Pre-Launch Advertising, Online Buzz and New-Product Sales

Ho Kim, Dominique M. Hanssens[*]

Anderson School of Management

University of California, Los Angeles

September 22, 2011

[*] Ho Kim is a PhD student in Marketing, UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, CA 90095 (Email: ho.kim.2013@anderson.ucla.edu, Phone: (1) 310-825-7873, Fax: (1) 310-206-7422). Dominique M. Hanssens is the Bud Knapp Professor of Marketing, UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, CA 90095 (Email: dominique.hanssens@anderson.ucla.edu, Phone: (1) 310-825-4497, Fax: (1) 310-206-7422). The authors acknowledge the helpful comments of Randy Bucklin, Xavier Dreze, and Suzanne Shu on a previous version of this paper.

Pre-Launch Advertising, Online Buzz and New-Product Sales

ABSTRACT

This study examines the effect of pre-launch new-product advertising on sales. Since sales are not observed in the pre-launch period, we propose online buzz as a metric that links the pre-launch advertising to sales. We adopt a two-stage approach: the first examines the time-varying effectiveness of pre-launch advertising on buzz; the second examines the influence of pre-launch buzz on sales. We apply our models to a movie dataset that consists of weekly advertising spending, blog postings and keyword search volume, movie characteristics, and box-office revenues. We find that (1) the pre-launch phase of a new movie starts with a consumer response regime that is subject to information discounting, and then switches to a non-discounting regime, (2) the effect of advertising deteriorates with the elapsed time since the first advertising, (3) previous advertising spending boosts the effectiveness of current advertising, (4) the effect of early advertising does not carry over to the subsequent periods, (5) pre-launch online buzz influences subsequent box-office revenues. Based on the findings, we illustrate an effective allocation of advertising budget in the pre-launch period.

Keywords: Pre-launch Advertising, Online Buzz, Discounting Behavior, Copy Wearout, Repetition Wearin.

INTRODUCTION

As the life cycles of new products get shorter and a good portion of their sales are realized in the first few weeks or months after release, the challenge of lifting initial-period sales is under a greater attention than ever. Products such as movies, books, game CDs/DVDs, and some consumer electronics are good examples. In these industries, peak sales are usually achieved in the first few weeks of release, which implies that pre-launch advertising is important. Realizing this, managers more often spend hefty amounts of pre-launch advertising without carefully measuring its effects on consumers. Such measurement is difficult because conventional performance metrics such as sales and profits do not exist pre-launch. New metrics are needed that are valid proxies or leading indicators of post-launch business performance.

The objective of this study is to understand the effect of pre-launch new-product advertising on sales by using readily available online buzz data. Specifically, we try to answer the following questions: How effective is the advertising that is run at different points in time during the pre-launch period, and what are the factors that drive the time-varying effectiveness of such pre-launch advertising? Is pre-launch buzz an influencer of new-product sales or merely a leading indicator of it? We develop several hypotheses on the effects of pre-launch advertising on buzz generation and test them by applying a model on movie releases in the United States.

The empirical results reveal the drivers of the time-varying effectiveness of pre-launch advertising. First, the pre-launch phase of a new movie starts with a consumer response regime that is subject to information discounting, and then switches to a non-discounting regime around 1.5 weeks before release. The effect of advertising conducted in the discounting regime is substantially smaller than that conducted in the non-discounting regime. Second, after controlling for consumers' time discounting behavior, we find that the effect of advertising deteriorates with the elapsed time since the first advertising – i.e. a copy *wearout* effect. Third, we find that previous advertising spending boosts the effectiveness of current advertising in generating online buzz – i.e. a repetition *wearin* effect. Fourth, the effect of early advertising, which is separated in time from the launch advertising, does not carry over to the subsequent periods, which limits its utility. Fifth, pre-launch online buzz influences subsequent box-office movie revenues.

In addition to generating the new findings above, the data in this paper are distinctive from previous studies. We show that the keyword search volume and blog postings are potentially informative pre-launch variables. By contrast, most previous studies use conversations from Usenet news group (Godes and Mayzlin 2004) or user reviews from Yahoo! Movie website (Chintagunta, Gopinath, and Venkataraman 2010; Liu 2006). User reviews or ratings restrict the pre-launch time span because they are posted shortly before or after product launch. Blog postings and keyword search volumes allow us to investigate time-varying advertising effects over a much longer pre-launch period.

This study treats only the consumer side of advertising effects, leaving its effect on supply side untapped. As Ho, Dhar, and Weinberg (2009) suggest, the early major advertising (e.g. Super Bowl advertising) may give signal to distributors so that the distributors feel they should allocate more screens for those movies. We do not explicitly treat those supply side effects in this study.

The rest of the paper is organized as follows: We review relevant research and develop hypotheses on the effects of pre-launch advertising on new-product revenues. We then introduce our data and describe pre-launch advertising practices and online buzz patterns in the US movie industry. We develop and estimate response models to investigate the effect of pre-launch advertising on buzz generation. In a separate section we provide a separate analysis to assess the influencer role of pre-launch buzz on consumers' movie consumption. Managerial implications on advertising resource allocation are provided in the subsequent section. We then formulate conclusion and areas for future research.

THEORY AND HYPOTHESIS DEVELOPMENT

During pre-launch period, conventional performance metrics such as sales are not observed, making it difficult to measure advertising effects. We propose that the volume of online buzz during the pre-launch period is a good proxy metric of new-product sales, and develop four hypotheses regarding the effect of advertising on pre-launch online buzz.

The Effect of Advertising on Online Buzz Generation

It is well known that advertising increases awareness, purchase intention, and ultimately sales. Naik, Mantrala, and Sawyer (1998) apply a dynamic advertising quality model to surveyed awareness data that vary according to the level of TV advertising. Their data show that the level of awareness increases while advertising is run and it decreases when advertising is ceased. This establishes two phenomena from real-world data: (1) Advertising increases awareness of the advertised product, (2) the awareness is mean-reverting or stationary over time. Draganska and Klapper (2010) examine the dual role of advertising – influencing awareness and influence preference. They use survey data from a tracking study on brand awareness along with data on advertising expenditure in the ground coffee category in Germany to show that advertising has a direct effect on brand awareness (inclusion in choice set) as well as an indirect effect on consumer preference (increase in utility). Zufryden (1996) develops a funnel model that links advertising spending to awareness, purchase intention, and finally purchase. He applies the model to movie box-office revenue data and shows that (1) movie advertising increases both awareness and purchase intention, and (2) the actual purchase is indirectly and positively influenced by the advertising through awareness and purchase intention. Several other studies verify that advertising generally increases awareness, purchase intention, and sales.

The current study uses consumers' blogging and search activities as measures of online buzz. These activities imply that they are at least aware of and are interested in the product. In other words, blog postings and keyword searching activities are behavioral consequences of awareness of and interest in a product. Since advertising increases awareness and purchase intention, it is likely that advertising also increases online buzz. This behavioral consequence should occur no matter when consumers are exposed to the advertising – whether before or after the product launch. We therefore hypothesize:

H1a: Advertising during the pre-launch period of a new product increases online buzz, as measured by the number of blog postings and the keyword search volume in major search engines.

Once consumers see an advertisement for a new product, they may respond immediately by posting their opinions on their blogs or by searching the new product on the Internet, or they may respond later on, for example due to lack of time or forgetting. One week later, they may see other blog postings about the product and engage in a search. Alternatively, they may visit the hot-issue section of a web portal and find the product ranks among the top search keywords, prompting them to blog or search. These and other scenarios illustrate that the advertising effect on buzz generation can carry over to future periods. However, the effect is not permanent in the sense that, once the advertising is ceased, online buzz reverts to its pre-advertising level. Therefore, we offer the following hypothesis:

H1b: The effect of advertising in one pre-launch period carries over to future periods. The effect of advertising is not permanent in that online buzz level eventually reverts to the pre-advertising level once the advertising is ceased.

Time-Varying Effects of Pre-Launch Advertising

Little (1979) summarizes five advertising response phenomena, one of which is that advertising effectiveness changes over time. Based on previous literature, we identify two distinct sources of time-varying effectiveness of pre-launch advertising: time-discounting behavior of consumers and wearin/wearout effects of repeated advertising.

Time-discounting behavior. Intertemporal decision theory supports the argument that time-to-launch impacts advertising effectiveness during the pre-launch period. According to Trope and Lieberman (2003), one common theme in intertemporal decision theory is that individuals often place higher value on a near-future reward than on a distant-future reward, even when the distant-future reward is larger (Ainslie and Haslam 1992; Read and Loewenstein 2000). In our context, this implies that consumers generate more buzz (i.e. blog postings and keyword searching) for movies that are released sooner than later because (1) the value of the movies that are released sooner looms larger than that of the later-released movies and (2) in general, greater value means greater interest, which is related to more word-of-mouth (WOM).

The time-to-launch effect is also supported by dual processing theory (Petty and Cacioppo 1979). According to this theory, consumers' involvement with a product or purchasing decision is a function of the personal relevance or the importance of that product or decision. Because movies that are released sooner are more relevant to consumers' decisions, consumers are more motivated to process the advertising of these movies, which increases the likelihood of generating buzz. Since the primary source of buzz is advertising, this implies that the effectiveness of advertising is greater as the movie release is approaching. Therefore, we hypothesize:

H2a: In the pre-launch period, the time discounting behavior of consumers influences the effectiveness of advertising on online buzz generation, in that the same amount of advertising spending generates greater buzz as the new-product release date is approaching.

Wearin vs. wearout effect. Previous empirical studies on established products showed that advertising impact changes over time because of wearin and wearout effects, and consumer forgetting (Bass et al. 2007; Bruce 2008; Naik, Mantrala, and Sawyer 1998). Based on these studies, we hypothesize equivalent time-varying effects in pre-launch advertising due to wearin/wearout, even after controlling for the time-discounting behavior of consumers.

The wearin effect means that previous advertising efforts help current advertising. It is supported by the theory of repeated stimuli exposures. Sawyer (1981) argues that repeated advertising can influence recall, attitude, and purchase intention. Janiszewski (1993) shows that mere exposure to a brand name or product package can induce consumers to have a more favorable attitude toward the brand. Shapiro, MacInnis, and Heckler (1993) find that priming an advertising can increase the likelihood that the product depicted in the advertising will be included in the consideration set even when consumers are unaware of the advertising or when consumers have at least very limited cognitive resource to process the advertising. On the empirical side, Bass et al. (2007) show repetition wearin, i.e. advertising effectiveness increases in advertising spending. These studies imply that previous pre-launch advertising for a new product may enhance the effectiveness of later advertising, which we term *repetition wearin effect*.

There is, however, an opposite force of repeated advertising, which is wearout. Wearout refers to a decline in the effectiveness of an advertisement (Grass and Wallace 1969; Greenberg and Suttoni 1973) because consumers have already been exposed to previous advertising and thus they find nothing new in the additional advertising, become bored or even irritated. When wearout is at work, the effectiveness of advertising declines as advertising spending is maintained (repetition wearout) and/or more time has passed since the first advertising (copy wearout). Many empirical studies (Bruce 2008; Naik, Mantrala, and Sawyer 1998) find such wearout effects. The two forces (wearin and wearout) are opposite in direction, and the composite effect depends on their relative magnitude. Therefore, we hypothesize:

H2b: The impact of current advertising on buzz generation can be either increasing or decreasing in past advertising efforts and the time since the first advertising. If the wearin effect of advertising is stronger than the wearout effect then current advertising's effectiveness is increased. Conversely, if the wearout effect is stronger, current effectiveness is decreased.

Pre-launch Buzz as a Predictor and Influencer of New-Product Sales

Previous empirical studies have found that pre-launch online buzz predicts movie box-office revenue (Karniouchina 2011; Liu 2006). We propose that pre-launch online buzz not only predicts but also influences new-product sales through herd behavior or information cascades. Consumers who read a blog about a new product or who observe a new product is a highly ranked search keyword may think the new product is a current hot subject among consumers. In general, the more blogs are posted about a new product, the more likely a consumer will be exposed to it. Similarly, the more a new product is searched for by other consumers, the more likely a consumer will become aware of the product (e.g. by clicking a keyword in the top search list in an internet portal). In these circumstances, the theory of herd behavior (Banerjee 1992) and information cascades (Bikhchandani, Hirshleifer, and Welch 1992) suggest that searching and blogging behavior in the pre-launch period can *influence* the new-product adoption in the post-launch period.

If pre-launch buzz is an influencer on the new-product adoption, the association between pre-launch buzz and new-product sales should be greatest in the opening week and then gradually decrease over the post-launch period. This is because in the early weeks of new-product launch consumers have little information about the new product and thus they are more likely to follow the wisdom of peer consumers accumulated in the pre-launch period. As time goes by, however, the relative impact of pre-launch buzz on new-product sales should diminish because other factors (e.g. previous adopters' reviews and ratings) dominate the pre-launch buzz. Therefore, we postulate as follows:

H3: Pre-launch buzz around a new product is not merely a predictor of its post-launch sales, it is an influencer of consumers' consumption. Pre-launch buzz is the most influential in the opening week and become less influential over time in the post-launch period.

Figure 1 summarizes our hypotheses and shows the relationship between the critical components of our model.

== Figure 1 about here ==

DATA

The U.S. movie industry is a good test case or our hypotheses because (1) most movie releases are by nature new products and (2) significant amounts are spent for pre-launch advertising. Our database contains 180 titles that were widely released in 2008 and 2009. We excluded two movies - *Food, Inc* and *Aaj Kal Love,* because they did not advertise before their release. We excluded an additional twenty movies whose pre-launch search volume index by Google was zero, resulting in a final dataset of 158 movies. Our search volume measure is calculated from the search volume indices of Google Insights and Google Trends. As such, a zero search volume measure of a movie does not necessarily mean that no one searched for the movie. Instead it implies that the movie lacked sufficient popularity to generate a non-zero search volume index defined by Google. Each movie is tracked from sixty weeks prior to release to five weeks after release. The main variables include weekly advertising spending, weekly blog postings, and weekly search volume measure.

Advertising

The main advertising for a movie - which we label *launch advertising* - typically starts several weeks before its release and runs without interruption until some point in the post-launch period. The weekly launch advertising spending tends to increase exponentially toward the release week and decreases rapidly thereafter. Launch advertising is the main source of information for movie viewers and is used by the major studios for all their releases. For several movies, however, this launch advertising is preceded by another type of advertising which we call *early advertising*. *Early advertising* starts much earlier, sometimes one year before the movie release, and its weekly spending schedule is either continuous or intermittent[1)]. If a movie has both early and launch, the early advertising is followed by one or more weeks of silence before the launch advertising starts.

Our advertising data covers both early and launch advertising in the major media, i.e. television, print, radio, and outdoor expenditure as collected by Nielsen. The average advertising spending of the 158 movies is $21.2 million with 53% of the advertising budget spent in the pre-launch periods. The advertising schedules vary greatly by movie. Among the 158 movies, 91 adopted early advertising, with the majority (63%) having only one run. One movie *The Soloist* did as many as five runs of early advertising before release. On average, the first early advertising starts 17 weeks before release, but the movie *Avatar* started advertising at least 123 weeks before release. Launch advertising starts, on average, six weeks before release, though the movie *A Christmas Carol* started twenty weeks before release. The average length of the silence period between the early advertising and the launch advertising is 4.3 weeks. In terms of budget allocation, the average pre-launch spending share of early advertising is 6.1%.

Blog Postings and Search Volume

The weekly blog postings for each movie in the US were collected from Google Blogs. To minimize noise in the data collection process, we constrained our search for postings to blogs with titles that contain at least one of the following words: *movies*, *films*, or *flicks*. Our general rule of constructing search keywords for blog postings is <movie title> + "movie"[2)]. For example, to find blog postings of the movie Avatar, we used the keyword "Avatar movie". For some movies with very long titles, reduced search keywords were used. For instance, to search for blog postings of the movie *Bad Lieutenant: Port of Call New Orleans*, we searched for the postings that contain the keyword 'Bad Lieutenant' in their title. For each week of each movie, we repeated the same search five times and used the mode or median of the number of blog postings so gathered.

For weekly search volume of movies, we relied on two publicly available services provided by Google: Google Trends (http://www.google.com/trends) and Google Insights (http://www.google.com/insights/search/#). *Google Trends* shows the weekly search volume index of the entered keyword. Like blog postings case, we used <movie title> + “movie” for the search keyword. For some movies whose titles are very long or for movies with subtitles, we modified the general rule to make sure that the collected search volume is a valid one in terms of the fit with the movie’s observed advertising and launch schedules.

Google Insights is similar to Google Trends except that it provides an industry category filter with which one can specify the industry from which the keyword search volume is measured. This is particularly useful for time-series analyses of movie search volumes, because the data collection noise due to irrelevant search is reduced. In our case, we restrict our search volume to the movie industry.

The index provided by Google is normalized and scaled to conceal the actual search volume of the keyword in the Google search engine. Both services normalize the search volume of a keyword by dividing by a common variable. However, the two services use different scaling methods. Google Trends scales the search volume so that the average search traffic of a keyword during the search period is one. Then the data are scaled to the average search traffic (represented as 1.0) during the time period chosen by the user. In Google Insights, the data are displayed on a scale of 0 to 100. Each index is divided by the highest point, or 100.

This normalization and scaling create a difficulty for analyzing the effect of search volume across movies. While the search volume index from Google Insights contains less noise due to the industry filter, it is relative to the peak volume within the movie, which makes a cross-section analysis impossible. The search volume index from Google Trends is also relative within a movie but there is a way of comparing the volume across different movies by use of a common basis keyword (Joshi and Trusov 2010). The search volume index of the basis keyword should be stable over time and at the same time it should not be too large overall, because, otherwise, the search volume indices of our movies will be masked as zero. With this in mind, we chose as the basis keyword 'Teri Hatcher', which is the name of a US-born actress who draws a medium and stable search volume compared to other keywords.

Constructing a Valid Search Volume Measure

From the search volume indices from Google, we develop a valid search volume *measure* so that we can do both time-series analysis within a movie and cross-section analysis across movies. The valid measure is constructed by matching the peak index from Google Insights (i.e. 100) with its counterpart from Google Trends. For example, assume the index from Google Trends is 8.5 in the week when Google Insights has a peak value of 100. Then, for a week where the Google Insight index is 10 our final measure of search traffic is 0.85(=8.5/100*10). This process assures that the time-series of a keyword search volume follows the Google Insights index, while the absolute level of the search volume is moderated by the index from Google Trends. This approach enables us to do both time-series analyses within a movie and compare the effects of search volume across movies.

Figure A1 in Appendix A exhibits the weekly advertising schedule and online buzz process constructed in the proposed way for the movies *Avatar*, *Transformers: Revenge of the Fallen*, and *Harry Potter and the Half-Blooded Prince*. Figure 2 shows the average weekly advertising spending and average volume of buzz of the 158 movies from 60 weeks before release to 5 weeks after release.

== Figure 2 about here ==

RESPONSE MODELS

The Effect of Advertising on Buzz Generation

Let A_{it} be the advertising spending and y_{it} be the volume of online buzz of movie i observed at time t. Then y_{it} may be decomposed in four components: online buzz generated by current and past early advertising ($\tilde{y}_{1it}$); online buzz generated by current and past launch advertising ($\tilde{y}_{2it}$), online buzz generated by other sources (u_{it}), and the time-invariant baseline of online buzz (κ_i), which is a function of idiosyncratic factors of the movie. We assume that u_{it} follows a white noise process. If the process u_{it} does not systematically differ across movies with different advertising schedules, then treating it as an unobserved random deviation would not cause any systematic bias in estimating advertising effectiveness. Therefore,

(1) $$y_{it} = \kappa_i + \tilde{y}_{1it} + \tilde{y}_{2it} + u_{it} .$$

Let A_{it}^P and A_{it}^M denote the amount of early advertising and that of launch advertising of movie i at time t, respectively. Then $\tilde{y}_{1it}$ and $\tilde{y}_{2it}$ are represented as follows:

(1-1)
$$\tilde{y}_{1it} = \beta_{i0} A_{it}^{P} + \beta_{i1} A_{i,t-1}^{P} + \cdots = \sum_{k=0}^{\infty} \beta_{i,t-k} A_{i,t-k}^{P}$$
$$\tilde{y}_{2it} = \gamma_{i0} A_{it}^{M} + \gamma_{i1} A_{i,t-1}^{M} + \cdots = \sum_{k=0}^{\infty} \gamma_{i,t-k} A_{i,t-k}^{M}$$

, where $\beta_{i,t-k}$ and $\gamma_{i,t-k}$ measure the effect of the early and launch advertising that have occurred *k* periods before. If we assume a geometric decay of the advertising effects, then eq. (1-1) is simplified as:

(1-2)
$$\tilde{y}_{1it} = \beta_{it} \sum_{k=0}^{\infty} \lambda_i^k A_{i,t-k}^{P} = \beta_{it} A_{it}^{P} + \lambda_i \, \tilde{y}_{1i,t-1}$$
$$\tilde{y}_{2it} = \gamma_{it} \sum_{k=0}^{\infty} \delta_i^k A_{i,t-k}^{M} = \gamma_{it} A_{it}^{M} + \delta_i \, \tilde{y}_{2i,t-1}$$

, where λ_i and δ_i represent the carryover rates of the early and launch advertising, respectively[3)].

In our data, we observe three different periods: the period before the first advertising starts; the period of early advertising (if applicable), and the period of launch advertising. By our definition of advertising types, early advertising occurs only in the pre-launch period, whereas launch advertising can be done both pre- and post-launch. Let y_{0it} be the observed online buzz level in the period before the first advertising starts, y_{i1t} be the level during early advertising period and y_{2it} be the level during the launch advertising period. Then, for movies with early advertising, it holds that

(2-1)
$$\begin{aligned} y_{0it} &= \kappa_i + u_{it} \\ y_{1it} &= \kappa_i + \tilde{y}_{1it} + u_{it} = \kappa_i (1-\lambda_i) + \beta_{it} A_{i,t}^{P} + \lambda_i y_{1,it-1} + \nu_{1it}, \\ y_{2it} &= \kappa_i + \tilde{y}_{1it} + \tilde{y}_{2it} + u_{it} = \kappa_i (1-\delta_i)(1-\lambda_i) + \gamma_{it} (A_{i,t}^{M} - \lambda_i A_{i,t-1}^{M}) + (\lambda_i + \delta_i) y_{2,i,t-1} - \lambda_i \delta_i y_{2,i,t-2} + \nu_{2it} \end{aligned}$$

, where $\nu_{1it} = u_{it} - \lambda_i u_{i,t-1}$ and $\nu_{2it} = u_{it} - (\lambda_i + \delta_i) u_{i,t-1} + \lambda_i \delta_i u_{i,t-2}$.

For movies without early advertising, we do not observe y_{i1t} and (2-1) is replaced by (2-2):

(2-2)
$$y_{0it} = \kappa_i + u_{it},$$
$$y_{2it} = \kappa_i + \tilde{y}_{2it} + u_{it} = \kappa_i(1-\delta_i) + \gamma_{it} A_{it}^{M} + \delta_i y_{2,i,t-1} + \nu_{3it}.$$

, where $\nu_{3it} = u_{it} - \delta_i u_{i,t-1}$.

Time-Varying Effectiveness of Advertising

We have previously discussed the two sources of time variation in the effectiveness of new-product advertising: the time-discounting behavior of consumers (Ainslie and Haslam 1992) and the wearin/wearout effect of repeated advertising (Naik, Mantrala, and Sawyer 1998). We incorporate these two sources in a multiplicative form: advertising effectiveness at time t = (time discounting factor at t) x (wearin/wearout effect of advertising at t), or Q_t=f(t)×q_t, where f(t) has a value between zero and one.

We address two important issues regarding consumer discounting behavior. First, there may be a point in time τ where the movie launch is sufficiently close that consumer discounting of the advertising no longer occurs. Then total advertising impact Q_t may be expressed as:

(3)
$$Q_t = \begin{cases} f(t,\tau) \times q_t, & \text{if } t \le \tau \\ q_t, & \text{if } t > \tau. \end{cases}$$

Second, some movies may be more vulnerable to time-discounting behavior than others. For example, consumers may be more responsive to an advertisement for a sequel movie because their expectation is likely to be higher (since only successful movies typically generate sequels). Thus, we allow f(t) to vary according to movie characteristics **x**.

We adopt a hyperbolic function as the specific functional form for f(t) (Soman et al. 2005). This is based on the assumption that movie viewers discount the utility of the future event (i.e. movie release) in a time-inconsistent manner during the pre-launch period (Ainslie and Haslam 1992). If we adopt a hyperbolic function, the time discounting factor is represented as $f(t) = \left(1 + \left(\varphi_0 + \boldsymbol{\varphi_1 x}\right) | t - \tau |\right)^{-1}$, where $|t-\tau|$ is the distance to the normal regime in week t[4] and $\boldsymbol{\varphi_1}$ captures the moderating effects of movie characteristics.

Another source of time-variation is wearin/wearout. To understand how to incorporate these effects, let us start with a continuous time model where q(t) is the advertising effectiveness at time t, net of the discounting factor, and AD(t) is the advertising spending at time t. If we assume the advertising effectiveness changes linearly in advertising spending (Pekelman and Sethi 1978), then it holds that dq(t)/dt=c+w·AD(t). The parameters c and w are interpreted as copy wearout (wearin) and repetition wearout (wearin), if they are negative (positive), respectively (Bass et al. 2007; Naik, Mantrala, and Sawyer 1998). Transforming the continuous time model to its corresponding discrete time model leads to the following specification: $q_t - q_{t-1} = c + w \cdot AD_{t-1}$ or $q_t = q_{t-1} + c + w \cdot AD_{t-1} = ct + w\sum_{l=1}^{t-1} AD_l$ with appropriate initial conditions. By augmenting the two sources, the effectiveness of advertising at time t is represented by :

$$\text{(3-1)} \qquad Q_t = \begin{cases} \left(1 + \left(\varphi_0 + \boldsymbol{\varphi_1 x}\right) | t - \tau |\right)^{-1} (c_0 + c_1(t - t_0) + w\sum_{l=\tau_0}^{t-1} AD_l), & \text{if } t \le \tau \\ (c_0 + c_1(t - t_0) + w\sum_{l=\tau_0}^{t-1} AD_l), & \text{if } t > \tau \end{cases}$$

, where t_0 is the first week of advertising. Figure 3(a) exhibits the effect of φ_0 on the discount factor f(t) with three different values: 0, 1, and 4. To avoid clutter, we set $\boldsymbol{\varphi_1}=\mathbf{0}$ and τ=0. Figure 3(b) exhibits how the two opposite forces – wearin vs. wearout – moderate the advertising effectiveness (q_t) over the course of the pre-launch period with various combinations of $c_0=3\times10^5$, $c_1=\pm10^3$, $c_2=\pm1$.

== Figure 3 about here ==

The general formula (3-1) is tailored for early and launch advertising. The effectiveness of movie i's early advertising in week t, β_{it} is derived by substituting movie i's characteristics and advertising schedule in eq. (3-1):

$$\text{(3-2)} \qquad \beta_{it} = \left(f_i(t,\tau)\cdot I(t\le\tau) + I(t>\tau) \right)\cdot\left(\psi_{0i}^P + \psi_1^P \cdot \max(t - t_{i0}^P, 0) + \psi_2^P \sum\nolimits_{l\le t-1} A_{il}^P \right)$$

, where t_{i0}^P is the week index for movie i's first early advertising and $f_i(t)=(1+(\varphi_0+\varphi_1\mathbf{x_i})\cdot|t|)^{-1}$. Likewise, the effectiveness of movie i's launch advertising in week t, γ_{it} is represented by :

$$\text{(3-3)} \qquad \gamma_{it} = \left(f_i(t,\tau)\cdot I(t\le\tau) + I(t>\tau) \right)\cdot\left(\psi_{0i}^M + \psi_1^M \cdot \max(t - t_{i0}^M, 0) + \psi_2^M \sum\nolimits_{l\le t-1} A_{il}^M \right).$$

, where t_{i0}^M is the week index for movie's first launch advertising. The parameters ψ_{0i}^P and ψ_{0i}^M represent the maximum effectiveness of advertising of movie *i* after controlling for time-discounting, copy wearout/wearin, and repetition wearin/wearout effects. They are estimated separately for each movie because various idiosyncratic factors of movies can influence their advertising effectiveness. The final model, which we apply to our movie dataset, is derived by substituting eq. (3-2) and (3-3) in eq. (2-1) and (2-2). This final model tests consumers' time discounting behavior (H2a) by φ_0, the copy wearin/wearout effect of advertising (H2b) by ψ_1^P and ψ_1^M, repetition wearin/wearout effect of advertising (H2b) by ψ_2^P and ψ_2^M. Table 1 summarizes the hypotheses and their corresponding test parameters.

== Table 1 about here ==

PARAMETER ESTIMATION

The model is estimated by a hierarchical Bayesian method. With this method, the complex error term can be explicitly estimated and the random effects (i.e. κ_i , $\psi_{0i}{}^P$, $\psi_{0i}{}^M$) and heteroskedasticity of errors are easily incorporated. Before applying the hierarchical Bayesian method, we address possible non-stationarity and endogeneity of the variables.

Preliminary Steps

Panel unit-root test. We conduct a panel unit-root test to examine whether we need to difference our variables before we apply the geometric distributed lag model (1-2). We apply the panel unit-root test developed by Levin, Lin, and Chu (2002) to test for a common unit-root process for advertising, blog postings, and search volume. The test equation is $\Delta y_{it}=\alpha y_{i,t-1}+\sum_l \beta_i \Delta y_{i,t-l}+\varepsilon_{it}$, where the appropriate number of lags is determined by Schwarz Bayesian Criterion (SBC)[5)]. Table 2 shows that the null hypothesis of a unit root is rejected at the conventional 5% significance level, both in early and launch advertising periods. Thus, we can apply the GL model to the level of advertising and online buzz variables.

== Table 2 about here ==

Endogeneity. There are three potential endogeneity issues in our model. First, both time-invariant omitted variables and time-varying omitted variables can be correlated with the explanatory variables. Second, the complex error term in eq. (2-1) and (2-2) is generally correlated with the lagged dependent variables. Third, the advertising spending schedule for a movie can also be endogenous. We address these one by one.

Movie characteristics are an example of time-invariant variables that are potentially correlated with both weekly advertising expenditure and buzz level. For instance, movies with greater production budget tend to have bigger advertising spending and consequently enjoy bigger buzz each week. Sequel movies tend to enjoy greater search volumes because consumers may be searching for both the sequel and the original release. To reduce the confounding effects of time-invariant omitted variables, we estimate the time-invariant component of each movie (i.e. κ_i) via random-effects modeling which treats κ_i as a random variable with a normal distribution. We include a sufficiently large number of data points in the period before the first advertising to help identify κ_i.

Second, the complex the error term in eq. (2-1) and (2-2) is correlated with the lagged dependent variable (i.e. $y_{i,t-1}$ is correlated with $u_{i,t-1}$ and $u_{i,t-2}$; $y_{i,t-2}$ is correlated with $u_{i,t-2}$). To treat this endogeneity, we apply a two-stage estimation method, explained in detail in Appendix C.

Third, some time-varying variables not observed by researchers may be observed by managers and used to adjust future advertising spending. If movie studios monitor the buzz or word-of-mouth of their movies and adjust their advertising spending based on this buzz, the advertising spending schedule is endogenous. We conduct a simple test for the existence of a feedback loop from past online buzz to current advertising spending, controlling for previous advertising spending, using the test equation: $A_{it}=\alpha_i+\beta_1 y_{i,t-1}+\beta_2 y_{i,t-2}+\beta_3 A_{i,t-1}+\varepsilon_{it}$. We include two lags for previous buzz because the error term in eq. (2-1) and (2-2) includes the second lag error term ($u_{i,t-2}$). The model is estimated by the feasible generalized least squares (FGLS) method to account for possible heteroskedasticity across movies. For standard error estimation, we use the White heteroskedasticity and autocorrelation consistent covariance estimation method. The estimation results in Table 3 indicate there is no discernible evidence of a feedback loop up to two lags.

== Table 3 about here ==

The above finding of no feedback loop is supported by previous studies as well. Based on interviews with studio executives, Elberse and Anand (2007) argue that adjusting advertising spending schedules over time is unlikely because studios typically buy the vast majority (90-95 %) of television advertising several months prior to a movie's release. It is difficult and expensive to purchase additional TV advertising time based on pre-release outcomes such as consumers' word-of-mouth. This institutional feature is also reported by Onishi and Manchanda (2010), who argue that advertising plans are set far in advance of a new-product launch and are usually not varied after launch. This implies that the advertising spending schedule is exogenously determined well in advance.

Estimation Results

The details of the Bayesian estimation method are described in Appendix B and the results are summarized in Table 4. In this table, Model 1 has no consumer time-discounting, Model 2 has time-discounting only for launch advertising and Model 3 has time discounting for all advertising (i.e. early and launch). Note that Models 2 and 3 have a separate time-discounting parameter φ_1 for sequel movies.

== Table 4 about here ==

The results show, first, that the model without time-discounting factor (i.e. Model 1) is dominated by the models with time-discounting factor (Model 2 and Model 3) in terms of deviance information criterion (DIC). This indicates that time-discounting plays an import role in generating online buzz in the pre-launch period. However, Model 3 is marginally better than Model 2 in terms of DIC which implies that the discounting factor for the early advertising is nearly flat. To have a better sense of model fit, we calculate the correlation coefficient between the predicted and the observed online buzz. The fit is satisfactory for Model 2 and Model 3, whereas Model 1 is significantly worse than the other two models. Second, we find a switching from the discounting regime to the normal (i.e. non-discounting) regime as the movie release is approaching. This occurs about 1.5 weeks before the movie's release. Thus consumers tend to ignore a good portion of advertising conducted far from release, gradually increase their response level as movie release is approaching, and start to fully utilize the advertising only one to two weeks before release. Third, the discount factor is moderated by a movie characteristic, the sequel nature of a movie. A sequel movie is less vulnerable to time-discounting behavior than its competing non-sequel movies. Generally the original movie of a sequel tends to be very successful. The appeal of the original movie may induce consumers to be more interested in and curious about a sequel than its competing original movies, making consumers more inclined to search for and blog about the sequel movie. Figure 4 shows how the discounting factor changes over time from 20 weeks before release (t=-20) to five weeks after release (t=5), assuming that the regime switching occurs a week before release. It shows not only that discounting is significant but also that sequel movies are less susceptible to the discounting behavior. For example, in generating search volume for an original (sequel) movie, the effect of advertising two weeks before release is only 58% (69%) of what would be achieved in a week in the normal regime. Thus, advertising is about 19% more effective for sequel movies than for original movies in the two weeks before release (Figure 4(c)).

== Figure 4 about here ==

Fourth, for both search volume and blog postings, we find statistically significant copy wearout and repetition wearin effects (i.e. repetition wearin) of launch advertising (scenario b-3 in Figure 3). That is, net of consumer discounting and repetition wearin effects, the effectiveness of *launch advertising* gradually deteriorates after the start of the advertising. This implies that starting launch advertising early is not beneficial because its effect is lost gradually over the course of the pre- and post-launch period. On the other hand, a strong *repetition wearin* effect is observed for search volume and blog postings. The repetition wearin effect implies that previous advertising enhances the effectiveness of the current advertising, making consumers generate more online buzz for the movie. However, we do not observe statistically significant copy and repetition effects for *early advertising*. Last, we observe strong carryover of launch advertising while there is no statistically strong evidence of carryover for the early advertising. Instead the effect of early advertising on online buzz generation is limited to the current period.

In Figure 5, we simulate the time-varying effectiveness (i.e. γ_{it}) of a given launch advertising schedule. We assume that the total advertising budget is $25M and the first advertising starts 12 weeks before release. To mimic common industry practice, we gradually increase the weekly advertising spending toward the release week (Figure 5(a)).

== Figure 5 about here ==

With this given advertising schedule, we observe the copy wearout effect prevails initially. Then, the repetition wearin effect starts to dominate copy wearout as sufficient advertising effort is accumulated (Figure 5(b-1)). However, these effects are again dominated by consumers' discounting behavior, resulting in the hyperbolic shape in Figure 5(b-2).

Comparing the effect of advertising on search volume versus blog postings provides an insight on consumers' responsiveness to advertising. First, blogging behavior is discounted more severely than search behavior. One explanation is that it generally takes more effort to post information on a blog than to search the information on the Internet. Thus, in terms of cost and benefit tradeoff, posting information about a product that will be released in the distant future is not as meaningful as searching for the information for most consumers. Second, the carryover rate of launch advertising is smaller for blogging behavior than for search behavior. This may imply the blogging behaviors among consumers may be more independent and self-driven than search behaviors.

In sum, our results support H1a (the contemporaneous effect of pre-launch advertising) for both early and launch advertising and support H1b (the carryover effects) only for launch advertising. H2a (time-discounting behavior) is strongly supported and H2b (time wearin/out, repetition wearin/out) is supported only for launch advertising.

PRE-LAUNCH BUZZ AS AN INFLUENCER OF MOVIE CONSUMPTION

In this section we show that pre-launch buzz is not merely a predictor but also an influencer to the box office revenue. The distinction is important because only if pre-launch buzz is an influencer can we make recommendations on pre-launch advertising scheduling to increase subsequent box office revenues (Eliashberg and Shugan 1997). To test the predictor vs. influencer scenarios, we measure the potential impact of pre-launch buzz on actual box office revenue for each week in the movie's life cycle. We apply a log-log model for each week where the dependent variable is the box office revenue and the explanatory variables include pre-launch buzz, marketing variables such as weekly advertising spending and number of screens, and various movie characteristics such as genre, studio, monthly seasonality, a holiday indicator, and pre-launch critics ratings. We estimate separate cross-sectional models by week (from opening week to the fifth week after release) and for the entire post-launch period. The model is eq. (4):

$$(4)\quad REVENUE_i = e^{\beta_0} \cdot AD_i^{\beta_1} \cdot SCREENS_i^{\beta_2} \cdot PREBUZZ_i^{\beta_3+\beta_4 SIMP\_TITLE_i} \cdot \prod_k \left(X_{ik}\right)^{\gamma_k} \cdot e^{\sum_l \delta_l \cdot Z_{il}} e^{\varepsilon_i}$$

, where i is the index for movie. $REVENUE_i$ is the box office revenue, AD_i is the advertising spending, $SCREENS_i$ is the number of allocated screens of movie i. $PREBUZZ_i$ the amount of pre-launch buzz of movie i up to one week before release[6)] and $SIMP\_TITLE_i$ is 1 if movie i's title consists of only one word, and 0 otherwise[7)]. X_{ik} is the kth continuous variable and Z_{il} is the lth dummy variable of movie i. $PREBUZZ_i$ is treated as exogenous because it is determined in the pre-launch period. However, the number of screens and the advertising expenditure can be endogenous (Basuroy, Desai, and Talukdar 2006; Elberse and Eliashberg 2003; Ho, Dhar, and Weinberg 2009).

A two-stage least square estimation is applied to treat the endogeneity of advertising expenditure and the number of screens. In the first stage, we regress the number of screens and advertising expenditure on exogenous variables as well as predetermined variables in the previous period. The exogenous variables include movie characteristics (genre, production budget, runtime, and studio), critics' ratings in prerelease period, pre-launch buzz, pre-launch advertising, national holiday dummy, and monthly seasonal dummy. Predetermined variables (i.e. endogenous but determined in the previous periods) variables include the number of screens, advertising expenditure, and box office revenue of the previous week as well as the remaining advertising budget in this week.

In the second stage, we substitute $SCREENS_i$ and AD_i in eq. (4) with the estimated number of screens and estimated advertising expenditure from the first stage. We do this analysis for each week from the opening week (t=0) to the fifth week after release (t=5) and examine the pattern of the coefficients. Two different pre-launch buzz variables – Google search volume index and blog postings – are used separately because they are highly correlated. Table 5 shows the estimation results for the focal coefficients – $\beta_1, \beta_2, \beta_3, \beta_4$.

== Table 5. About here ==

The estimation results show that the association between pre-launch buzz and weekly box office revenue is the strongest in the opening week, then diminishes over time and finally becomes insignificant. Thus we may conclude that the pre-launch buzz is an influencer on consumers' movie-going decisions, supporting H3. As time goes by, consumers rely more heavily on new and presumably more accurate information on the movie experience that becomes available as its cumulative attendance rises. For example, early viewers may spread their opinions through various media and later would-be-viewers attach more weight to these opinions than to the pre-launch buzz.

Another interesting finding is that the influence of search volume is greater and longer-lasting than that of blog postings. This may be explained by the fact that searching is a more general behavior than blog posting. That is, those who post their opinions about a movie are likely to search the movie, while not all searchers post their opinions on blogs. This finding demonstrates not only the information value of search behavior but also highlights the importance for studios of monitoring search volumes in the pre-launch periods.

MANAGERIAL IMPLICATIONS

Our findings have managerial implications on advertising budget allocation for new products, advertising of well-known brands vs. new brands, and the informational value of pre-launch buzz. First, we find four different forces are at work to affect the time-varying effectiveness of advertising: consumers' time-discounting behavior, copy wearout effect, repetition wearin effects of advertising, and the time at which switching from the discounting to the non-discounting regime occurs. These forces are relevant for the allocation of advertising in the pre- and post-launch period. Early pre-launch advertising not only has a lower contemporaneous effect due to consumers' time-discounting behavior, but also reduces the effects of later advertising due to copy wearout effect. Therefore starting advertising early is not advisable if its sole purpose is to maximize pre-launch buzz. On the other hand, once the advertising starts, it is better to concentrate the spending as much as possible, in order to benefit from the repetition wearin effect. Altogether, these findings argue that the advertising budget allocation for movies should be concentrated in the last few weeks prior to release, when consumers no longer discount the advertising information.

We illustrate the effect of different advertising budget allocations in Figure 5. Scenario 1 and Scenario 2 spend the same $22.9M on launch advertising in the pre-launch period. Scenario 1 starts its advertising eight weeks before release and has a big advertising campaign in week -5[8)]. Scenario 2, on the other hand, starts its advertising just four weeks before release. The advertising expenditure during weeks -8 through -5 of Scenario 1 is proportionately allocated to week -4 through week 0 under Scenario 2. Thus, Scenario 1 and 2 have the same total advertising budget, but under Scenario 2 the budget is more concentrated around the release week (See Figure 6(a)). Under these scenarios, we simulate the search volume with the parameter estimates in Table 4(a).

== Figure 6 about here ==

Scenario 1 generates 47.61 searches while scenario 2 generates 54.23. Thus, by concentrating the advertising around the release week, we can significantly increase consumer's search. In another simulations, we maintain the advertising schedule of Scenario 2 and calculate the required advertising budget that will generate 47.61 searches. The required budget is reduced to $20.3M. Thus in this specific experiment, we can save more than 10% of the advertising budget to generate the same search volume by allocating the budget more carefully.

Second, the finding that consumers discount sequel movies less severely than original movies implies the importance of brand equity in generating online buzz. That is, with successful original movies of the past, sequel movies may enjoy a healthy brand equity which helps generate more buzz with the same amount of advertising even long before their releases. This may be applicable to other industries. For example, in the Tablet PC category, Apple's current advertising for its next generation iPad, to be launched in six months, will generate significantly more buzz than the advertising from a new entrant, all else equal. Thus the weaker the position of a company in a market, the more it should concentrate its advertising around the release time of its new product in the market.

Last but not least, the finding that pre-launch online buzz is an influencer of box office revenue implies that such online buzz is a *link* between launch advertising and post-launch sales. As such, we can consider pre-launch buzz as a proxy measure for a new product's sales, especially for short life-cycle products such as movies. This gives firms a new opportunity to increase the effectiveness of their pre-launch advertising, by closely monitoring their pre-launch online buzz and adjusting their advertising accordingly.

CONCLUSION

This study has examined the effect of new-product advertising on online buzz generation and the effect of pre-launch online buzz on box office revenue of movies. Based on the literature, we developed hypotheses on the effect of pre-launch advertising and the role of pre-launch buzz on new-product sales, and tested them in two separate analyses. In the first analysis we used a geometric distributed lag model to incorporate the time-varying nature of the effectiveness of new-product advertising and test advertising's impact on buzz generation. In the second analysis, we developed a multiplicative model to assess the role of pre-launch buzz in movie-going behavior. We applied the models to a movie database consisting of weekly advertising spending, weekly number of blog postings, and weekly search volume measures during the pre- and post-launch periods of 158 movies that were widely released in 2008 and 2009.

The main findings include that (1) the pre-launch phase of a new movie starts with a consumer response regime that is subject to information discounting, and then switches to a non-discounting regime around 1.5 weeks before release, (2) the effect of advertising deteriorates with the elapsed time since the first advertising, (3) previous advertising spending boosts the effectiveness of current advertising in generating buzz, (4) the effect of early advertising does not carry over to the subsequent periods, (5) pre-launch online buzz influences subsequent box-office movie revenues. Based on these findings, we show that movie studies can significant increase their advertising impact by carefully scheduling their campaigns pre- and post-launch.

Our study is subject to some limitations. First, we do not collect valence information during the pre-launch period. Search volume itself does not provide any valence information; blog postings may contain valence, however due to data limitations we could not derive the content of each blog posting. Instead, we have the weekly number of blog postings on a specific topic. Some studies show that valence and/or user ratings play an important role in explaining new-product sales. Second, we examine the effect of pre-launch advertising only on the consumer side, ignoring its possible signaling effect on the supply side. For instance, significant early advertising can signal that a big blockbuster movie will be released in a certain week and thus change a competitor's decision to open in the same week. Similarly, the signal may persuade exhibitors (i.e. movie theaters) to allocate more screens for the movie (Ho, Dhar, and Weinberg 2009). Investigating these supply-side factors is a promising future research topic. Third, the current study did not formalize the optimal week-by-week allocation of advertising during the pre- and the post-launch period. Such dynamic optimization is another interesting topic for future research.

REFERENCES

Ainslie, George and Nick Haslam (1992), "Hyperbolic Discounting," In *Choice Over Time*, Loewenstein, George and Jon Elster, eds. New York: Russell Sage Foundation, 57–92.

Banerjee, Abhijit V. (1992), "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics*, 107 (August), 797-817.

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

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

Bikchandani, Sushil, David Hirshleifer, and Ivo Welch (1992), "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100 (October), 992-1026.

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

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

Dhar, Vansant and Elaine A. Chang (2009), "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing*, 23 (4), 300-307.

Draganska, Michaela and Daniel Klapper (2010), "Choice Set Heterogeneity and the Role of Advertising: An Analysis with Micro and Macro Data," Working Paper, Graduate School of Business, Stanford University.

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

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

Eliashberg, Jehoshua and Steven M. Shugan (1997), "Film Critics: Influencers or Predictors?," *Journal of Marketing*, 61 (April), 68-78.

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

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

Greenberg, Allan and Charles Suttoni (1973), "Television Commercial Wearout," *Journal of Advertising Research*, 13 (October), 47-54.

Ho, Jason Y. C., Tirtha Dhar, and Charles B. Weinberg (2009), "Playoff payoffs: Super Bowl Advertising for Movies," *International Journal of Research in Marketing*, 26 (September), 168-79.

Janiszewski, Chris (1993), "Preattentive Mere Exposure Effects," *Journal of Consumer Research*, 20 (December), 376-92.

Joshi, Amit and Michael Trusov (2010), "Intrinsic Media Transmittivity – A New Concept and Dynamic Metrics of the New Media," Working Paper, College of Business Administration, University of Central Florida.

Karniouchina, Ekatenia V. (2011), "Impact of Star and Movie Buzz on Motion Picture Distribution and Box Office Revenue," *International Journal of Research in Marketing*, 28 (March), 62-74.

Levin, Andrew, Chien-Fu Lin, and Chia-Shang James Chu (2002), "Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties," *Journal of Econometrics*, 108 (May), 1-24.

Little, John D. C. (1979), "Aggregate Advertising Models: the State of the Art," *Operations Research*, 27(4), 629-67.

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

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

Read, Daniel and George Lowenstein (2000), "Time and Decision: Introduction to the Special Issue," *Journal of Behavioral Decision Making*, 13 (2), 141-44.

Onishi, Hiroshi and Puneet Manchanda (2011), "Marketing Activity, Blogging, and Sales," SSRN Working Paper.

Petty, Richard E. and John T. Cacioppo (1979), “Issue Involvement Can Increase or Decrease Persuasion by Enhancing Message-Relevant Cognitive Responses,” *Journal of Personality and Social Psychology,* 37 (October), 1915–26.

Sawyer, Alan G. (1981), “Repetition, Cognitive Response, and Persuasion,” In *Cognitive Response to Persuasion*, Petty, Richard E., Thomas M. Ostrom, and Timothy. C. Brock, eds. New York: Psychology Press, 237-62.

Shapiro, Stewart, Deborah J. MacInnis, and Susan E. Heckler (1997), “The Effects of Incidental Advertising Exposure on the Formation of Consideration Sets,” *Journal of Consumer Research,* 24 (June), 94-104.

Soman, Dilip, George Ainslie, Shane Frederick, Xiuping Li, John Lynch, Page Moreau, Andrew Mitchell, Daniel Read, Alan Saywer, Yaacov Trope, Klaus Wertenbroch, and Gal Zauberman (2005), “The Psychology of Intertemporal Discounting: Why are Distant Events Valued Differently from Proximal Ones?,” *Marketing Letters*, 16 (3/4), 347-60.

Trope, Yaacov and Nira Liberman (2003), “Temporal Construal,” *Psychological Review*, 110 (3), 403-21.

Zufryden, Fred S. (1996), “Linking Advertising to Box Office Performance of New Film Releases: A Marketing Planning Model,” *Journal of Advertising Research*, 50(July/August), 29-41.

FOOTNOTES

1) Early advertising may also be used as an early signal to exhibitors to help secure wide screen coverage. Per conversation with studio executives, most screen contracts have already been secured prior to the advent of the early advertising.

2) We developed several different version of search keywords and found that the keyword <movie title> + "movie" is the most valid in the sense that it produced the highest fit with the movie's observed advertising and launch schedules.

3) With the model as given, we assume that the online buzz is stationary. Otherwise, we can difference the online buzz series before applying the model. In a separate sub-section, we show online buzz is stationary by a formal statistical test.

4) The time index has negative values in the pre-launch period, the value of zero in the release week, and positive values in the post-launch period, incrementing by one. For example, six weeks before release is represented by $t = -6$.

5) One exception in determining the appropriate number of lags is the advertising spending during the early advertising period in which case the maximum number of lags is constrained to be 1 due to data scarcity.

6) PREBUZZ is measured up to one week before release but not up to the release week to remove simultaneity issue between the buzz and box office revenue in the release week.

[7] For example, the movie *Up* is coded as 1, while the movie *Harry Potter and the Half-blooded Prince* is coded as 0. The variable SIMP_TITLE captures the simple title effect and enables to have clearer estimates for β_3 and β_4.

[8] Scenario 1 closely mimics the launch advertising schedule of the movie *Transformers: Revenge of the Fallen*.

Table 1. Hypotheses and Corresponding Parameters

Hypothesis	Parameters	Direction
H1a	$\beta, \psi_{0i}^{P}, \psi_{0i}^{M}$	$\beta, \psi_{0i}^{P}, \psi_{0i}^{M} > 0$
H1b	λ, δ	$0 < \lambda, \delta < 1$
H2a	φ_0	$\varphi_0 > 0$
H2b	$\psi_1^{P}, \psi_2^{P}, \psi_1^{M}, \psi_2^{M}$	Not known a priori

Table 2. Results of Panel Unit Root Tests

Online Buzz	Advertising period	N (Total obs.)	Test statistic	P-value
Advertising	Before launch ad	1971	-21.506	0.000
	During launch ad	4224	-14.391	0.000
Blog Postings	Before launch ad	7741	-23.004	0.000
	During launch ad	1879	-14.765	0.000
Search Volume	Before launch ad	5605	-3.465	0.000
	During launch ad	1896	-8.011	0.000

Table 3. Test of feedback loop (N=158, T=66)

(a) Search volume as pre-launch buzz

Independent variables	Estimates (t-stat.)
First lag advertising spending	0.771 (88.045)
First lag search volume	-5.855 (-1.179)
Second lag search volume	4.873 (1.252)

(b) Blog postings as pre-launch buzz

Independent variables	Estimates (t-stat.)
First lag advertising spending	0.770 (87.587)
First lag blog postings	0.036 (0.752)
Second lag blog postings	-0.016 (-0.424)

Table 4. Estimation Results

(a) Search Volume Measure (N=158, T=66)

		Model 1	Model 2	Model 3 (full model)
Early advertising	ψ_1^P	-1.96×10^{-5} [-1.08×10^{-4}, 6.18×10^{-5}]	-4.35×10^{-6} [-5.45×10^{-5}, 4.45×10^{-5}]	1.82×10^{-4} [-5.09×10^{-4}, 9.13×10^{-4}]
	ψ_2^P	7.75×10^{-8} [-5.09×10^{-8}, 2.11×10^{-7}]	6.84×10^{-8} [-1.64×10^{-8}, 1.55×10^{-7}]	2.85×10^{-7} [-8.73×10^{-7}, 1.42×10^{-6}]
	λ	3.51×10^{-4} [8.45×10^{-6}, 0.001]	3.92×10^{-4} [1.09×10^{-5}, 0.001]	3.92×10^{-4} [1.07×10^{-5}, 0.001]
Launch advertising	ψ_1^M	**-1.34×10^{-5}** **[-1.91×10^{-5}, -7.43×10^{-6}]**	**-3.98×10^{-5}** **[-5.45×10^{-5}, -2.59×10^{-5}]**	**-4.00×10^{-5}** **[-5.48×10^{-5}, -2.65×10^{-5}]**
	ψ_2^M	**1.07×10^{-8}** **[9.10×10^{-9}, 1.22×10^{-8}]**	**1.40×10^{-8}** **[1.11×10^{-8}, 1.70×10^{-8}]**	**1.38×10^{-8}** **[1.09×10^{-8}, 1.68×10^{-8}]**
	δ	**0.548** **[0.553, 0.574]**	**0.564** **[0.553, 0.574]**	**0.562** **[0.552, 0.572]**
Discount parameter	φ_0	N.A.	**0.677** **[0.619, 0.742]**	**0.729** **[0.664, 0.799]**
	φ_1	N.A.	**-0.273** **[-0.425, -0.116]**	**-0.275** **[-0.421, -0.099]**
Regime switching timing	τ	N.A.	**-1.505** **[-1.973, -1.023]**	**-1.494** **[-1.976, -1.024]**
-2log-lik		3812.810	-1564.830	-1590.470
Deviance Info. Criterion		4343.210	-1026.190	-1067.410
r*		0.845	0.860	0.860

(b) Blog Postings (N=158, T=66)

		Model 1	Model 2	Model 3 (full model)
Early advertising	ψ_1^P	-7.05×10^{-4} [-0.002, 0.001]	-8.19×10^{-4} [-0.002, 3.97×10^{-4}]	-0.002 [-0.016, 0.011]
	ψ_2^P	1.84×10^{-6} [-8.19×10^{-7}, 4.53×10^{-6}]	**2.31×10^{-6}** **[4.37×10^{-7}, 4.13×10^{-6}]**	-0.002 [-0.015, 0.011]
	λ	6.92×10^{-4} [1.93×10^{-5}, 0.003]	6.21×10^{-4} [1.74×10^{-5}, 0.002]	6.12×10^{-4} [1.53×10^{-5}, 0.002]
Launch advertising	ψ_1^M	**-2.51×10^{-4}** **[-3.36×10^{-4}, -1.74×10^{-4}]**	**-9.95×10^{-4}** **[-0.001, -6.77×10^{-4}]**	**-0.001** **[-1.29×10^{-3}, -7.94×10^{-4}]**
	ψ_2^M	**1.94×10^{-7}** **[1.72×10^{-7}, 2.16×10^{-7}]**	**1.19×10^{-7}** **[6.12×10^{-8}, 1.78×10^{-7}]**	**1.19×10^{-7}** **[6.58×10^{-8}, 1.71×10^{-7}]**
	δ	**0.156** **[0.138, 0.174]**	**0.242** **[0.228, 0.256]**	**0.241** **[0.227, 0.255]**
Discount parameter	φ_0	N.A.	**1.922** **[1.735, 2.117]**	**1.998** **[1.806, 2.225]**
	φ_1	N.A.	**-0.458** **[-0.792, -0.075]**	**-0.497** **[-0.852, -0.072]**
Regime switching timing	τ	N.A.	**-1.499** **[-1.979, -1.026]**	**-1.502** **[-1.973, -1.028]**
-2log-lik		70384.200	64739.100	64768.900
Deviance Info. Criterion		70867.300	65240.300	65239.200
r*		0.547	0.810	0.811

Note: 95 % probability interval in (); Bold numbers indicate the parameter estimates are statistically significant at 5% level. r is the correlation coefficient between the observed and the estimated values. We only estimate common slope for λ and δ.

Table 5. Box Office Estimation Results

(a) Pre-launch Buzz: Google Search Volume Measure

Week (No. Obs)†	R^2 (Adj. R^2)	AD (t-statistic)	SCREENS (t-statistic)	PREBUZZ (t-statistic)	SIMP_TITLE (t-statistic)
0	0.898	0.231***	0.680***	**0.195*** **	**-0.182*** **
(158)	(0.881)	(3.855)	(18.061)	**(3.571)**	**(-3.288)**
1	0.869	0.188***	0.719***	**0.167*** **	**-0.152*** **
(158)	(0.850)	(5.714)	(20.992)	**(3.291)**	**(-2.873)**
2	0.810	0.180***	0.728***	0.069	-0.076
(156)	(0.780)	(8.942)	(18.292)	(1.348)	(-1.427)
3	0.809	0.225***	0.819***	0.076	-0.099*
(150)	(0.779)	(9.069)	(16.238)	(1.452)	(-1.797)
4	0.849	0.204***	0.930***	0.042	-0.083
(133)	(0.822)	(7.392)	(16.576)	(0.759)	(-1.404)
5	0.879	0.201***	0.889***	0.051	-0.035
(108)	(0.851)	(6.359)	(13.626)	(0.898)	(-0.601)

(b) Pre-launch Buzz: Blog Postings

Week (No. Obs)†	R^2 (Adj. R^2)	AD (t-statistic)	SCREENS (t-statistic)	PREBUZZ (t-statistic)	SIMP_TITLE (t-statistic)
0	0.883	0.200***	0.703***	**0.115** **	**-0.068** **
(158)	(0.866)	(3.157)	(18.108)	**(2.200)**	**(-2.584)**
1	0.861	0.187***	0.738***	0.063	**-0.045* **
(158)	(0.841)	(5.537)	(21.256)	(1.328)	**(-1.852)**
2	0.800	0.195***	0.777***	0.030	-0.023
(156)	(0.770)	(8.148)	(17.071)	(0.652)	(-0.994)
3	0.808	0.222***	0.829***	0.065	-0.035
(150)	(0.778)	(8.945)	(16.785)	(1.375)	(-1.426)
4	0.851	0.195***	0.943***	0.002	-0.028
(133)	(0.824)	(7.193)	(17.446)	(0.030)	(-1.102)
5	0.879	0.198***	0.893***	0.075	-0.033
(108)	(0.852)	(6.558)	(14.460)	(1.388)	(-1.263)

*** significant at 1% level, ** significant at 5% level, * significant at 10% level.

†The number of observation decreases over time as movies with zero advertising expenditure in the week are discarded from the sample.

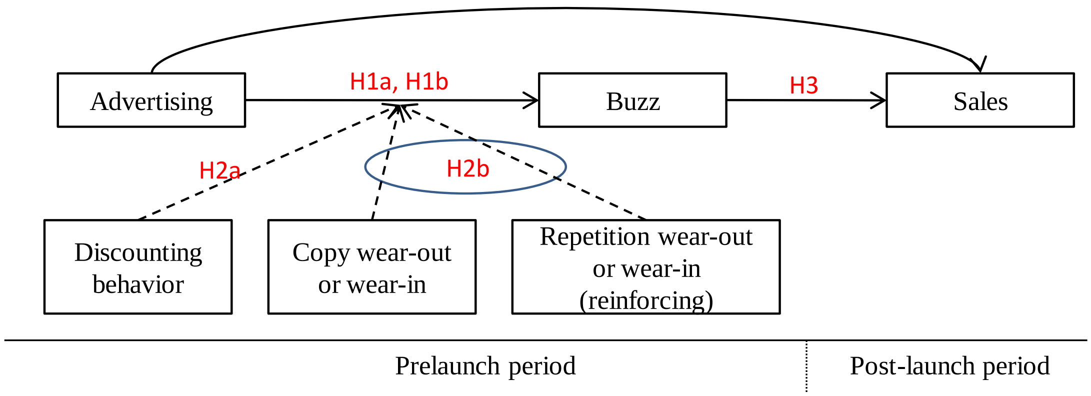
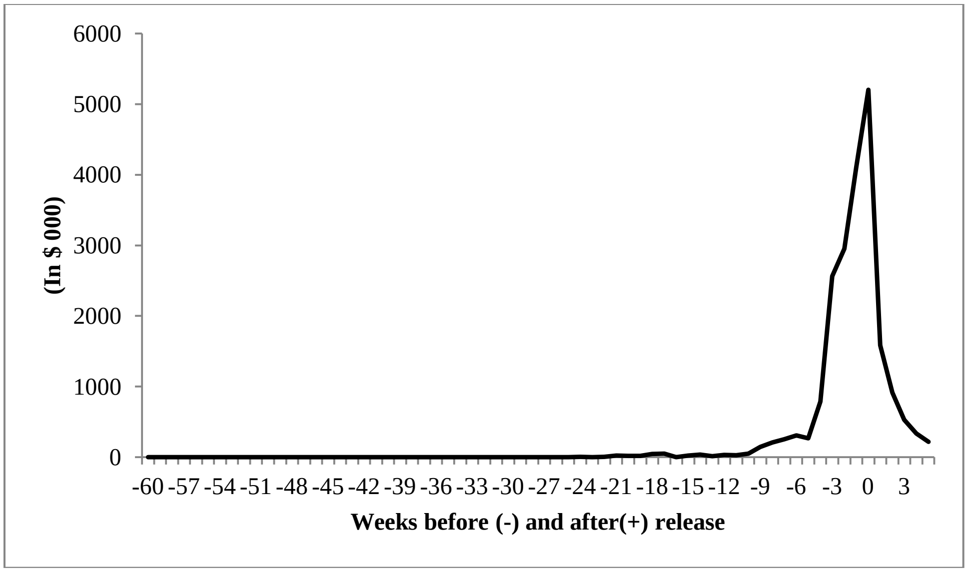
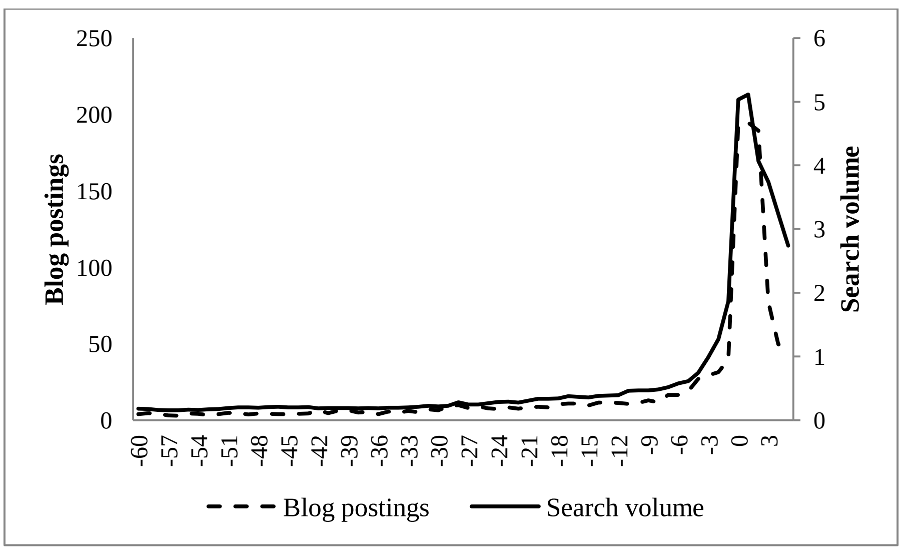


Figure 1. Hypotheses and the Relationship between Components of the Model

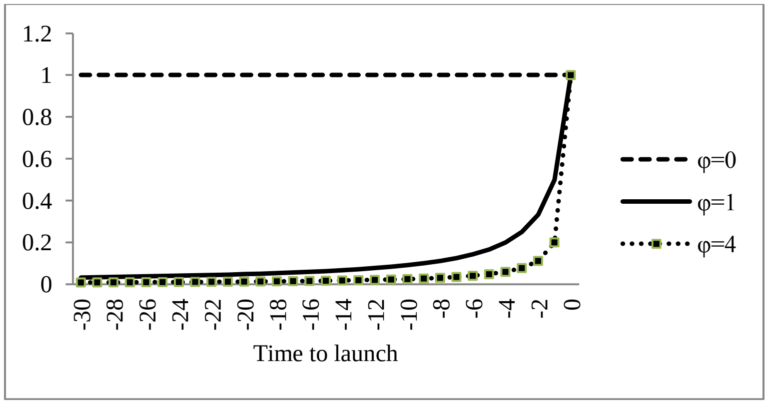


(a) Average Weekly Advertising Spending of Movies

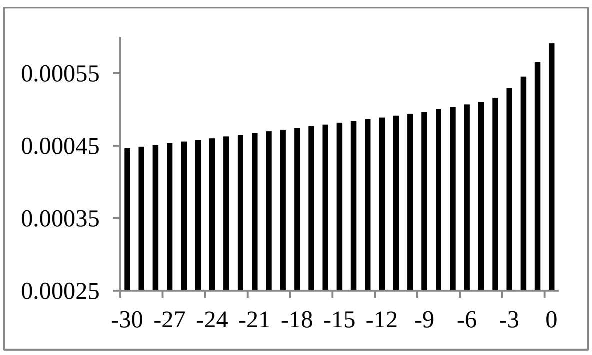


(b) Average Weekly Blog Postings and Search Volume Measure

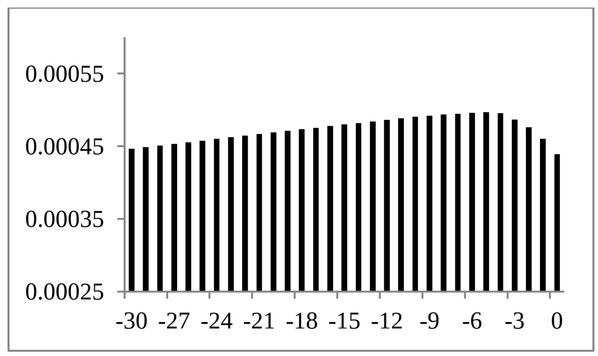
Figure 2. Average Weekly Advertising Spending and Buzz Volume



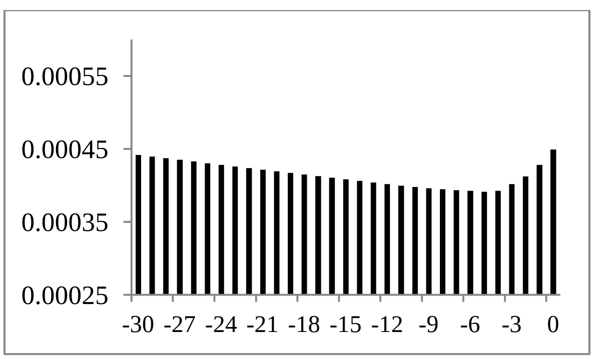
(a) The effect of φ on discount factor



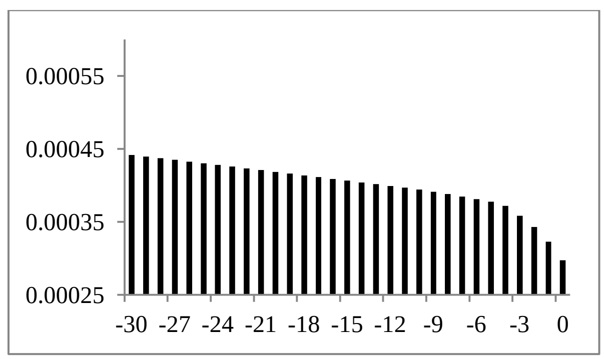
(b-1) Copy wearin and repetition wearin



(b-2) Copy wearin and repetition wearout



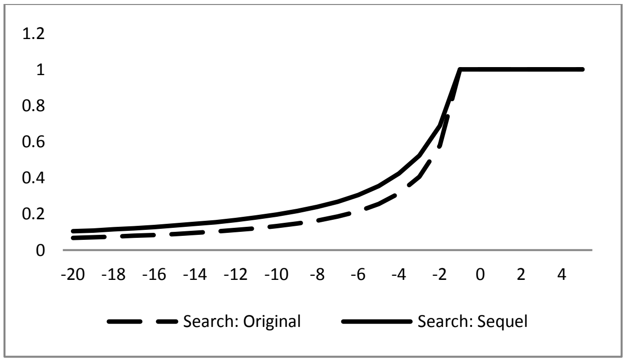
(b-3) Copy wearout and repetition wearin



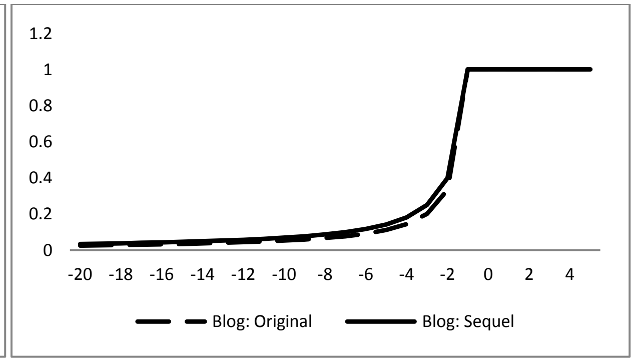
(b-4) Copy wearout and repetition wearout

(b) Various combinations of copy wearin/out and repetition wearin/out

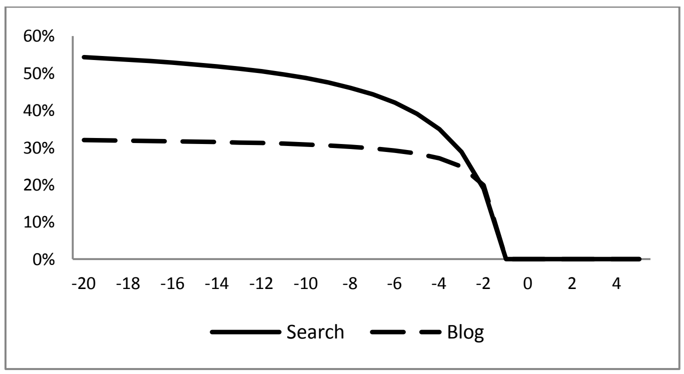
Figure 3. An Illustration of time-discounting effect and advertising wearin/wearout



(a) Discount factor for search volume

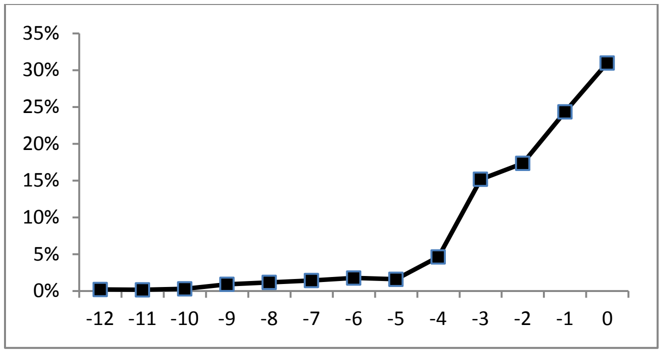


(b) Discount factor for blog postings

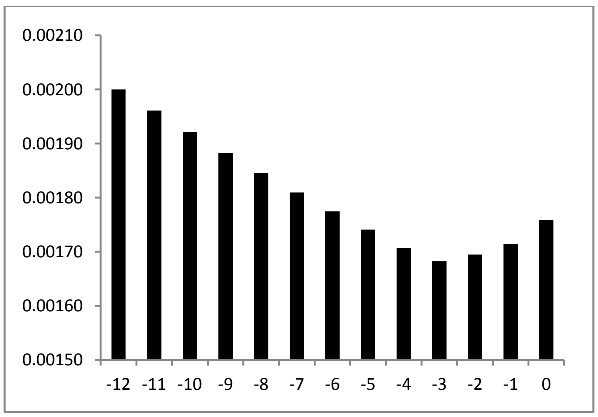


(c) Discounting factor ratio: Sequel/Original - 1

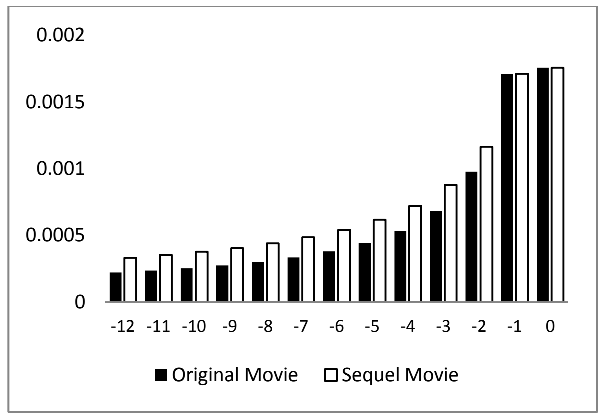
Figure 4. Changes in Discounting Factor and Discounting Factor Ratio



(a) Advertising budget allocation

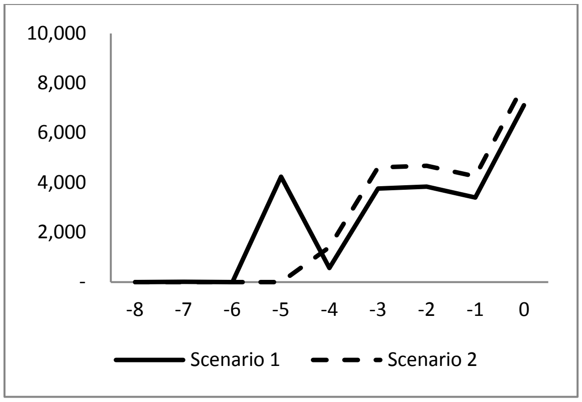


(b-1) Copy wearout and repetition wearin

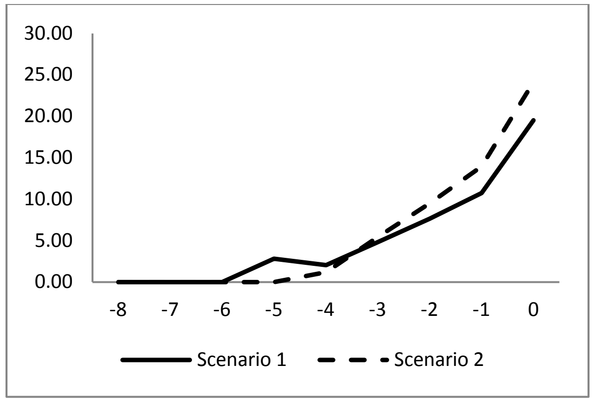


(b-2) Total effect considering discounting

Figure 5. Time-varying Effectiveness of Advertising in the Pre-launch Period



(a) Weekly advertising expenditure ($000)



(b) Search volume generation

Figure 6. Search Volume Simulation

Appendix A: Data Example

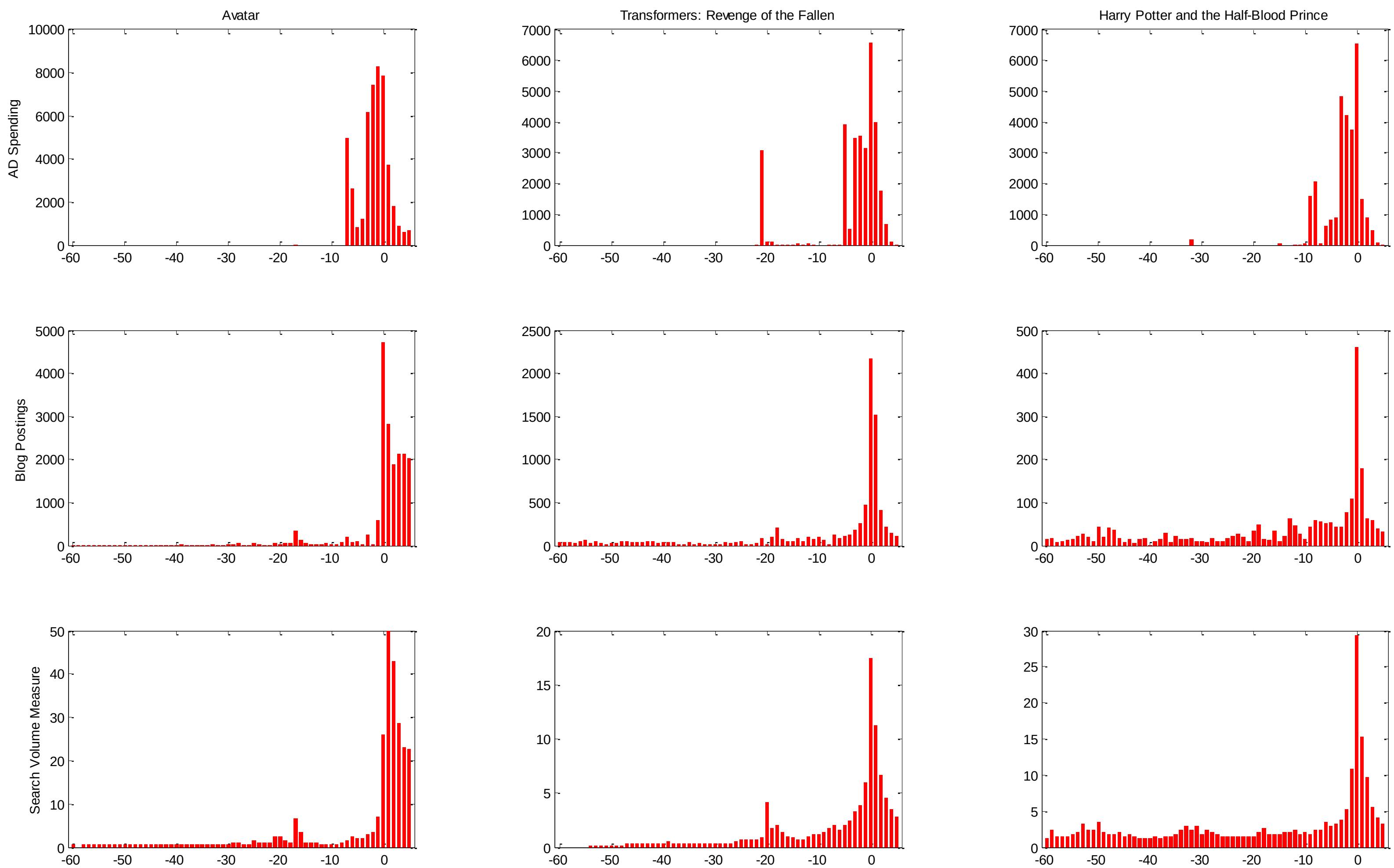


Figure A1. Examples of Advertising Spending, Blog Postings, and Search Volume during the Pre-launch Period

Appendix B: Hierarchical Bayesian Model

The final model is represented as in (A1) by use appropriate period dummy variables I_i^P, D_{0i}, D_{1i}, and D_{2i}.

$$\text{(A1)}\quad \begin{aligned} y_{it} &= \left[\kappa_i\right]\cdot D_{0i} \\ &+\left[\kappa_i(1-\lambda\cdot I_i^P)+\beta_{it}A_{it}+\lambda\cdot I_i^P\cdot y_{i,t-1}\right]\cdot D_{1i} \\ &+\left[\kappa_i(1-\delta)(1-\lambda\cdot I_i^P)+\gamma_{it}(A_{it}-\lambda\cdot I_i^P\cdot A_{i,t-1})+(\lambda\cdot I_i^P+\delta)y_{i,t-1}-\lambda\cdot I_i^P\cdot\delta y_{i,t-2}\right]\cdot D_{2i} \\ &+u_{it}-\left(\lambda\cdot I_i^P D_{1i}+\left(\lambda\cdot I_i^P+\delta\right)D_{2i}\right)u_{i,t-1}+\left(\lambda\cdot I_i^P\cdot\delta\cdot D_{2i}\right)u_{i,t-2} \end{aligned}$$

, where

$$\beta_{it}=\left(f_i(t,\tau)\cdot I(t\le\tau)+I(t>\tau)\right)\cdot\left(\psi_{0i}^P+\psi_1^P\cdot\max(t-t_{i0}^P,0)+\psi_2^P\sum_{l\le t-1}A_{il}^P\right),$$
$$\gamma_{it}=\left(f_i(t,\tau)\cdot I(t\le\tau)+I(t>\tau)\right)\cdot\left(\psi_{0i}^M+\psi_1^M\cdot\max(t-t_{i0}^M,0)+\psi_2^M\sum_{l\le t-1}A_{il}^M\right).$$

I_i^P is 1 if movie i has the early advertising. D_{0i}, D_{1i}, D_{2i} are period dummy variables of movie i, representing the period before the first advertising, the early advertising period, and the launch advertising period, respectively. Note in eq. (A1), we estimate only common slope for the carryover parameter, λ and δ.

For the movie characteristics that affect the discounting function, we consider the sequel nature of a movie as the movie characteristic that influences the discounting speed. Thus, $f(t,\tau)=\left(1+\left(\varphi_0+\varphi_0\cdot SEQUEL_i\right)|t-\tau|\right)^{-1}$, where $SEQUEL_i$=1 if movie $_i$ is a sequel of its original movie.

For the intercept κ_i and the baseline advertising effects ψ_{0i}^P and ψ_{0i}^M, we specify random effects to minimize confounding effects that would arise if we specify common effects for them.

We assume they follow normal distributions. For the means and variances we specify diffuse priors. Therefore,

$$\kappa_i \sim N(\mu^{\kappa}, \sigma^{\kappa}),$$
$$\mu^{\kappa} \sim N(0, 10^{-3}), \sigma^{\kappa} \sim IG(10^{-3}, 10^{-3})$$

$$\psi_{i0}^{P} \sim N(\theta_0^{P}, \xi_0^{P})$$
$$\theta_0^{P} \sim N(0, 10^{-3}), \xi_0^{P} \sim IG(10^{-3}, 10^{-3})$$

$$\psi_{i0}^{M} \sim N(\theta_0^{M}, \xi_0^{M})$$
$$\theta_0^{M} \sim N(0, 10^{-3}), \xi_0^{M} \sim IG(10^{-3}, 10^{-3})$$

For the rest of parameters, we specify a common slope for each parameter because the purpose of the study is to find the prevailing patterns in the data rather than to find individual heterogeneity. For each parameter, we specify appropriate diffuse priors. Therefore,

$$\varphi_0 \sim N(0\ , 10^{-3}),$$
$$\varphi_1 \sim N(0\ , 10^{-3}),$$

$$\psi_k^{P} \sim N(0, 10^{-3}), \qquad k = 1, 2$$
$$\psi_k^{M} \sim N(0, 10^{-3}), \qquad k = 1, 2$$

$$\lambda \sim Beta(1,1)$$
$$\delta \sim Beta(1,1)$$

For the errors, we assume they follow a multivariate normal distribution with mean zero and heteroskedastic variance. For the variance, we specify a diffuse prior. Therefore,

$$\mathbf{u_i} = \left(u_{i,-60}, \ldots, u_{i,5}\right) \sim MVN(\mathbf{0}, \sigma_i^{\mathbf{u}} \mathbf{I}_{66})$$
$$\sigma_i^{\mathbf{u}} \sim IG(1, 0.1)$$

Appendix C: Two-Stage Estimation

By construction of eq. (A1), the lagged dependent variables (i.e. $y_{i,t-1}$, $y_{i,t-2}$) are correlated with the lagged error terms ($u_{i,t-1}$, $i_{i,t-2}$), resulting in the endogeneity issue. To reduce the inconsistency in parameter estimates resulting from the endogeneity, we apply a two-stage estimation method. In the first stage, we filter the lagged dependent variables so that the filtered quantity (e.g. $y_{i,t-1}$) is free of its corresponding errors (e.g. $u_{i,t-1}$, $u_{i,t-2}$, $u_{i,t-3}$). This does not completely eliminate the endogenous variation in the lagged dependent variables because a lagged dependent variable (e.g. $y_{i,t-1}$)in the right hand side of eq. (A1) is again a function of its lagged dependent variables (e.g. $y_{i,t-2,}$ $y_{i,t-3}$). But this will reduce the endogeneity to some extent because the error terms are eliminated. Specifically, we estimate the parameters based on the assumption that the error is a Gaussian white noise. That is, $y_{it} \sim N(g(y_{i,t-1}, y_{i,t-2}; \varpi), \sigma_i^2)$, where ϖ is the set of parameters. Note that the lagged errors are not included in the function g(·). Therefore, the variation in the filtered value of y_{it} comes from the lagged dependent variables ($y_{i,t-1}$, $y_{i,t-2}$) but not from the errors. The fitted values of the lagged dependent variables ($\hat{y}_{i,t-1}, \hat{y}_{i,t-2}$) and the current error drawn from the above normal distribution ($\hat{u}_{it}$) are saved for each movie and week. In the second stage, we substitute the lagged dependent variables $y_{i,t-1}(y_{i,t-2})$ by their filtered counterparts ($\hat{y}_{i,t-1}, \hat{y}_{i,t-2}$) and the lagged errors ($u_{i,t-1}$, $u_{i,t-2}$) by their estimated counterparts ($\hat{u}_{i,t-1}, \hat{u}_{i,t-2}$) from the first stage. The parameters are estimated by assuming that the current error term follows a Gaussian white noise with variance σ_i^2. That is,

$$y_{it} \sim N(g(\hat{y}_{i,t-1}, \hat{y}_{i,t-2}, \hat{u}_{i,t-1}, \hat{u}_{i,t-2}; \varpi), \sigma_i^2)$$