMDB, RottenTomatoes, etc.) can push updates and news-like messages about movies to users via RSS or Twitter, but they do not push peer comments. Generally, user comments posted on discussion forums are not pushed to users, likely because of the large volume of posts, which may be interpreted as spam. Therefore posts made in user forums are basically not pushed to but pulled by moviegoers.

characters) received via instant messages, mobile phones, email or the web [51]. Twitter, for example, is the most popular form of microblog in US, Europe and parts of Asia. Individuals predominantly receive messages from sources they have elected to follow. Following a source implies that the short messages generated by the source will be automatically pushed onto the Twitter pages of all his/her followers [52]. [2]

We argue that it is important to compare the persuasiveness and impact of messages that consumers receive from these push and pull sources because they represent different ways through which consumers receive and process information about a product.

Choice Uncertainty and Information Behavior

As one of the strongest driving forces behind individuals' information behavior, choice uncertainty offers a coherent theoretical frame for understanding how online comments from various sources are factored into consumer decision making [e.g. 61, 87]. The concept of choice uncertainty dates back several decades [60, 73], and relates to information search, and decision making. It refers to "a consumer's uncertainty as to which alternative out of a considered set to choose" [89, p. 212], or the lack of confidence about the extent to which a single alternative fulfils the needs of a consumer. The more the alternatives, and the more equiprobable the various choices, the greater the degree of uncertainty [27]. Greater uncertainty also arises when information available about the choices deviate from consumers' ideal informational state [15, 87]. In an era of abundance, consumers' needs are diverse and their considered set is large, creating ambiguity of which choice presents the best alternative [27].

The linkage between choice uncertainty and external information search is widely acknowledged [84]. Decision theorists consider choice uncertainty as the key driver motivating information search for decision making [80, 83]. Psychologists highlight that greater choice uncertainty increases active search because of the urge to resolve conflicts arising from indecision [e.g. 9, 60]. There is clear empirical evidence attesting to the relationship between choice uncertainty and information seeking. Consumers with high

[2] According to Alexa (http://www.alexa.com/siteinfo/twitter.com), the bounce rate of Twitter is 44%, indicating that almost half of its users navigate away from Twitter after viewing only one page. Thus, the probability of engaging in search within Twitter is low. Researchers have also highlighted that limited searches are made on Twitter [36].

levels of choice uncertainty are more likely to consult others or consumer reports [89] and to increase their online search behavior [61]. Drawing on this research, we examine when online comments from different sources and platforms affect consumers' decision-making.

The Dynamic Influence of Online Comments

The effects of online comments on movie sales are not static. By assimilating new information and combining them with prior information, people actively and continuously update their views on various issues. Hence, the gap between individuals' ideal informational state and the information they possess decrease over time, as users gain more information [57]. The level of uncertainty that consumers experience thus correspondingly decreases as more information about each option becomes available. As their levels of confidence about the options change, consumers adopt different information behaviors [87].

The theory of information search states that consumers will weigh the benefits of information search with the associated search costs. The greater the need for information due to a higher level of choice uncertainty, the greater the perceived benefits of search, as more information helps consumers to better evaluate their options. Hence, when uncertainty is high, the benefits of search outweigh the costs of search [87]. Over time, as consumers continuously refresh their view of the world, the returns of search diminish as suggested in the economics of information theory, which states that consumers quickly reach the point where expected benefits of search no longer cover the costs of search [57, 87].

We believe the movie industry presents a particularly interesting context to examine how the effects of online comments differ over a product's lifecycle. The movie industry is characterized by short lifecycles and exponentially decaying adoption patterns [48]. The compressed lifecycle is expected to amplify the differential impacts of online comments at different stages of the product lifecycle. Recognizing this, prior research examining the commercial impact of online comments of movies has often provided a week-by-week breakdown in their analysis, to explore the dynamic aspects of the phenomenon [e.g., 69]. Table 1 shows a summary of the findings from these studies; we highlight two main conclusions from Table 1. First, there is much inconsistency in the findings of the impact of online comments over time. Second, there is a lack of coherent theory considering why online comments may differ in their impact across time.

Prior research usually examines the influence of online opinion on a product via the valence and volume of the online comments [e.g., 96, 98]. Valence refers to the sentiment expressed in the content of the comment. It is a continuum with two extremes indicating either absolute positive or negative attitude. The higher the valence, the more favorably the consumer perceives the product. Volume of online reviews refers to the quantity of comments, which reflects the frequency that individuals encounter information about a product. There are inconsistent findings in the literature concerning the impact of both valence and volume of online comments from both experts and peers over time.

In terms of expert reviews, Eliashberg and Shugan [35] found that the valence of expert reviews are not correlated with early box office but are associated with late and cumulative box office. However, Basuroy et al. [5] found that the valence of expert reviews are correlated with box office receipts throughout the first 8 weeks of movie screening, while Moon et al. [74] and Liu [69] found that experts had an impact only on early box office receipts.[3] In terms of peer comments online, prior research has found the relationship between valence of peer comments and movie sales to be either: (1) insignificant throughout all weeks of analysis [32, 53, 69]; (2) positive and significant only in the earlier weeks [4, 74]; (3) positive and significant in aggregate (not differentiated by weeks) [21], or even (4) negative in all weeks, in the case of microblogs [48]. The results regarding the impact of the volume of peer ratings is just as mixed. The relationship between volume of peer comments and movie sales has been found to be either: (1) positive and significant, especially in earlier weeks [4, 32, 69]; (2) insignificant in aggregate [21]; or even (3) negative across all weeks [48]. [4]

The confusion caused by the inconsistent findings regarding the influence of valence and volume of expert and peer comments online on movie sales is exacerbated by the lack of an overarching set of theoretical arguments to help us to understand why online comments from different sources and platforms might

[3] Although some researchers have juxtaposed expert and peer opinions in their studies, they do not explore whether and how these opinions affect box office at different stages [43, 74, 95].
[4] Table 1 listed two papers on push-based comments [4, 47], both of which collected data from Twitter for one or two weeks since movie release, and did not compare tweets with other online peer comments.

influence consumers' comments at different points in time. The discussions about the differential impact over time are usually post hoc, or merely descriptive with little emphasis on theoretical arguments to explain the findings [e.g., 69], while some studies focus on methodological improvements to resolve inconsistencies in the findings [e.g., 21, 42]. There is thus a significant need for research to provide a set of coherent theoretical arguments that might explain not only why online comments may differ across time, but also how that might differ for different sources and platforms providing online comments.

HYPOTHESIS DEVELOPMENT

We provide an overarching set of theoretical arguments, premised upon how uncertainty drives consumers' information seeking and processing behaviors over time. Prior research highlights that the initial weeks of a movie's release is usually characterized by a high level of uncertainty [18]. The lack of information and heavy advertising on the part of the studios may make it difficult for moviegoers to adequately assess the quality of movies and whether the chosen movie conforms to their taste [74]. There are thus significant perceived benefits to information search in the early stages of a movie's release. When consumers are faced with high choice uncertainty (i.e. uncertainty regarding which alternative to choose), they are more likely to increase their search [25, 89]. Hence, we argue that consumers are more likely to search and pay attention to both expert reviews and pull-based consumer reviews and online forums in the early stages of a movie's release. After the initial weeks of a movie's release, the box office sales and ranking of a movie becomes widely available and publicized [18, 33]. This provides important information to the consumer, significantly reducing the uncertainty about the quality and popularity of a movie, making it unnecessary to search for information actively, and reducing the impact of expert reviews and pull-based peer comments. Instead, we expect that in late stages, another form of information search – passive attention will dominate, through which consumers are confronted with information from repeated microblog posts. Awareness of the movie will be increased and motivate consumers to purchase.

In summary, we predict that the different levels of uncertainty that consumers face at initial and later stages of a movie's release affects their information seeking and processing behavior and thus the types of online information sources they pay attention to. Detailed arguments are provided below.

Expert Reviews and Early Box Office Receipts

Two alternative views on the influence of expert reviews have been proposed – experts as the influencer or the predictor [5, 35]. According to the influencer perspective, movie experts are opinion leaders, influencing uninformed consumers to follow experts' picks early in a movie's release; hence their opinions are expected to affect moviegoers mainly during the early period. In contrast, as predictors, movie experts are not expected to affect consumers' choice, but merely reflect their preferences, as they have similar tastes to the average consumer, or they have learned to express opinions representative of their readers. Accordingly, expert reviews are correlated with the box office revenue in the later weeks and the entire run. Empirical evidence have been found support for both the role of experts as predictors [35] and as influencers [74], while others found inconclusive evidence [95].

As the initial weeks of a movie's release are characterized by significant uncertainty, we argue that expert reviews provide information to reduce consumers' uncertainty. When a movie is initially released, there is a lack of information regarding the sensory and experiential aspects of movies. Information like the movie cast, directors, and budget is far from sufficient for consumers to assess the experience the movie will provide [31]. In contrast, expert reviews are able to provide a sneak preview of new movies, based on which consumers can make their purchase decision, as critics are often invited to early screenings of movies [5]. As professionals, experts are anticipated to deliver unbiased, reliable and informative reviews to inform purchase decisions [86]. In the later stages of a movie release, with the availability of more information and the publicity of the box office rankings, the level of uncertainty will be significantly reduced and the informational role of the expert will diminish. Hence, we expect the impact of expert reviews to fade over time. We thus expect expert reviewers to take the role of influencer rather than predictor. Hence:

H1: The valence of expert opinions has a positive effect on movie box office receipts during the early stages of a movie's release, and the effects will decrease over time.

Consistent with recent studies [e.g. 21, 43, 70, 74], we do not examine the volume of expert comments, as the number of experts reviewing a movie does not exhibit significant variance across movies, and studies have found that the number of expert reviews do not influence movie sales [5, 34].

Pull-based Peer Reviews from Forums and Early Box Office Receipts

In addition to expert reviews, consumers also rely on peer comments from pull-based platforms or online forums. The motivation to reduce uncertainty is likely to prompt active searching and browsing of information on forums. Pull-based websites like online forums, discussion boards and consumer ratings websites encourage active consumer participation in sharing opinions about specific products [11, 13, 50]. The sentiment embedded in peer comments are expected to exert a persuasive effect because of the perceived ideological similarity amongst consumers, given that both early consumers and those providing comments are peer consumers who are interested in watching movies in the early weeks of a movie's release [31, 76, 88, 90]. The sentiments embedded in online comments reveal information about product quality [64] or individuals' feelings about interacting with the product and the extent to which their expectations were met [19, 72], thus influencing consumers' attitudes about the movie.

While the valence of online opinion influences consumers by way of changing individuals' attitudes about a product, prior research has shown that the volume of online opinion influences consumers by way of signaling the popularity and attention that a particular movie is generating. A higher volume of online comments generates a higher level of awareness of the movie, attracting individuals' attention to the movie [14]. Hence, the volume of online comments from pull-based outlets in the early stages of a movie's release is an important source of information to reduce consumers' choice uncertainty.

However, as information about a movie (e.g. box office sales, movie ranking) becomes readily accessible to the public over time, uncertainty about a movie's quality and performance is significantly reduced. With reduced uncertainty, consumers will be more likely to discontinue active search when they believe they already have sufficient information about a product [58]. This echoes the view of decision theorists who describe economic behaviors as adaptive sequential decision making centered around tradeoffs between search costs and benefits [56]. When the market response to the focal movie becomes easily observed in later stages, the costs outweigh benefits of seeking and processing peer opinions on the product in pull based platforms, due to their lower informational value then. We therefore expect consumers to gradually discontinue pulling comments from forum-like platforms over time.

H2a: The valence of pull-based peer opinions has a positive effect on movie box office receipts in the early stages of a movie's release, and the effects will decrease over time.

H2b: The volume of pull-based peer opinions has a positive effect on movie box office receipts in the early stages of a movie's release, and the effects will decrease over time.

Push-Based Peer Reviews from Microblogs and Late Box Office Receipts

As consumers reduce their search for information, either from pull-based platforms like forums or from expert reviews at later stages of a movie's release, we expect consumers to activate other forms of information assimilation – namely passive attention. Prior studies tend to focus on active information search, implicitly regarding external information search to be the principle mode of information acquisition [45, 63]. Information acquisition, however, may happen without intentional seeking from external sources [92]. We call attention to another type of information behavior – being confronted with information, which is also termed passive attention [10, 92]. Passive attention refers to low involvement learning or attention due to interrupts, both of which suggests information acquisition without deliberate intentions [92]. When consumers acquire information in the mode of passive attention, they receive information without intentionally seeking them. Passive attention typically happens when one's attention is attracted unintentionally, as some existing behavior is interrupted [10, 92]. Passive attention will be dominant when consumers are not motivated to conduct external search, in the case of reduced choice uncertainty.

Push-based messages are displayed automatically, increasing the frequency with which a consumer is exposed to the product without request. When individuals review microblog messages pushed to them, they engage in undirected viewing – getting exposed to information with no specific informational need in mind. Such passive attention creates awareness of a product through repeated exposure to microblog messages pushed to them, increasing the availability and accessibility of the product in consumers' memory. After exposure to a product, subconscious psychological reactions can be provoked, as familiarity may increase people's inclination to develop a preference for a product [14, 26].

Microblog messages pushed to consumers trigger attention through passive monitoring, serving as reminders to consumers. Such messages serve to remind consumers about the movies and thus trigger the desire to visit the theatres. Hence, we suggest that such push-based messages serve a reminder rather than an informational role, in later stages of the movie's screening. We thus expect that the volume of push-based peer online opinion would be positively associated with box office sales only in the later stages. We do not expect the valence of push-based messages to influence the box office sales, as consumers are not searching for information in the push-based messages, but rather, the push-based messages trigger awareness and serve as reminders. The lower level of uncertainty at later stages of a movie's release imply that consumers are not motivated to put in extra effort to process, understand, and evaluate information. Hence, they may not process the messages in detail and pay attention to the sentiment of the messages they received. Therefore, we do not expect the valence of push-based peer opinions to affect movie sales, but hypothesize a positive effect of the volume of push-based on movie-going in later stages of a movie's release [4].

H3: The volume of push-based peer opinions has a positive effect on movie box office receipts in the later stages of a movie's release.

Table 2 summarizes the characteristics of the three types of online comments, and Figure 1 provides a summary of our hypotheses.

Insert Table 2 and Figure 1 about Here

RESEARCH METHODOLOGY

Data Collection

We combined various online sources to develop a dataset that allowed us to test the hypotheses. Broadly, three types of data were collected. First, we differentiated between expert and peer comments;

next for peer comments, we differentiated between pull-based peer comments and push-based peer comments. We adopted two main strategies to collect the data that was required. First, for expert and pull-based peer comments, we developed an online automated agent to crawl the Internet based on pre-specified search parameters (details later). Second, for push-based peer comments (e.g. Twitter), we partnered a social media monitoring and management company, which helped us to collect the comments from “live” feeds such as Twitter. Given the longitudinal nature of this study and as some online comments from microblogs are transitory; we conducted our data collection program *continuously* from October 2010 to October 2011. What is unique about our data collection methodology is that we considered numerous forms of online content. Unlike prior research, which focus on reviews from a single platform (e.g. Amazon or Yahoo!), our research considers a wide spectrum of online outlets thus providing a comprehensive view of the online activities surrounding any particular information channel.

In order to provide a comprehensive comparison of the impact of online opinions over time, we employ text mining techniques to analyze online opinions. Our study is not limited to comments with self-reported ratings, which constitute a small proportion of all posted online comments. Here we see compelling reasons to use text mining because using only online comments and reviews with self-reported rating scale results in a tremendous loss of data, and a natural biased censoring of the data. Without using text mining for sentiment analysis, one will naturally exclude consumer posts in social media outlets like forums, Twitter and discussion boards (that generally do not have rating scales), which now represents an overwhelming proportion of social media activity. This exclusion undermines the goal to acquire a comprehensive understanding of the effect of online comments across sources. Second, to compare across social media sources, we have to consider online comments from “live” feeds such as Twitter. Social media outlets like Twitter are highly dynamic and transitory; over time not all tweets are always publicly searchable and such data might not be easily captured after the information generating event. To ensure that we capture all relevant, time sensitive data, we maintain a constant “live” connection to such social media outlet and continuously track the information at source for over a year.

We collected data on major US cinematic movies that were screened from October 2010 to October 2011. First, we obtained the population list of movies released during the said period from IMDb (www.imdb.com) – a comprehensive online movie databases that monitors movies worldwide. We triangulated this list with an alternative site: Rotten Tomatoes (www.rottentomatoes.com) to ensure further accuracy. We excluded documentaries, movies released directly to DVDs (i.e. not screened in theatres), and movies that were released and screened only in film festivals, as these have limited market coverage and are not equally accessible to the entire US market. Such movies also tend to have limited online comments and lack daily box office records. In total, we collected data for 236 movies.

For this sample of movies, we generated a movie keyword list based on the names of movies. The movie keyword included the movie title in its entirety, as well as parts of the title omitting non-essential punctuations, articles, pronouns, prepositions and conjunctions whenever necessary. For example, keywords for "Transformers: Dark of the Moon" include the full movie title, and the phrases "Dark of the Moon" and "Transformers 3". For each movie, we capture all comments for two weeks prior to a movie's release, throughout the time it is screened in cinemas, and for two weeks after the movie has been removed from the cinemas.

Expert Reviews. These refer to movie reviews provided by experts, usually affiliated with mainstream press such as a major newspaper or magazine (e.g. Entertainment Weekly), radio station, or website (e.g. eFilmCritic.com). We considered various websites that aggregate movie reviews from different sources, which represent the majority of all expert reviews: Yahoo, Rotten Tomatoes, IMDb, Rolling Stone, PopMatters, and MRQE (www.yahoo.com/movies/;www.rottentomatoes.com; www.imdb.com; www.rollingstone.com/movies/; www.popmatters.com/pm/; www.mrqe.com/movies/). When these sites provide both expert as well as regular, peer user reviews, both types of reviews are *clearly* labelled and categorized based on the status of individuals posting the review. For example, in IMDb.com, expert reviews are classified as "critic" and non-expert reviews are classified as "user". We tested the comprehensiveness of movie reviews collected by randomly using Google search to test if it would surface movie reviews not included in our sample. Our ad-hoc searches did not yield any new reviews within the first page of search listings. Each week, we downloaded the reviewer's name, post date of review, review content, title of the review and also the review rating (if any) of the movies released in the preceding week. Our dataset covers expert reviews from more than 1,500 websites. We determined the valence of each expert review (details later), removing any duplicate reviews that may have been cross-posted on different websites.

Pull-Based Peer Comments – Online Forums. [5] [6] These refer to information and comments mainly generated via online forums, discussion boards, and consumer ratings websites, posted by consumers and moviegoers (not expert reviewers). To read these comments, users have to actively search on the Internet for them or browse the websites containing these comments; hence they are considered pull-based comments. Each week, we collected information by searching for consumer reviews using the movie keyword list, from major movie comments/review aggregator sites (IMDb, Rotten Tomatoes, Yahoo, Rolling Stone), discussion forums, and consumer rating websites.

Push-Based Peer Comments – Microblogs. New posts on Twitter constitute the source of push-based peer comments. [7] Twitter is subscription based – only those who subscribe to feeds provided by an individual will receive the information she or he sends. Thus most information propagated via Twitter is considered to represent information received from the parties we identify with or have interest in. To collect

---

[5] User reviews are collected from major movie websites (e.g. IMDb, RottenTomatoes, Metacritic, etc.), which categorize user reviews differently from expert reviews. Additionally, we collected comments from discussion forums indexed by BoardReader, which is a forum search engine. Specifically, we input keywords derived from the title of a focal movie, and collect results returned by BoardReader as raw data for user reviews on the movie. The returned results cover not only movie forums, but also forums covering various topics. Finally, we randomly inspected the dataset and verified that our data does not contain expert reviews.

[6] We recognize that various movie websites (e.g. IMDb.com) provide RSS feeds to subscribers and push out messages to these subscribers. We did not include these as push-based peer comments as these messages are mainly news articles about upcoming movies and not user comments about movies.

[7] Due to the sheer market share and popularity of Twitter in the microblog platform segment, many studies collect tweets to represent data from microblogging services [e.g., 59, 66] As Hennig-Thurau, et al. [48] pointed out, "although various microblogging services exist, Twitter has become synonymous with the concept". Therefore we believe Twitter, as the market leader, is properly representative of the microblog platform.

all relevant push-based comments, we maintain a constant “live” link to the platform provider, monitoring and archiving all relevant traffic. Every week, we input the movie keywords into a tool developed by a social media management company. The tool will monitor all public feeds from Twitter, and capture feeds that include words matching the list of keywords.

Data Cleaning

Our data collection methodology is comprehensive as we cast a wide search for social media content, with a total of almost 10 gigabytes of *relevant* online comments in plain text format collected for the data collection exercise. The challenge, however, is that this resulted in a demanding data cleaning and preparation processes on various fronts. Specifically, there is a non-trivial amount of irrelevant data collected in the process especially when the keywords for a movie are generic (e.g. Source Code) Nevertheless, we believe this methodology provides a granular, yet comprehensive snapshot of the social media content. In order to filter out data that contains the keywords but does not concern the movie, we carried out the procedure described in Appendix B.

Measures

Box office receipts. We collect the daily box office information from the Box Office Mojo website (www.boxofficemojo.com, a subsidiary of IMDb), for each day that the movie is screened in cinemas in the US. For each movie, we track the daily box office receipts for an average of 66 days.

Valence of Online Comments. In prior studies, it is a common practice to measure the valence of a review using self-reported numerical ratings of a review [e.g., 32, 69]. However, this often results in a loss of much of the data, as many reviews do not provide numerical ratings. In some cases, the numerical ratings may even distort the actual sentiment conveyed in the textual review as some websites provide a default rating of zero or a median rating (e.g. 3 out of 5) when reviewers do not provide any ratings, even though the sentiment conveyed in the text of a review is obviously positive or negative. Godes and Mayzlin [41] manually coded a sample of peer reviews, and found that the correlation between self-reported ratings and manually coded data is less than 0.2, suggesting a weak relationship. Review ratings also convey limited information about a reviewer's sentiment and consumers often read the review text rather than depend solely on the summary statistics provided in the ratings [20]. Prior research, recognizing that review ratings may provide inappropriate or limited information, has begun to use textual analysis to capture the valence of online reviews [e.g., 39].

To accurately measure the sentiment of a comment, we adopt a machine learning method for sentiment classification, similar to prior research [e.g., 17, 24, 85]. The essence of this technique involves randomly selecting a sample of the messages that have numerical review ratings attached to them by the message author. The sample of messages is then codified into a set of vectors for machine learning [93]. Using these trained vectors, we predict the probabilities that new messages would be positive (or negative) by the similarity in the message structure. The assumption made here is that messages with similar words are likely to have similar sentiment value. Classification techniques with similar theoretical basis has been applied in prior IS research, whereby the authors used a Naïve Bayes algorithm to train a subset of messages to recognize the significance of particular words that will impact the sentiment of the entire message [3, 44]. Our analysis tool gives us an estimate of the probability that a given text is positive, with 0 meaning absolutely negative and 1 meaning absolutely positive. To ensure that our sentiment measurement is robust, we triangulate our sentiment classification (i.e. positive or negative) against the sentiment classification provided by a social media management company's propriety sentiment analysis tool. We found relatively high inter-rater (computerized classification) reliability of 0.852, 0.822 and 0.926 for expert reviews, pull-based peer comments (forums) and push-based peer comments (microblogs) respectively. The correlation between expert numerical ratings and the estimates from the tool was as high as 0.821. Appendix C provides more details of the algorithm used.

Volume of Online Comments. We measure volume with the number of online comments concerning a particular movie. Volume figures are compiled on a daily basis for forums and microblogs. These figures are subsequently aggregated as required by the empirical specification.

Interaction Variables. To quantify the changes in effects of the valence and volume of online comments over time, we compute interaction variables between the valence and volume of the online comments and time (number of days since release). [8]

Control Variables. As our analyses show that fixed effects estimation is favored, control variables that are invariant across movie and time are dropped from our analysis as they violate full rank assumption. Nevertheless, we include a random effects model that accommodates controls for the sake of completeness. In this additional check, we include the following controls: (1) budget and popularity of the first and second leading artists, which are usually indicators of the hype surrounding a movie. IMDb Pro (pro.imdb.com) provides a STARmeter ranking for all actors and actresses, based on the inputs of millions of IMDb users. The more popular the actor/actress, the smaller his/her rank (i.e. the most popular actor/actress is ranked number 1). We use the STARmeter rankings to measure the popularity of the leading artists. (2) As per prior research, we also control for MPAA ratings (Ratings: R and PG) and genre. Consistent with prior studies [5, 12, 29, 69, 82], we select the two most common genres (Drama and Thriller) from our sample. Descriptions of all variables used in the analyses are provided in Table 3, while variable descriptive statistics and correlations are given in Table 4.

Insert Tables 3 and 4 about Here

DATA ANALYSIS AND RESULTS

We group the observations by movie and present the data generation process by the following functional form (see equation 1). The equation used in this research is:

$$Y_{it} = \alpha + X_{it}\beta + u_i + \varepsilon_{it} \qquad (1)$$

where Y denotes daily box office receipts, i denotes the movie, and t denotes the number of days since release. X is the vector of variables including valence and volume of online opinions, key interaction variables between valence, volume and time, and control variables. u_i represents movie level stochasticity, ε represents stochasticity and β represents estimated parameters. Multicollinearity is not a significant

[8] We center the main effects to compute the interaction terms. The beta values are identical with or without centering (with the exception of the intercept).

problem since the VIF values for all independent variables are less than 5. We first test if a fixed or random effects estimator will be more appropriate in estimating the movie level stochasticity. We estimate both fixed and random effects and found that the random effects estimators are inconsistent compared to the fixed effects estimators (Hausman χ^2 = 206.94, *p*-value < 0.001). Further, unlike random effects, fixed effects estimation does *not* require the assumption whereby the movie level error term, u_i , have to be independent of the explanatory variables, X_{it} which is more realistic in the setting given that individual movies are likely to be correlated with volume and valence of comments. Hence, we use the fixed effects specification in the additional estimators described below. Given that fixed effects estimation is favored, control variables that were invariant across movie and time were dropped from the functional form as they violate full rank assumption.

The functional form might suffer from endogeneity concerns as the volume and valence of user comments for a particular movie may be influenced by and correlated with unmeasured factors that are inherent with contributors of online comments (i.e. non-movie related factors but contributors related factors). To address this potential issue, we employed similar instrument variables as documented in Chintagunta et al. [21] (volume of pull-based comments from competing movies, rank of lead artist of competing movies, and rank of Director of competing movies). To construct the instruments for each relevant movie in the panel, we compile a list of movies that are screened in theatre in addition to the focal movie; we label these movies as "competitor movies". The competitor movies provide a good set of instrument variables given that they are likely to be provided by a similar set of online reviewers, hence, online contributors of comments of competitor movies are likely to be similar and correlated with the online contributors for the focal current movie. Given the similarity of online contributors, measures pertaining to competitor movies such as volume of pulled based comments, rank of lead artist, and rank of Director are likely to be correlated with the volume and valence of user comments of the current movie. Notably, the instruments from competitor movies are good indicators of the customer base, correlates with unobservable factors such as the activeness and level of online participation of the movie going audience, hence making them predictors of valence and volume of current movies. These instruments however are not likely to be correlated to the sales of the current movies, which are more likely to be driven by the characteristics of the *referent* movie. As presented in Hayashi [46] an efficient and consistent GMM estimator for the functional form is:

$$\beta_{GMM} = \left(X' \, Z S^{-1} Z' \, X\right)^{-1} X' \, Z S^{-1} Z' \, y \qquad (2)$$

Where Z represents the vector of instrument variables and S^{-1} is the optimal weighting matrix for the GMM estimator. In our estimation, we consider three scenarios – one baseline and two other with weighting matrices addressing different empirical concerns. We triangulate the findings of each alternative estimator for robustness. For the baseline case, we consider $S^{-1}$to be conditional homoskedastic; we label this estimator IV-GMM (I). Next, in the first of the two weighting matrices, we adopt the Eicker-Huber-White robust covariance estimator, as suggested in Hayashi [46]. This weighting matrix considers the variance-covariance structure of the instrument variables and the residuals of equation (1), which particularly addresses the concerns of endogeneity and heteroscedasticity of disturbances. Specifically:

$$S = \frac{1}{n}\left(Z' \, \widehat{\Omega} Z\right) \qquad (3)$$

where $\widehat{\Omega}$ represents the diagonal matrix of squared residuals ε^2 from equation (1) – the consistent but not necessarily efficient first step estimator. We label this set of estimators IV-GMM (II).

The second weighting matrix that we use considers the issue of serial correlations (in addition to endogeneity issues) between the user comment measures and the error term across time. Given that online comments posted are available for an extended period of time, daily sales of subsequent days might be correlated with the comments posted earlier. Although we control for timing differences in our functional form, the potential problem of autocorrelation requires more adjustments to the weighting matrix (3). The most direct way of addressing both autocorrelation and endogeneity problems is to present an autovariance weighting matrix, S_a, which comprises of elements that address both endogeneity and autocorrelation [6].

$$S_a = \Gamma_0 + \sum_{j=1}^{q} k\left(\frac{j}{q_i}\right)\left(\Gamma_j + \Gamma_j{}'\right) \qquad (4a)$$

with $\Gamma_j = E\left(\varepsilon_t \varepsilon_{t-j} Z_t^{'} Z_{t-j}\right)$ (4b)

where j represents the number of lags (in days), and q represents the number of days the movie was screened in total. Given that the first term, Γ_0 in S_a (of equation 4a) has zero lag, it simply accounts for the endogeneity through instruments Z. The remaining second term represents the autocorrelation adjustment approach of Newey and West [77] with $k(\cdot)$ representing the Bartlett kernel function. The above specification is similar to the one adopted by Chintagunta et al. [21] and the estimates presented are commonly known to be *HAC*: heteroskedasiticy and auto-correlation consistent estimates. The set of β estimates is obtained by using S_a as the weighting matrix for equation (2). We label this final set of estimators IV-GMM (III). Table 5 provides the estimates for our analysis.

Insert Table 5 about Here

In Table 5, we present five estimators [Random Effects, Fixed Effects, IV-GMM (I), IV-GMM (II) and IV-GMM (III)]. [9] Although the random effects estimator is inconsistent compared to the fixed effects estimator, we present it for comparison purposes, given it is the only estimator that considers movie level control variables. It is important to note that the random effects estimator is not expected to provide accurate estimates given the inconsistency. Notably, although some of the coefficients are not statistically significant compared to the final estimators the findings are based upon [i.e. IV-GMM (III)], the coefficients retain the same directionality. The inconsistency in the RE estimators result in some insignificant coefficients. The fixed effect estimators are also presented to establish the baseline case whereby we do not consider issues of endogeneity, heteroscedasticity and autocorrelation. For the other estimates, IV-GMM (I) considers the issue of endogeneity; IV-GMM (II) accounts for both endogeneity and heteroscedasticity and IV-GMM (III) accounts for endogeneity, heteroscedasticity and autocorrelation. We show all five estimators to highlight the progressive effects of adding additional controls, but the results from the IV-GMM (III) are

[9] The analyses reported here use a 7-day moving average of the valence and volume of the reviews as key independent variables. To validate the robustness of our analyses, we computed 5, 6, 8, 9, 10, 11, 12, 13 and 14-day moving averages for the reviews' valence and volume, and found qualitatively similar results for each of the 5 estimators: RE, FE, IV-GMM (I), IV-GMM (II) and IV-GMM (III).

expected to be the most robust. We base our findings on IV-GMM (III), but we note that the coefficients of estimators basically have same directionality.

The results suggested good model fit and explanatory power for the variance in daily box office receipts, and provided support for the hypotheses. The R^2 for all models range from 0.303 to 0.409. Although the random effects model have more variables as it included movie and time invariant control variables, the R^2 is only marginally higher compared to the other fixed effects models. This suggests marginal contribution of explained variance by the additional control variables beyond the fixed effects.

Based on Aiken and West [1], we probe the interaction effects. Figures 2 and 3 plot the relationships between valence of experts, pull-based and push-based peer comments over time (days since movie release), and volume of pull-based and push-based peer comments over time respectively.

Insert Figures 2 and 3 about Here

Findings

All three GMM estimates yield qualitatively similar results. For conciseness, we discuss the results from model IV-GMM (III) as these results take care of issues related to endogeneity, heteroscedasticity and autocorrelation. We first examine how the valence of expert comments influences box office receipts. Table 5 shows that the valence of expert reviews is positively and significantly associated with a movie's box office receipts (β = 2.537, p-value <0.001). The effect size of the expert review valence is the highest among all three sources (expert, push and pull) of online comments. The positive impact of the valence of expert comments decreases over time (β = -0.356, p-value <0.001), as shown in Figure 2, providing support for H1.

Next, we examine how the valence and volume of peer opinions in pull-based platforms (forums) influence movie sales. Table 5 shows that the both the valence (β = 1.584, p-value<0.001) and volume (β = 0.142, p-value<0.001) of forum comments are positively associated with movie sales. Similar to the trends shown in expert reviews, we found that the effects decrease over time, as indicated by the negative and significant coefficients of the interaction terms of Valence * Ln(Days) (β = -0.217, p-value<0.05) and

Volume * Ln(Days) (β = -0.031, p-value<0.001) for forum comments. These results suggest that although increased volume and positive comments from pull-based platforms in the prior week improved movie sales, there is a decline in the impact of volume and valence of the comments as time progresses, as shown in Figures 2 and 3. This provides support for H2a and H2b.

Finally, we examine how the valence and volume of peer opinions in push-based platforms (microblogs) influence the box office receipts of movies. Table 5 shows that the coefficients of the valence and volume of microblog comments are not significant ($\beta_{valence}$ = 0.248, p-value >0.05; β_{volume} = 0.002, p-value>0.05), and does not influence box office receipts. The coefficient of Push Valence * Ln(Days) (β = 0.032, p-value>0.05) is insignificant, but Push Volume * Ln(Days) (β = 0.012, p-value<0.05) is significant and positive. These results, as shown in Figures 2 and 3, suggest that at the beginning, although the volume and valence of weekly push-based comments do not have a significant impact on movie sales, over time, volume in push-based platforms has an increasingly positive impact on movie sales, supporting H3.

Figure 4 shows the summarized results of all the hypotheses. The findings thus far suggest that although online comments in general have a significant impact on movie sales, the impact of comments from different platform and sources are not uniform across platforms and vary over time. Expert comments have a strong impact on sales in the initial product release and the influence decays over time. Pull-based comments from forums exhibit a similar trend – they have a significant impact on sales during the initial release of the movie, and the impact reduces over time. Push-based comments from microblogs, on the other hand, have the opposite trend. They do not have a significant impact on sales at the beginning of the movie release, but their contribution to sales (via volume) improves over time.[10]

Insert Figure 4 about Here

[10] In our data, there are a small number of comments (less than 0.7%) that span both pull- and push- based platforms. For example, expert reviews might be tweeted or repeated in forums, and forum comments may be forwarded via microblogs and vice versa. Based on the final source where the information is captured, such messages would have been classified as pushed or pull-based messages even if they originated from a different source. However, to ensure that messages that span across platforms do not confound our findings, we perform a robustness check by dropping such data. The analysis, shown in Online Appendix D, shows that our results remain unchanged.

DISCUSSION

Although prior studies examine the impact of online comments, they tend to focus on one or limited online platforms, and they do not provide insights into how online comments from multiple platforms of differing nature may influence product sales differently [4, 55, 73]. Through a comprehensive data collection effort, we consider not only different types of online comments, but also multiple sources of data for each type of review to obtain a complete picture of how different types of online comments affect movie sales. In total, we collected expert comments from more than 1,500 websites, pull-based peer comments from 4 major movie aggregator sites and various discussion forums, and push-based peer comments from a key microblog: Twitter. By segregating the influence of online comments by source, and by examining their influence on movies' box office receipts across time, we offer a set of coherent theoretical arguments that explain when and why online comments originating from different sources and platforms affect movie sales. With prior research focusing only on one or limited sources of information, exhibiting much inconsistent findings and lacking theory to explain the findings, we believe that our research provides a significant contribution to the literature, advancing the area of research both theoretically and empirically.

On the premise that there is higher uncertainty about movie quality in the initial weeks of a movie's screening, we propose that the higher uncertainty imply that consumers will engage in more active search to reduce the uncertainty. Hence, consumers will be more reliant on pull-based expert reviews and peer comments for information about the movie quality and experience. In line with these arguments, we found that the valence of expert reviews, as well as both the valence and volume of pull-based peer comments from forums, have a positive and significant influence on box office receipts in the early stage of movie screening, and the influence decreases over time.

Over time, consumers are able to obtain more information about a movie, especially via the highly publicized box office rankings that become available after the first week of a movie's screening. Such information, along with greater word of mouth, decreases the uncertainty with respect to movie quality and popularity. As predicted by the theory of information search, individuals weigh the costs and benefits of search. As information uncertainty reduces over time, the reduction in benefits of search results in decreased search by the consumers. As a result, we argue that in the later stages after a movie's release, consumers are less likely to engage in active search or rely on pull-based expert and peer reviews.

In the later stages of a movie's release, we argue that another form of information search dominates – passive attention. As more information becomes available about the movie, individuals are not motivated to engage in active search. Through passive attention, individuals are confronted with information from repeated microblog messages, which increase individuals' awareness of the movie, such that consumers are motivated to watch the movie due to the reminders they receive. Hence, we argue that push-based platforms like microblogs are likely to be more significant in influencing later box office receipts. In line with these arguments, we found that the volume of push-based peer comments became more positive and significant over time. As expected, the valence of microblog comments did not play a significant reminder role, highlighting that the volume of push-based messages increases the awareness of consumers, regardless of whether the messages are positive or negative. Overall, our findings extend the extant literature by considering a wider scope of sources and more granular temporal sampling.

Limitations

We leave some worthwhile issues for future research. First, the strength of our research approach lies in the comprehensiveness of the data collection around the same set of movies from different platforms over a one-year period. In comparing across different social media platforms, we aggregated the sentiment reflected in the social media comments from the same type of platform or source. There, however, could be differences in the extent to which specific forum sites, due to its unique set of characteristics, tend to be more influential in affecting movie sales. Similarly, some microblog commenters and some expert reviewers may be more influential than others. Hence, in addition to the source and information delivery mechanism, the influence of specific comments, commenters and sites are likely to vary. In other words, some comments may be read by a lot more consumers and are likely to make a greater impact. However, it is impractical for us to collect data on the number of views of a particular comment, given the scale of our data collection. The analysis of influential sites and commenters will also imply a different research design. These are interesting areas for future research.

Second, our study is focused on online comments and did not include offline conversations. The objective of our research is to compare online comments from different outlets. Future research can extend our work by juxtaposing online and offline comments to gain more insights into the relationship between the two and their effects on product sales.

Finally, although our study is based on cinematic movies, we believe that the findings will be applicable to the context of most experience goods that have similar cost, demand and consumption patterns such as books and music albums. Nevertheless, it is important for future research to replicate and extend our study to other contexts. As the popularity of social media grows rapidly, the role of online comments becomes increasingly important. It is thus meaningful to explore whether and how consumers of different products react to online opinions.

Implications for Research

From a theoretical point of view, this research contributes to the list of emerging work on online WOM, with a focus on the dynamic influence of online WOM on economic outcomes. Extant literature has shown that the valence and volume of online opinions significantly influence product sales [e.g., 28, 49, 67], providing evidence that online opinions significantly affect consumers in their purchase decision for products and services. But significant inconsistencies have surfaced regarding when different types of online comments will influence movie sales, and there is a lack of theory providing a coherent set of explanations. Our study fills this research gap and provides a comprehensive understanding of how online comments influence box office receipts across time in several ways.

First, we provide a comparison of how different sources of information influence box office receipts. To the best of our knowledge, we are one of the first studies to provide a comparison of the effects of push and pull-based peer comments on product sales, in addition to a comparison of expert versus peer comments. We theorize and found that different sources of online comments differ in *when* they influence movie sales. This highlights that consumers regard differently and adopt different information processing approaches for dealing with distinct sources of online comments. Future research should thus consider the type of online platform and the characteristics of the online platform, and how it may influence users' processing of information from the platform.

Further, our results show the importance of comparing not only across different sources but also over time. The inconsistencies in prior research findings and the lack of explanations about when and why online comments affect box office receipts highlight the need for a theoretically driven approach to understand the problem. We present a theoretically derived model explaining how uncertainty at different stages of a movie's release results in different information search and processing behavior. We use this model to guide our hypotheses development and testing. We believe that future research can further expand this theoretical model to examine whether it can explain results in other settings regarding the influence of peer and other comments on consumer decisions.

Finally, our empirical data collection is among the most comprehensive published studies to date, as we collect different types of online comments, and we use multiple sites to collect each source of information. The comprehensiveness of the data collection provides us with more confidence that our findings are not influenced by the tendency of certain sites to attract certain types of consumers, and for us to conclude that online comments from different sites indeed have an influence on box office receipts.

Implications for Practice

Our research informs marketers that online opinions are not all equal. They differ not only in the extent to which they persuade or inform consumers, but also in terms of when they exert an influence. Based on the results of this study, we recommend marketers to employ a portfolio approach towards social media marketing where they focus on different channels at different times. Marketers should monitor the sentiment of pull-based peer comments during the initial release period of a product. In the subsequent stages of a product's release, however, they would need to pay more attention to push-based platforms for online comments. In addition, our results also show that expert reviews continue to play a critical role in influencing moviegoers, especially in the initial period of a movie's release. Hence, it is important for marketers to continue to ensure the availability of expert reviews for their products early on, and not neglect this group of reviewers in their pursuit of a marketing strategy focused on social media.

Conclusion

While prior research has examined the impact of online comments on product sales, limited research has examined whether information from different online platforms and from different sources differ in their influence on product sales. Our results provide important insights into when and why information from different sources and online platforms influence movie sales. Drawing on the theory of information search, we provide a coherent theoretical framework that explains how information uncertainty drives consumers' search behaviors. The comprehensiveness of our data collection, coupled with the use of a set of coherent theoretical arguments, provide both the theory and evidence explaining that pull-based online comments from experts and peers influence movie sales in the early stages of a movie's release, while push-based online comments from peers influence movie sales in the later stages. Our research demonstrates that online opinions are not always persuasive and useful, and our findings provide insights into when consumers are likely to pay attention to which types of online opinions.

REFERENCES

1. Aiken, L.S., and West, S.G. *Multiple regression: Testing and interpreting interactions*. Sage, 1991.
2. Amblee, N., and Bui, T. Harnessing the Influence of Social Proof in Online Shopping: The Effect of Electronic Word of Mouth on Sales of Digital Microproducts. *International Journal of Electronic Commerce*, 16, 2 (2011), 91-114.
3. Antweiler, W., and Frank, M.Z. Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance*, 59, 3 (2004), 1259-1294.
4. Asur, S., and Huberman, B. Predicting the future with social media. *Arxiv preprint arXiv:1003.5699* (2010).
5. Basuroy, S., Chatterjee, S., and Ravid, S.A. How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing*, 67, 4 (2003), 103-117.
6. Baum, C.F., Schaffer, M.E., and Stillman, S. Instrumental variables and GMM: Estimation and testing. *Stata journal*, 3, 1 (2003), 1-31.
7. Bayus, B.L. Word of mouth: the indirect effects of marketing efforts. *Journal of Advertising Research*, 25, 3 (1985), 31-39.
8. Berger, J., and Iyengar, R. Communication Channels and Word of Mouth: How the Medium Shapes the Message. *Journal of Consumer Research*, 40, 3 (2013), 567-579.

9. Berlyne, D.E. *Conflict, arousal, and curiosity*. McGraw-Hill, 1960.
10. Bettman, J.A. Consumer information acquisition and search strategies. *The effects of information on consumer and market behaviour*, Chicago: American Marketing Association, 1978.
11. Blackshaw, P., and Nazzaro, M. Consumer-generated media (CGM) 101: Word-of-mouth in the age of the web-fortified consumer. *A Nielsen BuzzMetrics White Paper, Second Edition, Spring* (2006).
12. Boatwright, P., Basuroy, S., and Kamakura, W. Reviewing the reviewers: The impact of individual film critics on box office performance. *Quantitative Marketing and Economics*, 5, 4 (2007), 401-425.
13. Brown, J., Broderick, A., and Lee, N. Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, 21, 3 (2007), 2-20.
14. Cacioppo, J., and Petty, R. Effects of message repetition on argument processing, recall, and persuasion. *Basic and Applied Social Psychology*, 10, 1 (1989), 3-12.
15. Case, D.O. *Looking for information: A survey of research on information seeking, needs and behavior*. Emerald Group Publishing, 2012.
16. Chaiken, S., and Maheswaran, D. Heuristic processing can bias systematic processing: Effects of source credibility, argument ambiguity, and task importance on attitude judgment. *Journal of Personality and Social Psychology*, 66, 3 (1994), 460-473.
17. Chang, C.C., and Lin, C.J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2, 3 (2011), 1-27.
18. Chen, Y., Liu, Y., and Zhang, J. When Do Third-Party Product Reviews Affect Firm Value and What Can Firms Do? The Case of Media Critics and Professional Movie Reviews. *Journal of Marketing*, 76, 2 (2012), 116-134.
19. Chen, Y., and Xie, J. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, 54, 3 (2008), 477-491.
20. Chevalier, J.A., and Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43, 3 (2006), 345-354.
21. Chintagunta, P.K., Gopinath, S., and Venkataraman, S. The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, 29, 5 (2010), 944-957.
22. Clemons, E.K., Gao, G.G., and Hitt, L.M. When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, 23, 2 (2006), 149-171.
23. Cones, J.W. *Film finance & distribution: a dictionary of terms*. Silman-James Press, 1992.
24. Cortes, C., and Vapnik, V. Support-vector networks. *Machine learning*, 20, 3 (1995), 273-297.
25. Cox, D.F. Risk handling in consumer behavior: An intensive study of two cases. *Risk taking and information handling in consumer behavior* (1967), 34-81.
26. Crano, W., and Prislin, R. Attitudes and persuasion. *Annual Review of Psychology*, 57 (2006), 345-374.
27. Crawford, J.L., and Haaland, G.A. Predecisional information seeking and subsequent conformity in the social influence process. *Journal of Personality and Social Psychology*, 23, 1 (1972), 112-119.
28. Cui, G., Lui, H.-K., and Guo, X. The Effect of Online Consumer Reviews on New Product Sales. *International Journal of Electronic Commerce*, 17, 1 (2012), 39-58.
29. Dellarocas, C., Zhang, X.M., and Awad, N.F. Exploring the value of online product reviews in forecasting sales: the case of motion pictures. *Journal of Interactive Marketing*, 21, 4 (2007), 23-45.
30. Dinner, I.M., Heerde, H.J.V., and Neslin, S.A. Driving Online and Offline Sales: The Cross-Channel Effects of Traditional, Online Display, and Paid Search Advertising. *Journal of Marketing Research*, 51, 5 (2014), 527-545.
31. Donnavieve, S., Satya, M., and Sivakumar, K. Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing*, 19, 3 (2005), 15-37.
32. Duan, W., Gu, B., and Whinston, A. Do online reviews matter?--An empirical investigation of panel data. *Decision Support Systems*, 45, 4 (2008), 1007-1016.
33. Elberse, A., and Eliashberg, J. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Science*, 22, 3 (2003), 329-354.
34. Eliashberg, J., and Shugan, S.M. Film critics: Influencers or predictors? *The Journal of Marketing* (1997), 68-78.

35. Eliashberg, J., and Shugan, S.M. Film critics: Influencers or predictors? *The Journal of Marketing*, 61, 2 (1997), 68-78.
36. Elsweiler, D., and Harvey, M. Engaging and maintaining a sense of being informed: Understanding the tasks motivating twitter search. *Journal of the Association for Information Science and Technology*, 66, 2 (2015), 264-281.
37. Fitzsimons, G.J., and Lehmann, D.R. Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Science*, 23, 1 (2004), 82-94.
38. Forman, C., Ghose, A., and Wiesenfeld, B. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, 19, 3 (2008), 291-313.
39. Ghose, A., Ipeirotis, P.G., and Li, B. Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, 31, 3 (2012), 493-520.
40. Ghose, A., and Todri, V. Towards a Digital Attribution Model: Measuring the Impact of Display Advertising on Online Consumer Behavior. *MIS Quarterly*, 40, 4 (2015), 1-20.
41. Godes, D., and Mayzlin, D. Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, 23, 4 (2004), 545-560.
42. Godes, D., and Silva, J.C. Sequential and temporal dynamics of online opinion. *Marketing Science*, 31, 3 (2012), 448-473.
43. Gopinath, S., Chintagunta, P.K., and Venkataraman, S. Blogs, Advertising, and Local-Market Movie Box Office Performance. *Management Science*, 59, 12 (2013), 2635-2654.
44. Gu, B., Konana, P., Raghunathan, R., and Chen, M. Research Note—The Allure of Homophily: Evidence from Investor Responses on Virtual Communities. *Information Systems Research*, 25, 3 (2014), 604-617.
45. Gu, B., Park, J., and Konana, P. Research Note—The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products. *Information Systems Research*, 23, 1 (2012), 182-196.
46. Hayashi, F. *Econometrics*. Princeton, NJ: Princeton University Press, 2000.
47. Hennig-Thurau, T., Marchand, A., and Hiller, B. The relationship between reviewer judgments and motion picture success: re-analysis and extension. *Journal of Cultural Economics*, 36, 3 (2012), 249-283.
48. Hennig-Thurau, T., Wiertz, C., and Feldhaus, F. Exploring the "Twitter Effect:" An Investigation of the Impact of Microblogging Word of Mouth on Consumers' Early Adoption of New Products. *Available at SSRN 2016548* (2012).
49. Huang, G.-H., and Korfiatis, N. Trying Before Buying: The Moderating Role of Online Reviews in Trial Attitude Formation Toward Mobile Applications. *International Journal of Electronic Commerce*, 19, 4 (2015), 77-111.
50. Huang, P., Lurie, N.H., and Mitra, S. Searching for experience on the web: an empirical examination of consumer behavior for search and experience goods. *Journal of Marketing*, 73, 2 (2009), 55-69.
51. Java, A., Song, X., Finin, T., and Tseng, B. Why we twitter: understanding microblogging usage and communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web Mining and Social Network Analysis*, ACM, California (2007), 56-65.
52. Kaplan, A.M., and Haenlein, M. The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, 54, 2 (2011), 105-113.
53. Karniouchina, E.V. Impact of star and movie buzz on motion picture distribution and box office revenue. *International Journal of Research in Marketing*, 28, 1 (2011), 62-74.
54. Kendall, J.E., and Kendall, K.E. Information delivery systems: an exploration of web pull and push technologies. *Communications of the AIS*, 1, 14 (1999), 1.
55. Kim, J.S., and Ratchford, B.T. Consumer Choice and Use of Multiple Information Sources for Automobile Purchases. *International Journal of Electronic Commerce*, 16, 3 (2012), 7-40.
56. Kohn, M.G., and Shavell, S. The theory of search. *Journal of Economic Theory*, 9, 2 (1974), 93-123.
57. Kuhlthau, C.C. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42, 5 (1991), 361–371.
58. Kulviwat, S., Guo, C., and Engchanil, N. Determinants of online information search: a critical review and assessment. *Internet Research*, 14, 3 (2004), 245-253.

59. Kwon, J., and Han, I. Information Diffusion with Content Crossover in Online Social Media: An Empirical Analysis of the Social Transmission Process in Twitter. *Journal of Management Information Systems*, 2013, pp. 3292-3301.
60. Lanzetta, J.T. Information acquisition in decision making. *Motivation and social interaction-cognitive determinants* (1963), 239-265.
61. Lee, C.H., and Cranage, D.A. Customer Uncertainty Dimensions and Online Information Search in the Context of Hotel Booking Channel. *Journal of Hospitality Marketing & Management*, 19, 19 (2010), 397-420.
62. Lee, G., and Raghu, T.S. Determinants of Mobile Apps' Success: Evidence from the App Store Market. *Journal of Management Information Systems*, 31, 2 (2014), 133-170.
63. Lee, K., Lee, B., and Oh, W. Thumbs Up, Sales Up? The Contingent Effect of Facebook Likes on Sales Performance in Social Commerce. *Journal of Management Information Systems*, 32, 4 (2015), 109-143.
64. Li, X., and Hitt, L. Self selection and information role of online product reviews. *Information Systems Research*, 19, 4 (2008), 456-474.
65. Li, X., Hitt, L.M., and Zhang, Z.J. Product Reviews and Competition in Markets for Repeat Purchase Products. *Journal of Management Information Systems*, 27, 4 (2011), 9-41.
66. Li, Z., and Li, C. Twitter as a social actor: How consumers evaluate brands differently on Twitter based on relationship norms. *Computers in Human Behavior*, 39 (2014), 187-196.
67. Liang, T.-P., Li, X., Yang, C.-T., and Wang, M. What in Consumer Reviews Affects the Sales of Mobile Apps: A Multifacet Sentiment Analysis Approach. *International Journal of Electronic Commerce*, 20, 2 (2016), 236-260.
68. Liu, C.Z., Au, Y.A., and Choi, H.S. Effects of Freemium Strategy in the Mobile App Market: An Empirical Study of Google Play. *Journal of Management Information Systems*, 31, 3 (2014), 326-354.
69. Liu, Y. Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70, 3 (2006), 74-89.
70. Luo, L., Chen, X., Han, J., and Park, C.W. Dilution and enhancement of celebrity brands through sequential movie releases. *Journal of Marketing Research*, *Forthcoming* (2010).
71. Ma, D. Push or Pull? A Website's Strategic Choice of Content Delivery Mechanism. *Journal of Management Information Systems*, 32, 1 (2015), 291-321.
72. Markopoulos, P.M., and Clemons, E.K. Reducing Buyers' Uncertainty About Taste-Related Product Attributes. *Journal of Management Information Systems*, 30, 2 (2013), 269-299.
73. Milliken, F.J. Three Types of Perceived Uncertainty About the Environment: State, Effect and Response Uncertainty. *Academy of Management Review*, 12, 1 (1987), 133-143.
74. Moon, S., Bergey, P.K., and Iacobucci, D. Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. *Journal of Marketing*, 74, 1 (2010), 108-121.
75. Moul, C.C. Measuring Word of Mouth's Impact on Theatrical Movie Admissions. *Journal of Economics & Management Strategy*, 16, 4 (2007), 859-892.
76. Mudambi, S.M., and Schuff, D. What makes a helpful online review? A study of customer reviews on Amazon. com. *Management Information Systems Quarterly*, 34, 1 (2010), 185-200.
77. Newey, W.K., and West, K.D. Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28, 3 (1987), 777-787.
78. Panaligan, R., and Chen, A. Quantifying movie magic with google search. *Google Whitepaper—Industry Perspectives+ User Insights* (2013).
79. Pornpitakpan, C. The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34, 2 (2004), 243-281.
80. Raiffa, H. Decision analysis: Introductory lectures on choices under uncertainty. *Reading*, Mass.: Addison-Wesley, 1968.
81. Reinstein, D.A., and Snyder, C.M. The influence of expert reviews on consumer demand for experience goods: a case study of movie critics. *The journal of industrial economics*, 53, 1 (2005), 27-51.
82. Sawhney, M.S., and Eliashberg, J. A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, 15, 2 (1996), 113-131.
83. Schlaifer, R.O. *Probability and statistics for business decisions*. New York: McGraw-Hill, 1959.

84. Schmidt, J.B., and Spreng, R.A. A proposed model of external consumer information search. *Journal of the Academy of Marketing Science*, 24, 3 (1996), 246-256.
85. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., and Williamson, R.C. Estimating the support of a high-dimensional distribution. *Neural computation*, 13, 7 (2001), 1443-1471.
86. Senecal, S., and Nantel, J. The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80, 2 (2004), 159-169.
87. Shiu, E.M.K., Gianfranco, W., Hassan, L.M., and Deirdre, S. Consumer uncertainty, revisited. *Psychology & Marketing*, 28, 6 (2011), 584-607.
88. Susarla, A., Oh, J.-H., and Tan, Y. Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube. *Information Systems Research*, 23, 1 (2012), 23-41.
89. Urbany, J.E., Dickson, P.R., and Wilkie, W.L. Buyer Uncertainty and Information Search. *Journal of Consumer Research*, 16, 2 (1989), 208-215.
90. Wang, A. Consensus and disagreement between online peer and expert recommendations. *International Journal of Internet Marketing and Advertising*, 4, 4 (2008), 328-349.
91. Wilson, E., and Sherrell, D. Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal of the Academy of Marketing Science*, 21, 2 (1993), 101-112.
92. Wilson, T.D. Information behaviour: An interdisciplinary perspective ☆. *Information Processing & Management*, 33, 4 (1997), 551-572.
93. Wu, T.-F., Lin, C.-J., and Weng, R.C. Probability estimates for multi-class classification by pairwise coupling. *The Journal of Machine Learning Research*, 5 (2004), 975-1005.
94. Xie, K., and Lee, Y.-J. Social Media and Brand Purchase: Quantifying the Effects of Exposures to Earned and Owned Social Media Activities in a Two-Stage Decision Making Model. *Journal of Management Information Systems*, 32, 2 (2015), 204-238.
95. Zhang, X., and Dellarocas, C. The Lord Of The Ratings: Is A Movie's Fate is Influenced by Reviews? *Proceedings of the International Conference on Information Systems* (2006).
96. Zhang, X., Fuehres, H., and Gloor, P.A. Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". *Procedia - Social and Behavioral Sciences*, 26 (2011), 55-62.
97. Zhang, Z., Ye, Q., Law, R., and Li, Y. The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29, 4 (2010), 694-700.
98. Zhu, F., and Zhang, X. Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74, 2 (2010), 133-148.

Figure 1. Research Model Overview

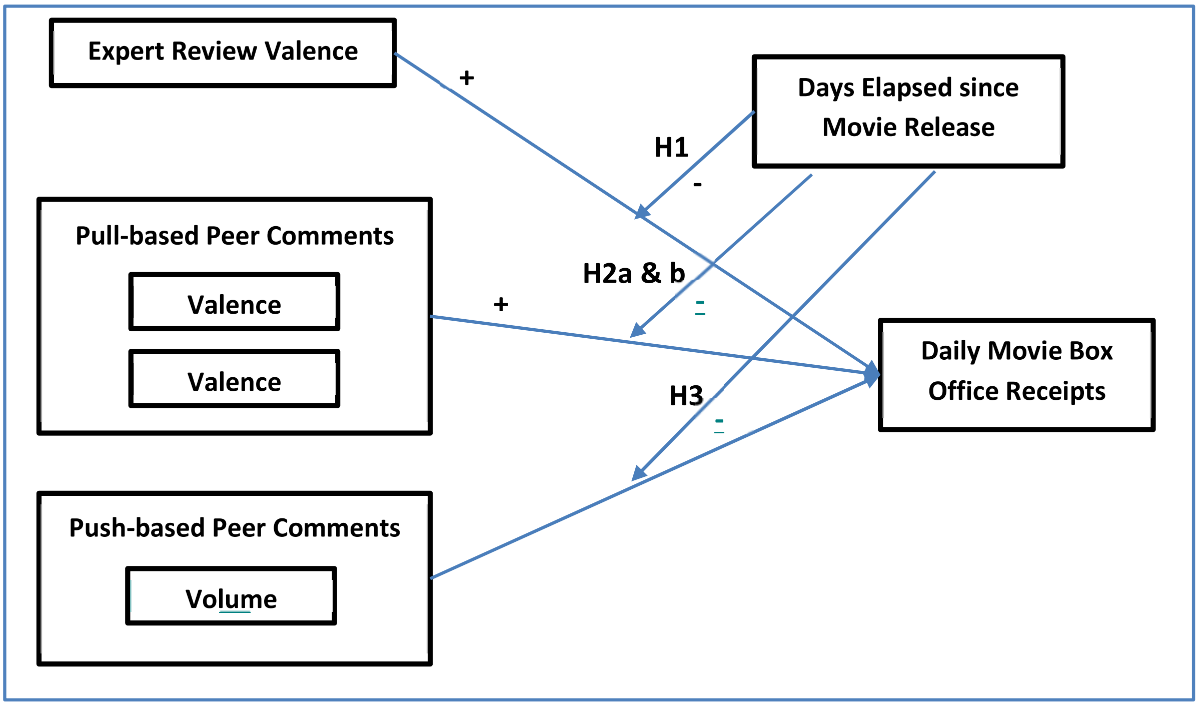

Figure 2. Interaction Plot of Online Comments Valence and No. of Days since Movie Release

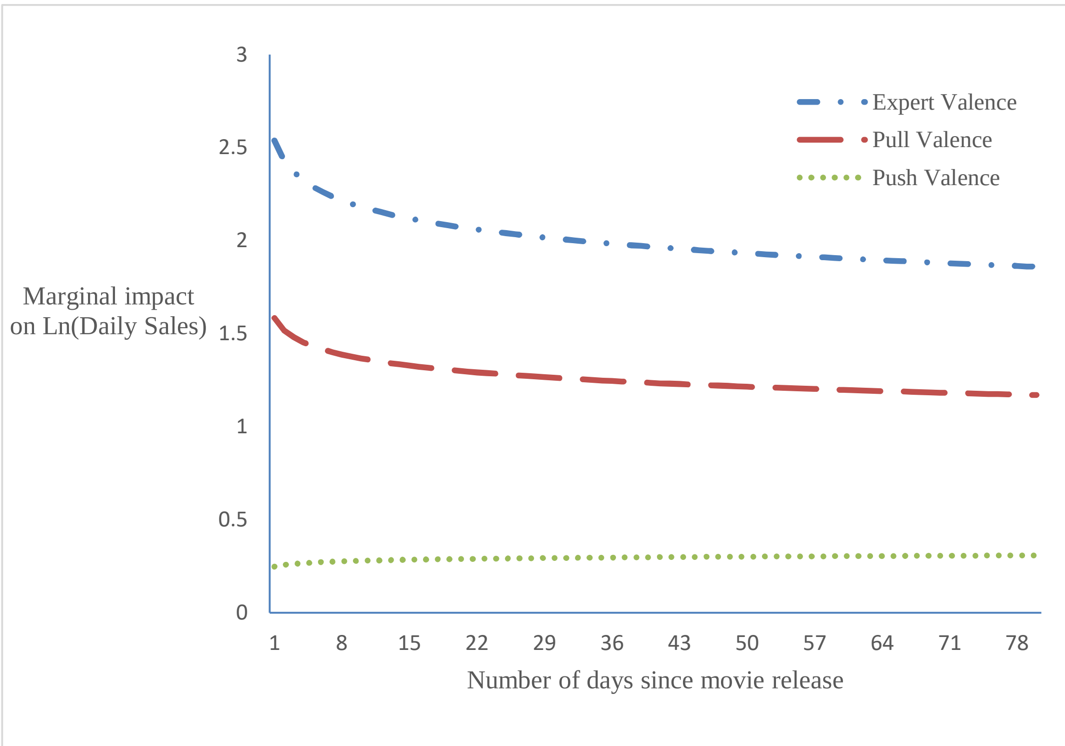

Figure 3. Interaction Plot of Online Comments Volume and No. of Days since Movie Release

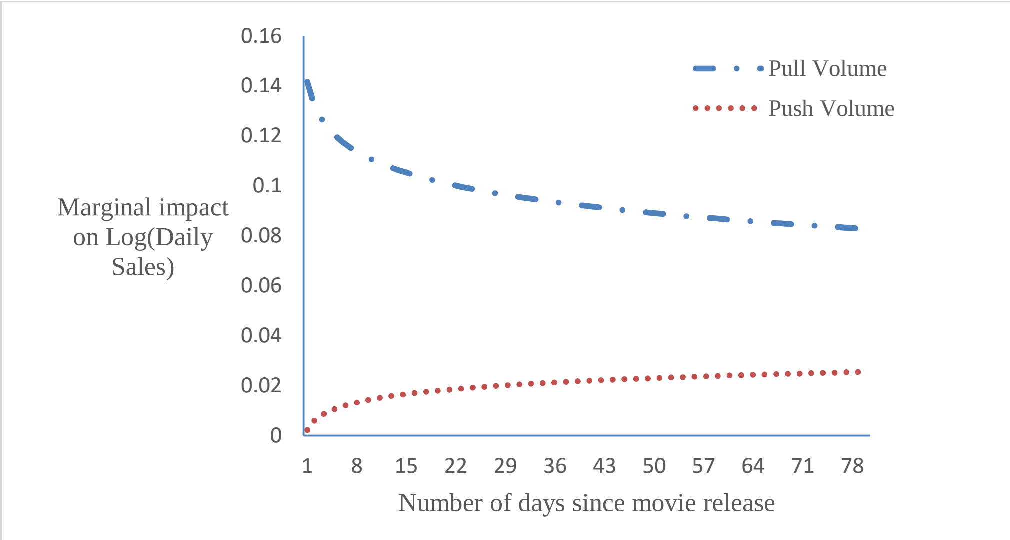

Figure 4. Summary of Results

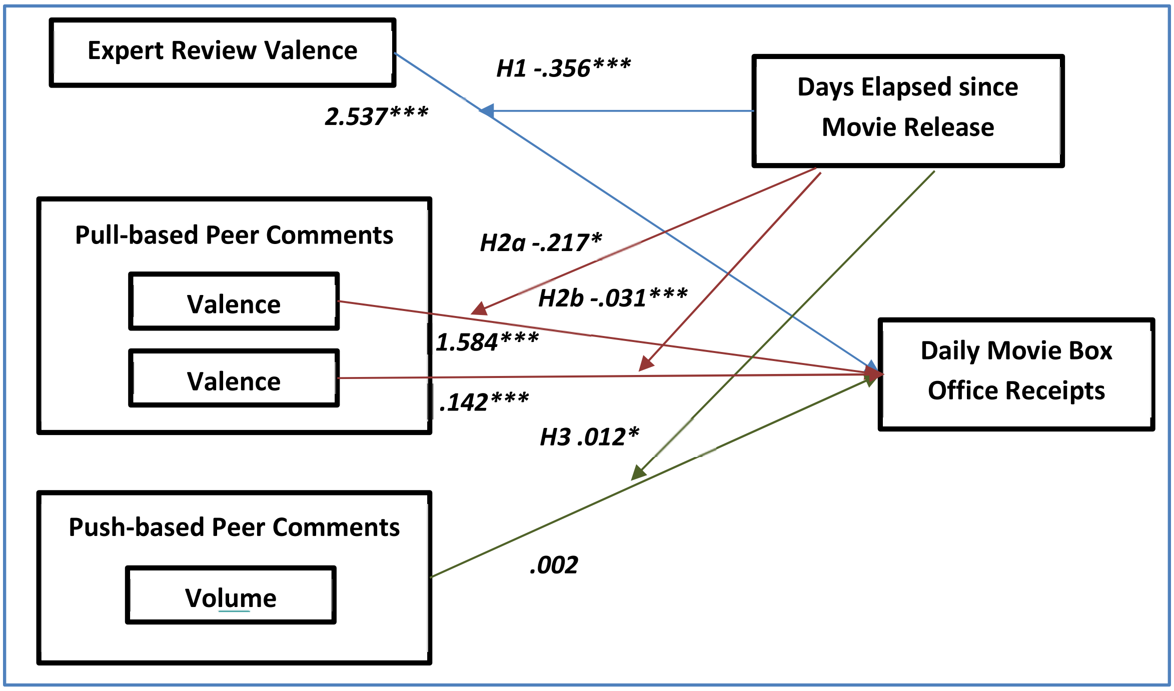

Table 1. Summary of Findings on Influence of Online Comments Valence and Volume on Movie Box Office Receipts Over Time

	Dependent Variable: Box Office Receipts Over Time	
	Early →	**Late**
Valence of Expert Reviews		
Eliashberg & Shugan (1997) [35]	Insignificant	+
Moon et al., (2010)[74]	+	Insignificant
Liu (2006)[69]	+	Insignificant
Zhang & Dellarocas (2006)[95]	+	Insignificant
Gopinath et al. (2013) [43]	+ (aggregate)	
Basuroy et al. (2003)[5]	+	+
Hennig-Thurau, Marchand, & Hiller (2012)[47]	+	Not tested
Valence of Pull-Based Comments (Forums)		
Liu (2006)[69]	Insignificant	Insignificant
Karniouchina (2011)[53]	Insignificant	Insignificant
Duan et al. (2008)[32]	Insignificant	Not tested
Moon et al. (2010)[74]	+	Not tested
Chintagunta et al. (2010)[21]	+ (aggregate)	
Gopinath et al. (2013)[43]	+ (aggregate)	
Zhang & Dellarocas (2006)[95]	+ (aggregate)	
Volume of Pull-Based Comments (Forums)		
Duan et al. (2008)[32]	+	Not tested
Liu (2006)[69]	+	Insignificant
Chintagunta et al. (2010)[21]	Insignificant (aggregate)	
Gopinath et al. (2013)[43]	Insignificant (aggregate)	
Zhang & Dellarocas (2006)[95]	Insignificant (aggregate)	
Valence of Push-Based Comments (Microblogs)		
Asur & Huberman (2010)[4]	+	Not tested
Hennig-Thurau, Wiertz, et al. (2012)[48]	+	Not tested
Volume of Push-Based Comments (Microblogs)		
Asur & Huberman (2010)[4]	+	Not tested
Hennig-Thurau, Wiertz, et al. (2012)[48]	-	Not tested

Note: Generally, early period is defined as around Weeks 1-4.

Table 2. Characteristics of the Three Types of Online Comments

Variable	Perceived ideological similarity	Search costs	Activation of passive attention	When they matter	How they matter
Expert reviews	Low	High	No	Early	Valence-persuade
Pull-based peer comments	High	High	No	Early	Valence-persuade Volume- inform
Push-based peer comments	High	Low	Yes	Late	Volume-inform

Table 3. Description of Variables used in Empirical Model

Variable	Description
Dependent Variable	
Daily Sales	Daily box office receipts from North America market (in US$)
Key Independent Variables	
Expert valence	The average sentiment of cumulative expert reviews
Pull valence	The average sentiment of pull-based peer comments for preceding 7 days #
Push valence	The average sentiment of push-based peer comments for preceding 7 days #
Pull volume	The number of pull-based peer comments for preceding 7 days #
Push volume	The number of push-based peer comments for preceding 7 days #
Controls	Note: Citations indicate the prior studies that have also included that control variable.
Days	Number of days since movie is released[69]
Budget	The budget of the movie (in US$)[5, 12, 29, 33, 69, 95]
Rank1	The popularity ranking of the leading artist [5, 12, 29, 33]
Rank2	The popularity ranking of the 2^{nd} leading artist [5, 12, 29, 33]
MPAA-PG	Movie is rated PG in MPAA ratings[5, 12, 29, 64, 82]
MPAA-R	Movie is rated R in MPAA ratings[5, 12, 29, 64, 82]
Drama	Movie genre is Drama[29, 69, 81, 82]
Thriller	Movie genre is Thriller[29, 69, 81, 82]

Note: # The results reported in this paper use a 7-day moving average of the valence and volume of the reviews. We also computed the 5, 6, 8, 9, 10, 11, 12, 13 and 14-day moving averages for the reviews' valence and volume as a robustness checks and found qualitatively similar results for all the estimators in used in this study.

Table 4.Descriptive Statistics and Correlation of Variables

Variables	Mean	Std. Dev.	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Ln (Daily Sales)	5.445	1.134													
2 Expert valence	0.737	0.173	0.17												
3 Pull valence	0.749	0.111	0.14	0.43											
4 Push valence	0.697	0.13	0.05	0.19	0.14										
5 Pull volume	1.684	2.44	0.41	-0.05	-0.04	-0.04									
6 Push volume	3.348	2.071	0.24	0.12	0.09	0	0.39								
7 Ln(Days)	3.433	1.021	-0.55	0.17	0.07	0.07	-0.56	-0.16							
8 Ln(Budget)	17.088	1.303	0.05	-0.1	-0.02	0.07	0.1	0.19	0.07						
9 Ln(Rank1)	4.829	1.932	-0.07	-0.1	0.01	-0.02	-0.04	-0.32	-0.11	-0.5					
10 Ln(Rank2)	5.674	1.807	-0.08	-0.1	0.02	-0.02	-0.06	-0.31	-0.12	-0.53	0.64				
11 MPAA-PG	0.141	0.348	-0.04	-0.09	-0.06	0.05	-0.02	-0.09	0.06	0.26	0.08	-0.06			
12 MPAA-R	0.363	0.481	0.02	0.1	0.17	-0.04	-0.05	-0.07	0	-0.3	0.13	0.14	-0.31		
13 Drama	0.513	0.5	0.01	0.24	0.12	0.11	-0.08	-0.01	-0.04	-0.41	0.09	0.12	-0.29	0.1	
14 Thriller	0.278	0.448	-0.04	0.03	-0.07	-0.09	0.05	0.21	-0.01	0.15	-0.17	-0.12	-0.21	-0.04	-0.12

Note: All correlation values greater than 0.02 are also statistically significant at *p*-value<0.05

Table 5. Impact of Social Media Comments Across Different Platforms on Movie Sales

DV: Ln(Daily Sales)	Random Effects		Fixed Effects (no endogeneity correction)		IV-GMM (I) (endogeneity correction)		IV-GMM (II) (endogeneity + heteroskedasticity)		IV-GMM (III) (endogeneity + heteroskedasticity + autocorrelation)	
	Coeff.	**S.E.**	**Coeff.**	**S.E.**	**Coeff.**	**S.E.**	**Coeff.**	**S.E.**	**Coeff.**	**S.E.**
Expert valence	2.279***	0.235	1.300***	0.340	2.549***	0.153	2.538***	0.192	2.537***	0.245
Pull valence	0.617**	0.233	0.480*	0.226	1.595***	0.242	1.586***	0.289	1.584***	0.365
Push valence	-0.275	0.201	-0.621**	0.194	0.239	0.214	0.249	0.229	0.248	0.293
Pull volume	0.040***	0.012	0.039***	0.012	0.142***	0.014	0.142***	0.014	0.142***	0.018
Push volume	0.020	0.011	0.021	0.011	0.002	0.013	0.002	0.014	0.002	0.018
Ln(Days)	-0.807***	0.008	-0.811***	0.008	-0.573***	0.009	-0.573***	0.010	-0.572***	0.013
Expert valence X Ln(Days)	-0.385***	0.040	-0.379***	0.038	-0.358***	0.044	-0.356***	0.054	-0.356***	0.068
Pull valence X Ln(Days)	-0.159*	0.067	-0.230***	0.063	-0.219**	0.070	-0.218**	0.081	-0.217*	0.103
Push valence X Ln(Days)	0.110*	0.053	0.205***	0.050	0.034	0.059	0.031	0.062	0.032	0.079
Pull volume X Ln(Days)	-0.007*	0.004	-0.007	0.004	-0.031***	0.004	-0.031***	0.004	-0.031***	0.005
Push volume X Ln(Days)	-0.009*	0.003	-0.009**	0.003	0.012***	0.004	0.012**	0.004	0.012*	0.005
Ln(Budget)	0.036	0.043								
Ln(Rank1)	-0.020	0.029								
Ln(Rank2)	-0.102**	0.032								
MPAA-PG	-0.072	0.148								
MPAA-R	0.025	0.095								
Drama	-0.050	0.097								
Thriller	-0.203*	0.100								
Intercept	8.187***	0.861	8.230***	0.030	7.386***	0.033	7.384***	0.039	7.383***	0.050
R^2		0.409		0.303		0.404		0.404		0.404

Note: ***represents p-value<0.001; ** represents p-value<0.01; * represents p-value<0.05. All F-statistics are significant at p-value<0.001.Random Effects (RE) and Fixed Effects (FE) estimations presented for comparison purposes. RE inconsistent based on Hausman test. **n**=15,665