Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Blogs, Advertising, and Local-Market Movie Box Office Performance

Shyam Gopinath, Pradeep K. Chintagunta, Sriram Venkataraman

MANAGEMENT SCIENCE
Vol. 59, No. 12, December 2013, pp. 2635–2654
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2013.1732

Blogs, Advertising, and Local-Market Movie Box Office Performance

Shyam Gopinath
David Eccles School of Business, University of Utah, Salt Lake City, Utah 84112,
shyam.gopinath@business.utah.edu

Pradeep K. Chintagunta
Booth School of Business, University of Chicago, Chicago, Illinois 60637,
pradeep.chintagunta@chicagogsb.edu

Sriram Venkataraman
Kenan-Flagler Business School, University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina 27599, sriraman_venkataraman@kenan-flagler.unc.edu

We measure the effects of pre- and postrelease blog volume, blog valence, and advertising on the performance of 75 movies in 208 geographic markets in the United States. We attribute the variation in blog effects across markets to differences in demographic characteristics of markets combined with differences across demographic groups in their access and exposure to blogs as well as their responsiveness conditional on access. We study the effects of prerelease factors on opening day box office performance and of pre- and postrelease factors on box office performance one month after release. Our estimation accounts for confounding factors in the measurement of these effects via the use of instrumental variables. We find considerable heterogeneity in the effects across consumer- and firm-generated media and across geographic markets, with gender, income, race, and age driving across-market differences. Release day performance is impacted most by prerelease blog volume and advertising, whereas postrelease performance is influenced by postrelease blog valence and advertising. Across markets, there is more variance in advertising and blog valence (postrelease) elasticities than there is in blog volume (prerelease) elasticities. We identify the top 20 markets in terms of their elasticities to each of these three instruments. Further, we classify markets in terms of their sensitivities across these three instruments to identify the most sensitive markets that studios can target with their limited release strategies. Finally, we characterize the extent to which studios could have improved their limited release strategies by identifying the overlap between the actual release markets and the most responsive ones. We find that at the time of first-release studios cover only 53% of the most responsive advertising markets and 44% of the most responsive markets to prerelease blog volume in their limited release strategies, implying considerable room for improvement if these were the only metrics to assess those strategies.

Key words: consumer-generated media; social media; online word of mouth; blogs; advertising; motion pictures; endogeneity; instrumental variables
History: Received September 21, 2010; accepted February 8, 2013, by Gérard P. Cachon, marketing. Published online in *Articles in Advance* August 19, 2013.

Introduction

An emerging literature in marketing, information systems, and computer science tries to understand the impact of consumer-generated word-of-mouth (WOM) on demand. These studies rely on content from online social networking sites, blogs, newsgroups, etc., as proxies for consumer-generated WOM (Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Liu 2006, Mishne and Glance 2006, Liu et al. 2007, Dhar and Chang 2009, Trusov et al. 2009, Sadikov et al. 2009, Chintagunta et al. 2010, Chen et al. 2011, Moe and Trusov 2011, Dewan and Ramaprasad 2012, Onishi and Manchanda 2012, Stephen and Galak 2012) and demonstrate the impact of these factors on outcomes such as sales. Some studies contrast the effect of online consumer-generated content with traditional media, such as newspaper coverage, television content or advertising (Chintagunta et al. 2010, Onishi and Manchanda 2012, Stephen and Galak 2012).

The main objective of this study is to understand differences across geographic markets (designated market areas (DMAs)) in the relative impacts of nationwide consumer-generated movie blog volume and valence (a positive or a negative sentiment) and of firm-generated advertising on box office performance of a new movie. The literature on measuring the effects of these factors on sales, or more specifically on movie performance, has looked at the aggregate impact at the national level largely because of the lack of more granular data. Previous research in other contexts has shown that there exists considerable heterogeneity across geographic markets in their responsiveness to marketing mix; for example, Dubé and Manchanda (2005) and Chintagunta et al. (2010) find that advertising effects vary across geographic market clusters and DMAs, respectively. However, the literature on understanding geographic differences in the effects of online word-of-mouth in general, and blogs specifically, is limited. As we discuss later, understanding these differences could help studios make more informed decisions on movie rollout and advertising.

Why would local geographic markets differ in the impact of blogs in these markets? We hypothesize that demographic differences across markets can potentially explain differences in the effects of blogs via their impacts on three sets of factors: (i) access and usage of Internet content, (ii) exposure to blogs, and (iii) responsiveness conditional on exposure. Without Internet access, consumers will not be able to access blogs. There is an influential stream of literature that has looked at the digital divide—differences in Internet access across different demographic groups. Research companies like Pew Research (Lenhart and Fox 2006) find that Internet adoption varies by age, gender, education, and race. Academic research reflects such findings; see, for example, Hoffman and Novak (2000) and Goldfarb and Prince (2008). The latter also discusses a different aspect—usage conditional on access. Consistent with other research (e.g., Sinai and Waldfogel 2004), they find high-income, educated people to be more likely to have adopted the Internet; however, conditional on adoption, low-income, less educated people spend more time online. Thus, one way in which the effects of blogs can vary across markets due to demographic differences across them is the relationship between demographic variables and access to and usage of the Internet.

Even if all households had access to and used the Internet, not all of them would be exposed to blogs. In a survey of 1,000 Internet users in the United States from 2004 to 2006, Edelman (2007) found that 27% of Internet users read blogs (mostly in political, technology, and entertainment domains). Blog readers skew to a younger demographic with 18- to 24-year-olds being the biggest consumers followed by 45- to 54-year-olds, then by 25- to 34-year-olds and 55- to 64-year-olds, and finally by 35- to 44-year-olds and those over age 65. Forrester Research (2006) found that 33% of adults online read blogs: 59% of adults in the 18–21 age group and 54% in the 22–26 age group read blogs, with much smaller percentages from higher age groups contributing to the readership. In a survey of 17,159 blog readers (not a random but a self-selected sample) conducted during May 17–24, 2005, *blogads*, a blog advertising company, found that the age group contributing the most to blog readership was the 31–40 age group (29.4%), followed by the 41–50 age group (23.1%), and then the 25–30 age group (16.8%). Further, 79% of the readers were male and 22% had incomes in the $60,000–$90,000 range. Research by Lenhart and Fox (2006), Raine (2004), and Copeland (2004) also provides insights into the demographic makeup of blog readers. Collectively, these studies reveal that there is variation across demographic groups in their exposure to blogs; however, there is some disagreement in the exact nature of the variation across studies. For our purposes here, we note that there is likely to be variation across geographic areas in blog readership to the extent that these areas differ in their demographic makeups.

The third factor through which demographic differences across geographic markets can manifest themselves as differential blog sensitivities is responsiveness conditional on exposure. Based on the peer-influence literature, Aral and Walker (2012) use a randomized experiment on Facebook to identify profiles of users who are influential (those who can influence) and those who are susceptible (those who can be influenced) to social influence. The main variable of interest here is the volume of messages received. They find that younger users are more susceptible to influence than are older users. Men are more influential than women, whereas women are less susceptible to influence than men. Further, married individuals are the least susceptible to influence in the decision to adopt the product offered. According to the Edelman (2007) study, 28% of blog readers have responded to a blog's "call to action," although the study does not describe the demographic makeup of those influenced by blogs. Johnson and Kaye (2004) discuss the greater credibility attached by blog readers to information from blogs relative to information from traditional sources, although they find that these latter sources are viewed as being credible as well. Banning and Sweetser (2007) find, however, that there are no differences in the perceived effects of blogs relative to those for more traditional media. Neither study looks at the variation in the effects across demographic groups. There is some research that directly looks at geographic differences in consumers' preferences and sensitivities to marketing activities; for example, Jank and Kannan (2005) show how consumers across various geographic regions exhibit differences in their preferences and price sensitivities for physical (print) as opposed to electronic (PDF) goods.

In the advertising domain, based on self-reported information, Shavitt et al. (1998) find that males, younger people, lower income, and lower education and nonwhite consumers hold a more favorable attitude toward ads than other demographic groups do. On the other hand, researchers such as Gleser et al. (1959) and Aronson (1972), among others, make the point that women are more likely to be influenced by advertising. Holbrook (1978) suggests that the more educated are less likely to be influenced by advertising because they will be more evaluative and engage in more information processing. Thus, although it is reasonable to assume that responsiveness is likely to vary across demographic groups and hence across geographic markets, more research is required to make that linkage.

Taken together, the above discussion suggests that the presence of blogs online can influence box office performance by how they are related to various demographic factors. Because different geographic markets have different demographic profiles, this translates into geographic markets responding differentially to online content. In this study, based on the literature cited above, we focus on age, gender, population, race, and income, which could account for heterogeneity in the effects of blogs across markets. Since we do not directly observe the factors—access and usage, exposure and response conditional on exposure—we instead quantify the extent to which the heterogeneous blog effects across geographic markets, if any, can be attributed to these five variables.

A unique feature of the movie industry is that a significant portion of sales of movies occurs at the time of release; for example, the opening weekend gross for *Shrek 2* was 24.7% of the overall domestic box office receipts. Consequently, it is of interest to understand the impact of blogs and advertising on opening day–opening market performance separately from their effects on the subsequent performance of the movie. If the focus is on opening box office performance, then the key box office influencers of interest are *prerelease* blogs and advertising, where prerelease blogs are based on *expectations* of a movie's quality. Subsequent box office performance of the movie will depend on the prerelease blogs and advertising and also on postrelease blogs and advertising. Further, postrelease, one also needs to consider the impact of other user-generated content such as online user reviews.

We assemble a unique data set of 75 movies (accounting for 90% of box office revenue in our data) released in 2004 for our empirical analysis. The data contain daily market-level box office information for each movie for the one-month period from the date of first release, along with movie and market characteristics. Blog information, defined as the number of blogs about the movie in the month preceding the movie's release and in one month postrelease, as well as the proportion of these blogs that is positively valenced, was manually collected and coded. We also obtained online user reviews from Yahoo! user reviews for the one month, postlaunch period. In addition, we observe advertising expenses for a movie from the inception of advertising until one month postrelease. Since we are interested in the opening box office performance of the movie and how it is driven by prerelease advertising and blogs, one dependent measure is the opening day–opening market box office performance of these movies. For postrelease box office performance, we use the performance of the movie on the day that is four weeks after the movie is released (for the same first-release markets).

A key challenge with estimating the effects of online content and offline advertising on box office performance is the presence of unobservable factors that are not explicitly accounted for, e.g., political events, death/suicide of actors, etc., that would influence box office performance and be correlated with online content. Researchers have used instrumental variables to deal with this issue (e.g., Chintagunta et al. 2010). Our choice of dependent measures at two discrete time points allows us to deal with potential endogeneity issues that might arise via the use of plausible instruments for blogs and advertising. On the other hand, although using the data on the entire time series of box office performance from launch provides more information, it is a challenge to assemble corresponding time-varying instruments. The problem could be further exacerbated in the presence of serial correlation in the unobservables.

Our paper contributes to the literature on understanding the impact of firm-generated media, i.e., advertising and consumer-generated media (namely, blogs and user reviews) on the performance of a new product. We characterize this impact on performance both at the time of launch as well as in a postlaunch period. Importantly, we seek to understand how these effects vary across geographic markets that have different demographic profiles. Understanding this variation will enable studios to better assess their launch market strategies, especially for movies that do not get the widest release. For example, our analysis reveals that studios do not necessarily target the most responsive markets in terms of advertising and blogs when they pick their first-release markets for the movies.[1]

[1] Our study is distinct from that of Chintagunta et al. (2010) on several dimensions. First, our primary interest is in blogs that exist both before and after the introduction of a new product. Contrasting the effects of user-generated content at the two stages of the product's life cycle generalizes the postrelease analysis of Chintagunta et al. (2010). Second, that study looked at the effects of only one aspect of content—user reviews—which are available only after a product's release. By contrast, we generalize this by

We acknowledge that the rapid deployment of broadband and the access to mobile platforms since the time period of our data has had a definite impact on the availability and access to a variety of consumer-generated media beyond the ones we consider. This potentially limits our ability to translate our results to the present. Nevertheless, the variation in effects across geographic markets is likely to persist even with more recent data.

Data Description

Our data are for 75 movies released in 2004 for each of the DMAs that the movies are first released in. Table 1 gives the variable definitions along with their descriptions and key summary statistics.

(i) *Box office data*: Gross receipts for a title in a market is our dependent variable ($GROSS^t$, $t = PRE$, $POST$). *PRE* refers to opening day, whereas *POST* refers to the date four weeks after release. We focus only on that subset of DMAs where the movie is released the first time it is ever released in the U.S. market. We have 9,382 observations (movies $*$ markets) in the *PRE* and 6,923 observations in the *POST* data. Market-specific theater counts (*THEATERS, THEATERS_POST*) are the numbers of theaters in which the movie is playing on the days our dependent variables are measured. We use cumulative national advertising expenditures to date as our advertising variable ($ADVERTISING^t$, $t = PRE, POST$).

(ii) *Blog measures and user ratings*: Google blogs (see http://blogsearch.google.com/) and Technorati (see http://technorati.com/) are two widely used blog search engines. For this research, the blog data are collected from Google blogs because we found that Google blogs performs slightly better than Technorati with respect to the number of blogs for movies across different genres. Also, one might expect a great degree of overlap between the blogs from these two search engines. So for each movie, all blogs were collected for the one-month prerelease and one-month postrelease periods. The blogs were then classified, similar to prior research (e.g., Godes and Mayzlin 2004, Liu 2006), as either positive or negative by human coders.[2] This allows us to make sure that we are screening out irrelevant blogs (e.g., blogs for *The Alamo*, which was released in the 1960s). We use two measures to capture the impact of blogs—blog volume (*BLOGVOL_PRE, BLOGVOL_POST*), which is the number of movie-related blogs generated, and blog valence (sentiment), which is the ratio of the numbers of positive and negative blogs (*BLOGPROP_PRE, BLOGPROP_POST*). Examples of the nature-of-blog conversations are given in Figure 1. There are some distinguishing features for these blogs. Blogs 1 and 2 mention the movie as a part of a *consideration set*. Blog 1 mentions the *reason* for wanting to see the movie, whereas blogs 2 and 3 do not. Blogs 3 and 4 indicate possible *peer influence*. Finally, blog 5 indicates that the sentiment expressed need not always be positive. It is important to note that all of these blogs are publicly available and so are accessible to all potential moviegoers. We also recognize (based on the above description) that the blogs can be further categorized beyond the two variables we use, but we leave that issue for future research.

User ratings are only available after the release of a movie and hence only relevant for the postrelease analysis. While several sites publish users' ratings of movies, previous research has used those from Yahoo! Movies (see Chintagunta et al. 2010, Duan et al. 2008, Dellarocas et al. 2007), which we do as well. We compute the volume of reviews (*RATINGVOL*) and the valence of reviews that are measured on a scale from 1 to 13 (*RATINGVAL*). Chintagunta et al. (2010) also use a similar procedure; they find that a higher rating leads to a larger box office postrelease in *non-first-release markets*.

(iii) *Movie characteristics* (*Z*), *market characteristics* (*D*), *competition, and other factors*: We collect data on various movie characteristics (denoted collectively by *Z*), such as genre (*ACTION, DRAMA, COMEDY,* etc.); stars (*STARS*—the star power data are from *The Hollywood Reporter*, specified as the number of cast/crew members of the movie who are featured in the top 100 entertainers list for that year); Motion Picture Association of America (MPAA) rating (*PG*, *PG13*, etc.); and critics' reviews (*CRITICS*—the metascore, i.e., weighted critics' score from metacritic.com; this score ranges from 0 to 100 and is a weighted average of critics' scores from about 30 top publications and critics and has been used by previous research; Basuroy et al. 2003). These data are obtained from Nielsen EDI. Other than the *CRITICS* variable, the other variables only influence the average box office performance across markets.

Next, we look at market characteristics (collectively denoted as *D*). Recall that the objective in including these variables is to account for the roles of access, exposure conditional on access, and response conditional on exposure. For the access variable, we have

looking at blogs before launch, and both blogs and user reviews after launch. Finally, the two studies look at complementary sets of markets. Chintagunta et al. (2010) only look at sequentially released markets *after* the markets of initial release, whereas we focus only on first-release markets.

[2] Demographic information on the bloggers is not available or codified. Most of the blogs were strongly positive or strongly negative. However, there were some instances where blog postings were mixed, i.e., contained both positive and negative sentiment. In such scenarios, the postings were classified as either positive or negative depending on the overall sentiment of the posting.

Table 1 Descriptive Statistics

Variable	Description	Min	Max	Mean	Std. dev.
$OPENINGGROSS_{ij}$	Total opening gross for movie i in market j (in dollars)	11.50	2,152,649.57	6,973.07	26,434.85
$GROSS\_POST_{ij}$	Total gross for movie i in market j on the 30th day after initial release (in dollars)	4.00	281,646.60	3,189.32	8,854.43
$THEATERS_{ij}$	Total number of theaters in which the movie i is released in market j	1.00	149.00	9.08	13.87
$THEATERS\_POST_{ij}$	Total number of theaters on the 30th day after initial release for movie i in market j	1.00	137.00	6.10	9.65
$COMPSTARS_{ij}$	Average star power of competing movies on the opening day for movie i in market j (0–6)	0.00	2.50	1.66	0.64
$COMPSTARS\_POST_{ij}$	Average star power of competing movies on the 30th day after initial release for movie i in market j (0–6)	0.00	2.08	1.84	0.11
$COMPREVIEW_{ij}$	Average critic score of competing movies on the opening day for movie i in market j (0–100)	9.00	84.00	47.19	12.31
$COMPREVIEW\_POST_{ij}$	Average critic score of competing movies on the 30th day after initial release for movie i in market j (0–100)	23.75	60.28	48.27	3.76
$COMPAGE_{ij}$	Average age of competing movies on the opening day for movie i in market j	1.00	231.00	36.13	34.71
$COMPAGE\_POST_{ij}$	Average age of competing movies on the 30th day after initial release for movie i in market j	3.20	82.12	38.72	14.15
$BLOGVOL_{i}$	Number of blogs in the prerelease month for movie i	2.00	437	111.20	102.63
$BLOGVOL\_POST_{i}$	Number of blogs in the postrelease month for movie i	3.00	610	182.43	149.76
$BLOGPROP_{i}$	Ratio of number of positive blogs to the number of negative blogs in the prerelease month for movie i	1.56	16.20	3.24	5.36
$BLOGPROP\_POST_{i}$	Ratio of number of positive blogs to the number of negative blogs in the postrelease month for movie i	1.37	14.18	3.08	4.89
$RATINGVOL_{i}$	Number of user ratings (on Yahoo! Movies) in the postrelease month for movie i	1.00	1,920	197.12	300.45
$RATINGVAL_{i}$	Average valence of use ratings (on Yahoo! Movies) in the postrelease month for movie i	4.45	12.34	9.57	1.56
$ADVERTISING_{i}$	Advertising in the prerelease month for movie i (in 1,000 dollars)	251.00	34,025.60	13,751.87	1.97
$ADVERTISING\_POST_{i}$	Advertising expenditure in the postrelease month for movie i (in 1,000 dollars)	148.41	32,859.53	8,103.08	1.69
$STARS_{i}$	Star power of movie i (0–6)	0.00	4.00	1.97	0.41
$CRITICS_{i}$	Critic score for movie i (0–100)	9.00	84.00	47.07	16.06
$BROAD_{j}$	Percentage of households with broadband connection in market j	1.95	53.29	23.59	9.34
$WOMEN_{j}$	Percentage of women in market j	49.06	54.44	51.59	0.83
$YOUNG_{j}$	Percentage of population who are less than 34 years old in market j	28.97	48.88	37.66	3.13
POP_{j}	Log(Population in market j)	10.25	16.53	13.18	1.10
INC_{j}	Log(Household Income in market j)	10.12	11.02	10.54	0.17
$CAUCASIAN_{j}$	Percentage of Caucasians in market j	6.99	99.20	80.98	15.95
$ASIAHISP_{j}$	Percentage of Asians and Hispanics in market j	0.49	92.60	8.93	13.81
$COMEDY_{i}$	Dummy variable for comedy genre	0.00	1.00	0.27	0.46
$ACTION_{i}$	Dummy variable for action genre	0.00	1.00	0.33	0.47
$DRAMA_{i}$	Dummy variable for drama genre	0.00	1.00	0.31	0.42
PG_{i}	Dummy variable for PG MPAA rating	0.00	1.00	0.24	0.43
$PG13_{i}$	Dummy variable for PG13 MPAA rating	0.00	1.00	0.47	0.50
$HOLIDAY_{i}$	Dummy variable indicating whether movie i was playing on a holiday	0.00	1.00	0.25	0.44
AGE_{it}	Age of movie i in week t (in weeks)	0.00	4.00	1.87	1.05
Number of observations—Preanalysis	9,382				
Number of observations—Postanalysis	6,923				
Number of movies	75				
Number of markets (DMAs)	208				

Figure 1 Examples of Blog Postings

Sarah (**redandblack13**) wrote,
@ 2004-04-04 19:38:00

Everyone take me to the following movies during the month of april
—the prince and me
—13 going on 30
—mean girls

I know, all the classic teen movies . . . mean girls has lindsay lohan . . . why is she in all the good movies. now it seems like i'm going because shes in it. oh well. you all should know thats not why i'm going . . . .

Ashley (**interceptor_cjs**) wrote,
@ 2004-05-04 22:57:00
My Movie List
Movies that I think would like to see
1. Van Helsing (I think this one looks ok)
2. Troy (this movie look great and I love the music)
3. Shrek 2 (I just love shrek)
4. Raising Helen (looks like a cute movie)

Phillip (**ugafan07**) wrote,
@ 2004-01-21 23:04:00
I'm **Definately** going to see *The Butterfly Effect* this weekend!!!!!
Who's with me?

Nick (**nick_a_bear**) wrote,
@ 2004-10-21 20:36:00

OH THE GRUDGE COMES OUT TOMORROW AND ME AND TREVOR AND WHOEVER ELSE WHO WANTS TO COME IS INVITED TO GO SEE IT WITH US. I CANT WAIT THAT MOVIE LOOKS AWESOME ESPECIALLY THAT LITTLE MEOWING BOY. CALL ME OR WHATEVER IF YOU WANT TO COME. YOU KNOW THE NUMBER. AND IF YOU DONT THAT MEANS I PROBLY DONT WANT YOU TO HAVE IT ANYWAY. JUST KIDDING. BUT SERIOUSLY THOUGH. BYE.

D (**tawang**) wrote,
@ 2004-07-16 00:10:00

Who, Robot?
If the ads and trailers and such for "I, Robot" aren't insulting enough, the credit "Suggested by the book by Isaac Asimov" is ridiculous. All they've done is borrow a couple of character names and the infamous "Three Laws of Robotics," and they've grafted them onto a separate script with similar themes that the studio bought a few years ago.

---

some information that is directly related to access—the proportion of people in a geographic market who have access to broadband (*BROAD*). As noted in the introduction, we also look at the following variables: gender, operationalized as the proportion of women in a market (*WOMEN*); age, operationalized as the proportion of the population in a market under age 34 (*YOUNG*); income, operationalized as the logarithm of the average market income (*INC*); and population, operationalized as the logarithm of the market's population (*POP*). We also include two variables related to race: the proportions of Caucasian (*WHITE*) and Asian and Hispanic (*ASIAHISP*) members in a market. The market-level broadband-penetration data are from the Current Population Survey (2003)—a household survey conducted jointly by the Census Bureau and the Bureau of Labor Statistics (http://www.census.gov/cps/). The other variables are obtained from the U.S. Census Bureau.

As noted by Ainslie et al. (2005), it is important to control for competitive considerations when studying box office performance. In our case, since we look at only a subset of movies, we cannot use the logit model proposed in that study. Hence, we need to control for competitive effects via the inclusion of additional variables. Similar to Chintagunta et al. (2010), we include three variables to control for competitive effects due to other movies playing in the market on the day of interest. The first variable is the average "star power" of *all* the competing movies (not just the ones in our set). These variables based on the same measure of star power as in *STARS* are *COMPSTARS* and *COMPSTARS_POST*. The second competition variable is the average of critics' scores for competing movies. We compute the average of the *CRITICS* score across all competing movies (not just the ones in our set) in a given market (*COMPREVIEW*, *COMPREVIEW_POST*). We expect a higher value of both variables to lower a movie's box office performance in a market. The third variable we use is the average age of all competing movies on the day of interest (*COMPAGE* and *COMPAGE_POST*). The lower this average, the smaller is the predicted gross of the focal film.[3]

We also gather information on other factors such as weather (*WEATHER, WEATHER_POST*) on the days of interest. For this, we record the amount of rain and snow in that market on the release and postlaunch dates. Interestingly, these variables did not have a significant impact on performance. Additionally, we also include variables such as seasonal factors (*SEASON*), holidays (*HOLIDAY*), etc. Thus, we have a comprehensive set of data to address the questions that are of interest to us.

Correlations Among Blog Measures and Advertising

Since a potential driver of blogs is studio advertising, we look at how these variables are correlated in the prerelease (postrelease) periods. We find that the correlation between advertising and blog volume is 0.24 (0.20) and that between advertising and blog valence is 0.23 (0.26). Finally, the correlation between the two blog measures is 0.22 (0.31). These correlations indicate that there is idiosyncratic information contained in the blog measures and that controlling for advertising, although important, does not preclude the importance of the blog measures in explaining box office performance. This appears to be different from the Japanese market analyzed by Onishi and Manchanda (2012).

Methodology

Model

Our main variables of interest that influence box office performance—advertising, blog volume, and blog valence—only vary at the movie level, and we seek to understand their differential effects across markets. Our econometric model is therefore specified and estimated in two stages. In the first stage, we exploit differences in movie performance across markets to uncover the *differences* in the effects of consumer- and firm-generated media across markets (along with movie fixed effects). In the second stage, we project the movie fixed effects onto the advertising and blog measures to recover the *means* of their effects across markets. Such a two-stage approach facilitates understanding the variation of effects across markets while accommodating unobservable factors that might be correlated with the advertising and blog measures.

[3] We also investigated including a competitive performance measure—the average box office performance of other movies of the same genre in the market during the previous month. In the robustness section, we show that the effect of the variable is not statistically significantly different from zero; further, in the presence of serial correlation in our dependent variable, inclusion of the box office performance of rival movies can be problematic.

Our unit of analysis is a movie (i)–market (j) combination (ij). Two sets of analyses are conducted—the first is right after the release of the movie ($t = PRE$), and the other four weeks after the movie's release ($t = POST$). The revenue (*GROSS*) for movie i in market j is written as

$$\begin{aligned}
&\log(GROSS_{ij}^t)\\
&\quad = \alpha_i^t + \rho_j^t + \Gamma_1^t \log(BLOGVOL_PRE_i) \times D_j\\
&\qquad + \Gamma_2^t \log(BLOGPROP_PRE_i) \times D_j\\
&\qquad + \Gamma_3^t \log(ADVERTISING_i^t) \times D_j + I^{POST}\\
&\qquad \times \Gamma_4 \log(BLOGVOL_POST_i) \times D_j + I^{POST}\\
&\qquad \times \Gamma_5 \log(BLOGPROP_POST_i) \times D_j + I^{POST}\\
&\qquad \times \Gamma_6 \log(RATINGVOL_i) \times D_j + I^{POST}\\
&\qquad \times \Gamma_7 \log(RATINGVAL_i) \times D_j\\
&\qquad + \Gamma_8^t \log(CRITICS_i) \times D_j + I^{PRE}\\
&\qquad \times \beta_1^{PRE} \log(THEATERS_{ij}) + I^{POST}\\
&\qquad \times \beta_1^{POST} \log(THEATERS_POST_{ij}) + I^{PRE}\\
&\qquad \times \beta_2^{PRE}\, COMPSTARS_{ij} + I^{POST}\\
&\qquad \times \beta_2^{POST}\, COMPSTARS_POST_{ij} + I^{PRE}\\
&\qquad \times \beta_3^{PRE}\, COMPREVIEW_{ij} + I^{POST}\\
&\qquad \times \beta_3^{POST}\, COMPREVIEW_POST_{ij} + I^{PRE}\\
&\qquad \times \beta_4^{PRE}\, COMPAGE_{ij} + I^{POST}\\
&\qquad \times \beta_4^{POST}\, COMPAGE_POST_{ij}\\
&\qquad + \Gamma_{9j}^t\, GENRE_i + \Gamma_{10i}^t DEMOS_j + I^{PRE}\\
&\qquad \times \beta_5^{PRE}\, WEATHER_{ij} + I^{POST}\\
&\qquad \times \beta_5^{POST}\, WEATHER_POST_{ij} + e_{ij}^t
\end{aligned}$$

where

$$\begin{aligned}
D_j = \big(&BROAD_j,\ WOMEN_j,\ YOUNG_j,\ POP_j,\\
&INC_j,\ WHITE_j,\ ASIAHISP_j\big)'. \qquad (1)
\end{aligned}$$

The above equation implies that box office performance at the time of release ($t = PRE$) depends on prerelease blogs, advertising, critics' reviews, number of theaters, competitor variables, and weather. Postrelease ($t = POST$), it depends on pre- and postrelease blogs, user ratings, total advertising to date, number of theaters, competitor variables, and weather. Thus, we explicitly separate out the pre- and postrelease online WOM in the postrelease analysis. The t superscript on the parameters implies that different effects are estimated for the opening versus the postrelease time period. In Equation (1), the overall intercept is denoted by α_i, which denotes the movie fixed effect; ρ_j denotes the market fixed effect; *BLOGVOL_PRE* (*BLOGVOL_POST*) and *BLOPGPROP_PRE* (*BLOGPROP_POST*) denote the volume of blogs and valence of blogs for the one month prior to (and after) release. *ADVERTISING* refers to prerelease advertising in the *PRE* period and prerelease plus one-month-postrelease advertising in the *POST* period; *RATINGVOL* and *RATINGVAL* refer to the volume and valence of user ratings from Yahoo! in the one-month-period postrelease; *CRITICS* refers to critic scores garnered by movie i; and D_j refers to the set of demographic characteristics of market j. Thus, Γ_1^t, Γ_2^t, Γ_3^t, Γ_4, Γ_5, Γ_6, Γ_7, Γ_8^t refer to the *interaction* effects of the blog, advertising, ratings, and critics variables with the demographic variables on the performance of movie i in market j. Given the inclusion of an overall movie intercept, this means that α_i will contain the *main* effects of the variables *BLOGVOL_PRE, BLOGPROP_PRE, ADVERTISING, BLOGVOL_POST, BLOGPROP_POST, RATING_VOL, RATING_VAL,* and *CRITICS* across markets. In the estimation, we operationalize D_j as the deviations of the demographic levels for each market from the averages across markets. This way, the main effects contained in α_i will also absorb the mean effects of demographics across markets. Note that instead of interacting the various box office drivers with the variables in D_j, we could interact them directly with market fixed effects (see the robustness section). We prefer our current approach (i) because of its relative parsimony and (ii) because our ultimate interest is in how responsiveness varies across markets along the variables identified in D_j.

The discrete variables I^{PRE} and I^{POST} take the value 1 in the *PRE* and *POST* periods, respectively, and 0 otherwise. *COMPSTARS*, *COMPREVIEW*, *COMPAGE*, and their associated parameters reflect the effects of competition, and e_{ij} is the movie-market unobservable. In addition to movie-specific factors and market-specific factors, there could be movie-market-specific factors that could result in the error term e_{ij} being correlated across markets. For example, the studio may decide to release a movie with a certain set of characteristics in markets where the studio has knowledge of some unobserved demand drivers. To account for such factors, we include interactions between the market fixed effects and movie characteristics (e.g., genre) and between the movie fixed effects and market characteristics (e.g., demographics). These are represented by $\Gamma_{9j}GENRE_i + \Gamma_{10i}DEMOS_j$, and account for any inherent "fit" between the movie and the launch market, without which one may incorrectly attribute differential box office performance across markets to market-level response heterogeneity to consumer- and firm-generated media. The motivation for including these variables comes from our examination of the raw data where we observe, for example, that certain genres of movies tend to be released first in certain geographic markets. Stated differently, our objective in including these factors is to account for studios' endogenous release decisions that are reflected in our data; that is, our identifying assumption is that conditional on these variables, there is no error correlation across markets.[4] Next, the movie fixed effect, α_i can, in turn, be written as

$$\begin{aligned}
\alpha_i^t = {} & \alpha^t + \lambda^t Z_i + \theta_1^t \log(BLOGVOL_PRE_i) \\
& + \theta_2^t \log(BLOGPROP_PRE_i) \\
& + \theta_3^t \log(ADVERTISING_i^t) + I^{POST} \\
& \times \theta_4 \log(BLOGVOL_POST_i) + I^{POST} \\
& \times \theta_5 \log(BLOGPROP_POST_i) + I^{POST} \\
& \times \theta_6 \log(RATINGVOL_i) + I^{POST} \\
& \times \theta_7 \log(RATINGVAL_i) + \theta_8^t \log(CRITICS_i) \\
& + \eta_1 SEASON_{1i} + \eta_2 SEASON_{2i} + \eta_3 SEASON_{3i} \\
& + \eta_4 HOLIDAY_i + \zeta_i; \\
& \theta_\ell^t = \tilde{\theta}_\ell^t + \Gamma_\ell^t \bar{D}, \quad \ell = 1, 2, 3, 8; \\
& \theta_\ell = \tilde{\theta}_\ell + \Gamma_\ell \bar{D}, \quad \ell = 4, 5, 6, 7;\ t = PRE, POST;
\end{aligned} \tag{2}$$

where θ_ℓ denotes the average effect across markets of the corresponding variable on box office performance, and $\bar{D}$ is the vector of mean values of demographic variables across markets (recall that D_j in the estimation of Equation (1) is operationalized as deviations from the mean); the vector Z_i includes all movie characteristics such as stars, rating, etc. We also include dummy variables for the season in which the movie is released as well as for the release date following a holiday. Together, Equations (1) and (2) reflect the various effects of interest to us. The parameters in these equations can be estimated sequentially while accounting for potential endogeneity effects.

Endogeneity Issues vis-à-vis Equation (1). The potential source of endogeneity here is the correlation between the number of theaters in which the movie is released and the error term. As noted previously, studios may be picking markets for a movie based on some knowledge of the movie-market unobservable. We accounted for such factors via the inclusion of the interaction effects in Equation (1) between movie characteristics and market fixed effects and between

[4] Nair et al. (2010) include prescriptions of other physicians in the area around the focal physician to account for these effects. We discuss the inclusion of a similar variable in the robustness section.

market demographics and movie fixed effects. By the same token, it is possible that $THEATERS_{ij}$ is correlated with e_{ij} (Elberse and Eliashberg 2003) since studios could also pick the number of theaters to release based on knowledge of the unobservables. While the interaction effects we include in (1) could account for this effect to some extent, we also instrument for theaters as follows.

We use as an instrument for the number of theaters the *number of temporary* (*not permanent*) *closures of theaters in the market during the week in which the focal movie is released in the market* (*THEATERS_TMPCLOSURES*). Theaters close for durations from one to several weeks at a time for painting, refurbishing, etc. The variable shifts the supply of theaters available to a studio without being correlated with any unobserved aspect of a particular movie's demand. The identifying assumption is that conditional on the various fixed effects and interactions included in the analysis, the covariation of the number of theaters with temporary closures is due to theater supply shocks and not demand shocks (Wooldridge 2002). Although not detailed here, we find that around 7% of theaters in our data experienced temporary closures with no appreciable pattern of seasonality in these closures even at the market level.

Endogeneity Issues vis-à-vis Equation (2). Of more serious concern is the estimation of the parameters in Equation (2). This is a cross-sectional regression across movies. Here, the movie-specific unobservable ζ_i might be correlated with the blog or advertising measures. If such correlations exist, they will contaminate our estimates of the mean effects (across markets) of the various measures of interest and of the overall market-specific effects as well. The intuition behind any choice of instruments is to have variables that are correlated with the blog and advertising measures but uncorrelated with the movie-specific unobservable.

For the blog volume we use two instruments. The first is the *number of holidays in the one-month period before the measurement of the dependent variable* that we are using to create our *BLOGVOL_PRE* and *BLOGVOL_POST* measures. The idea is that a larger number of holidays provides a greater opportunity for potential bloggers to engage in blogging activity, thereby pushing up the number of blogs for the movies released at the end of that month.[5] The identifying assumption here is that conditional on the movie's characteristics, seasonality, etc., covariation of the volume of blogs with the number of holidays is due to the supply of time for blogging and not due to unobserved demand factors. The second instrument we use is the *number of bad weather days in the one-month period before the measurement of the dependent variable*. A larger number of such days in markets where people are more interested in movies overall (not the specific focal movie) implies greater opportunities for potential bloggers to engage in blogging activity. The identifying assumption here is similar to that for the holidays instruments.

Finding instruments for the valence measures is trickier—we need variables that influence the proportion of bloggers liking or disliking a particular movie without those variables directly influencing box office performance. The instruments are based on the *match between the demographic profile of the lead cast of the movie and the demographic profile of the universe of bloggers*. The intuition is that bloggers are likely to express a more positive sentiment when there is a match between their characteristics and those of the movie. The bloggers are described using three demographic measures—race (white, African American, and others), age ("young" is less than age 34 and "old" is at least age 34), and gender (male and female). Similarly, for each movie we have seven indicator variables (three for race, two for age, and two for gender). The indicator variables take a value of 1 if the lead actor is of the corresponding demographic classification; for example, for movie i with a young white male lead, we would have the following value for the indicator variables: white, 1; African American, 0; others, 0; young, 1; old, 0; male, 1; female, 0. Using the demographic measures (Raine 2004) and the movie level demographic indicators, the instrument *BPROP_DEMOG* is as follows:

$$BPROP\_DEMOG_i = \sum_{s=1}^{S} Ind_{is} * DemogProp_s, \qquad (3)$$

where Ind_{is} is an indicator variable that takes a value of 1 (0 otherwise) if the lead actor in movie i has characteristic s, and $DemogProp_s$ is the proportion of the blogging population with characteristic s.

A downside to the above measure is that it is time invariant; we do not have separate instruments for *BLOGPROP_PRE* and *BLOGPROP_POST* if we wish to include them separately in the analysis as suggested by Equation (1). One option is to combine both pre- and postmeasures when conducting the *POST* analysis. In our case, the *PRE* analysis revealed that prerelease blog valence did not impact box office performance (either with or without instrumenting for this variable). Consequently, for the postrelease analysis, we only instrumented for *BLOGPROP_POST* using the instrument in Equation (3).

Our instrument for advertising is a movie's *production budget*. The production budget drives that

[5] We are only assuming that people are *more likely* to blog when they are off work (not that people *only* blog when they are off work).

part of a movie's advertising budget that is set by studios before the movie is made. The actual advertising budget, determined after the movie is completed, is adjusted by studios with some knowledge of the movie-specific unobservable. The covariation between the production budget and observed advertising expenditures can be attributed to factors other than the unobservable characteristics of movies. This is our identifying assumption.

In Table 2 we present the results of the regression of the endogenous variables on the instruments and other exogenous variables and fixed effects. Table 2(a) shows that our instrument for theaters is strongly correlated with the endogenous variable—the more temporary closures, the fewer the number of theaters in which the movie is released. Furthermore, the F-statistic exceeds 10, the rule of thumb suggested by Stock and Watson (2010) for single endogenous variable regressions. Thus, we have some assurance that this is not a weak instrument.

Table 2(b) shows the regressions of the endogenous variables in Equation (2) on the set of instruments and exogenous variables. Once again, it appears that (i) the instruments are acting in the directions predicted (the greater the number of holidays and bad weather days in the 30 days preceding a movie's release and postrelease, the higher the volume of blogs at the time of release and one month after release; the better the demographic match between bloggers and the movie's stars, the greater the proportion of positively valenced pre- and postrelease blogs; and the higher the production budget, the greater the advertising budget); (ii) the instruments do not seem to be "weak"; and (iii) the test for overidentifying restrictions has a p-value of 0.412 (pre) and 0.344 (post) that well exceeds 0.1, implying that the instruments satisfy the required orthogonality conditions. All of these suggest that we have a reasonable set of instruments for our analysis. The table also shows that blog volume is correlated with factors such as stars, genre, and season—perhaps indicative of the factors that generate blog creation.

Empirical Results

Movie-Market-Level Factors and Drivers of Market-Level Responsiveness to Blog and Advertising Measures

We focus on the second-stage coefficients from the movie-market-level equation (1). The results are in Tables 3(a)–3(c). Table 3(a) shows the effects of movie-market factors in driving opening as well as one-month postrelease box office performance in the first-release markets. Note that these regressions control for movie as well as market fixed effects and potential endogeneity in the *THEATERS* variable. Further, we have interactions between the fixed effects and market demographics and genre, respectively. After controlling for these factors, Table 3(a) shows that the number of theaters in which a movie is released in a market has a statistically significant effect on opening box office. This finding is consistent with prior research in the area (e.g., Elberse and Eliashberg 2003). In marginal effect terms, increasing the number of theaters by 10% in each of these markets increases opening box office by 6.12% (opening day) and 11.84% (one month after release) on average. The second statistically significant effect in Table 3(a) is a competitive effect. The higher the critics' ratings of the movies playing in the opening markets at the

Table 2(a) **First-Stage Estimates for *THEATERS***

DV=LOG(*THEATERS*)	Release day ($t=PRE$)	Postrelease ($t=POST$)
Variable	Estimate (std. error)	Estimate (std. error)
LOG(*THEATERS_TMPCLOSURES*) (instrument for *THEATERS*)	-0.048^{***} (0.011)	-0.075^{***} (0.018)
COMPSTARS	−0.257 (0.279)	−0.145 (0.298)
COMPREVIEW	-0.033^{**} (0.014)	−0.004 (0.014)
COMPAGE	0.000 (0.001)	0.000 (0.0001)
R^2		
Overall	0.93	0.96
Instrument	0.14	0.12
F-test		
Overall	334.74^{***}	635.45^{***}
Instrument	$1{,}289.52^{***}$	$1{,}107.94^{***}$

Note. Fixed effects are not reported.
$^{**}p<0.05$; $^{***}p<0.01$.

Table 2(b) First-Stage Estimates for *BLOGVOL*, *BLOGPROP*, and *ADVERTISING*

	Release day ($t = PRE$)			Postrelease ($t = POST$)		
	BLOGVOL	*BLOGPROP*	*ADVERTISING*	*BLOGVOL*	*BLOGPROP*	*ADVERTISING*
BLOGVOL INSTRUMENT 1: *Number of holidays in the one-month period before the measurement of the dependent variable*	0.007*** (0.002)	0.001 (0.002)	0.003 (0.002)	0.022*** (0.007)	−0.002 (0.002)	0.146 (0.185)
BLOGVOL INSTRUMENT 2: *Number of the measurement of the dependent variable*	0.528*** (0.154)	−0.122 (0.086)	−0.015 (0.078)	0.196*** (0.063)	−0.314 (0.236)	0.301 (0.197)
BLOGPROP INSTRUMENT: *Match between the demographic profile of the lead cast of the movie and the demographic profile of the universe of bloggers*	0.002 (0.002)	0.005*** (0.001)	−0.001 (0.001)	0.229 (0.421)	0.003*** (0.001)	−0.451 (0.348)
ADVERTISING INSTRUMENT: *Production budget*	0.005 (0.004)	0.003 (0.002)	0.456*** (0.095)	−0.001 (0.009)	−0.008 (0.007)	0.657*** (0.254)
STARS	0.302* (0.168)	0.084 (0.152)	0.227* (0.124)	0.163 (0.154)	0.077 (0.168)	0.081 (0.132)
CRITICS	0.007 (0.005)	0.023*** (0.005)	0.005 (0.004)	0.018** (0.008)	0.019*** (0.006)	0.005 (0.004)
PG	−0.338 (0.246)	0.058 (0.236)	0.162 (0.191)	−0.155 (0.280)	−0.251 (0.272)	0.415** (0.224)
PG13	−0.182 (0.219)	−0.079 (0.205)	0.005 (0.173)	−0.276 (0.224)	−0.026 (0.228)	0.169 (0.176)
ACTION	−0.631* (0.349)	0.561* (0.312)	0.095 (0.262)	−0.587* (0.320)	0.127 (0.358)	0.105 (0.251)
COMEDY	−0.354 (0.360)	−0.586* (0.311)	0.332 (0.264)	−0.381 (0.288)	−0.546* (0.304)	0.384 (0.298)
DRAMA	−0.218 (0.349)	−0.265 (0.314)	0.093 (0.261)	−0.015 (0.421)	−0.230 (0.354)	0.096 (0.285)
HOLIDAY	0.252 (0.214)	−0.324 (0.278)	0.220 (0.205)	0.198 (0.176)	−0.214 (0.191)	0.172 (0.154)
SEASON 1	0.402 (0.286)	0.482* (0.250)	0.179 (0.208)	0.389 (0.291)	0.414* (0.225)	0.225 (0.348)
SEASON 2	0.722** (0.306)	0.589** (0.279)	0.430* (0.231)	0.220 (0.387)	0.471* (0.269)	0.059 (0.253)
SEASON 3	0.174 (0.272)	0.410 (0.254)	0.272 (0.210)	0.352 (0.270)	0.547** (0.254)	0.071 (0.128)
R^2						
Overall	0.55	0.32	0.29	0.45	0.41	0.40
Instrument(s)	0.49	0.13	0.25	0.23	0.13	0.27
F-test						
Overall	6.50***	2.83***	2.46**	3.76***	3.49***	3.38***
Instrument(s)	72.12***	12.07***	25.68***	9.67***	10.17***	24.83***

Note. All endogenous variables and *CRITICS* in logs.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

time of the focal movie's release, the smaller is its opening box office performance. In marginal effects terms, we find that if the critics' reviews of the competitive movies improve by 10 points (ranging from 19 to 84), the focal movies' market-level box office performance reduced by 5.18% (opening day) and 1.59% (one month after release) on average.

Tables 3(b) (*PRE*) and 3(c) (*POST*) help us understand the heterogeneity in the effects of our focal variables—blog volume, blog valence, and advertising—across the various opening markets. First, we note the remarkable consistency of our results across Tables 3(b) and 3(c). The effects of demographic factors on consumer- and firm-generated media are quite consistent across opening day and sales after one month. This provides us with some reassurance about how demographic variables might influence the effects of consumer blogs and firm advertising.

Next, note that differential access to blogs in the different markets is partially accounted for by broadband penetration in those markets. Our results (both

Table 3(a) Movie-Market-Level Factors

	Release day ($t = PRE$)	Postrelease ($t = POST$)
Variable	Estimate (std. error)	Estimate (std. error)
THEATERS (LOG)	0.612419*** (0.010618)	1.184352*** (0.016392)
COMPSTARS	−0.041296 (0.032474)	−0.231896 (0.315791)
COMPREVIEW	−0.005329*** (0.001586)	−0.0016*** (0.000428)
COMPAGE	2.5e−05 (8.7e−05)	0.001694 (0.001217)
R^2	0.987	0.992

Note. Fixed effects are not reported.
$^{***}p < 0.01$.

Table 3(b) Market-Level Factors Driving Responsiveness to Blogs and Advertising—Release Day ($t = PRE$)

Variable	*BROAD*	*WOMEN*	*YOUNG*	*POP*	*INC*	*WHITE*	*ASIAHISP*
BLOGVOL_PRE	−0.0002	−0.6025	0.3080***	0.0019	−0.0017	0.0032	0.0977***
	(0.0002)	(0.3942)	(0.0942)	(0.0047)	(0.0229)	(0.0371)	(0.0350)
BLOGPROP_PRE	−0.0013***	−4.2580***	−0.6329***	0.0057	−0.1584***	−0.5270***	−0.4311***
	(0.0004)	(0.5479)	(0.1388)	(0.0068)	(0.0339)	(0.0572)	(0.0533)
ADVERTISING	−0.0008**	−4.3622***	−0.3503**	0.0045	−0.2287***	0.1770***	−0.0852***
	(0.0004)	(0.6051)	(0.1470)	(0.0075)	(0.0362)	(0.0585)	(0.0368)
CRITICS	1.8e−05	−0.0432	0.0239***	0.0008**	0.0137***	−0.0001	−0.0132***
	(1.9e−05)	(0.0285)	(0.0068)	(0.0003)	(0.0017)	(0.0019)	(0.0025)

Note. Blog, advertising, and critics variables are in logged form.
** $p < 0.05$; *** $p < 0.01$.

Table 3(c) Market-Level Factors Driving Responsiveness to Blogs and Advertising—PostRelease ($t = POST$)

Variable	*BROAD*	*WOMEN*	*YOUNG*	*POP*	*INC*	*WHITE*	*ASIAHISP*
BLOGVOL_POST	3.9e−05	−0.1339	0.4881***	−0.0040**	0.0561	0.0251	0.2329***
	(3.5e−05)	(0.1029)	(0.1462)	(0.0018)	(0.0528)	(0.0834)	(0.0776)
BLOGPROP_POST	−0.0004***	−1.2392***	−0.5684***	−0.0210	−0.3643***	−0.2736**	−0.3190***
	(0.0001)	(0.3262)	(0.1704)	(0.0149)	(0.0731)	(0.1158)	(0.1085)
BLOGVOL_PRE	0.0013	−1.4433	0.0257	−0.0121**	0.1645	0.0957	0.0906
	(0.0025)	(1.0846)	(0.2661)	(0.0056)	(0.1528)	(0.0825)	(0.0765)
BLOGPROP_PRE	−5e−06**	−0.1030**	−0.0051	−0.0008	−4.1e−05**	−0.031425	−0.0005
	(2.4e−06)	(0.0421)	(0.0074)	(0.0023)	(2.3e−05)	(0.042281)	(0.0022)
ADVERTISING	−0.0026**	−1.8468**	−0.4999	−0.0059	−0.1377***	0.1699**	−0.2396**
	(0.0011)	(0.7376)	(0.5394)	(0.0213)	(0.0236)	(0.0681)	(0.1027)
CRITICS	3.7e−05**	−0.0968***	0.0129***	0.0019**	0.0189***	−0.0128**	−0.0078**
	(1.8e−05)	(0.0173)	(0.0047)	(0.0008)	(0.0038)	(0.0059)	(0.0035)

Note. Blog, advertising, and critics variables are in logged form.
** $p < 0.05$; *** $p < 0.01$.

Tables 3(b) and 3(c)) reveal that although some of the effects of broadband penetration are statistically significantly different from zero, their magnitudes are such that differences in broadband penetration do not seem to affect differences in blog volume and valence elasticities across markets (combined with the numbers in Table 1, the differences are typically in the third decimal place). There is some effect on advertising elasticities postrelease in that greater penetration lowers advertising elasticities. One explanation of the limited impact of our metric of broadband penetration is that it only reflects penetration into homes. It is possible that consumers actually access the content away from home, in which case the finding is not surprising.

Having more women in a market makes the market less sensitive to blog valence and less sensitive to advertising. The popular press has alluded to women engaging more and being more sensitive to WOM (e.g., Emerson 2011). Our findings are not consistent with the latter notion. The results appear to be consistent with the more recent results of Aral and Walker (2012) indicating lower susceptibility of women to WOM; however, note that they were looking at message volume. The results are consistent with the Shavitt et al. (1998) finding of more favorable attitudes toward advertising by men, although it is important to note that attitudes may not translate into responsiveness.

Having more young consumers in the market enhances the impact of blog volume but lowers the effect of blog valence. This former finding is consistent with the higher exposure of youth to blogs, as indicated by Forrester Research (2007) and Edelman (2007), and with Aral and Walker (2012) on volume of message effects. Further, unlike the Shavitt et al. (1998) finding, having a younger demographic in a market lowers the impact of prerelease advertising on opening day box office performance, although much has changed in the media landscape since the time of that study.

We also see strong income effects across markets—high-income markets are less responsive to blog valence and to advertising. Consequently, studios engaged in spot advertising might consider reducing their advertising in the higher-income markets as well as those with more women and with younger consumers. The advertising effects appear consistent with Holbrook (1978), given the strong positive correlation between education and income levels. The same pattern also holds in markets with more Asians and Hispanics in the population, although those markets seem more sensitive to "buzz," i.e., the blog volume effect. Finally, we find that markets with larger white populations are more sensitive to advertising after controlling for the other factors. Studios might want to direct more of their advertising to these markets. At the same time, markets with larger white populations are also less sensitive to blog valence.

Our results in Table 3 reveal several interesting patterns of relationships between demographic characteristics and the effectiveness of blog volume, valence, and advertising. To determine the relative magnitudes of the effects of the various demographic interactions, we compute the percentage change in box office performance for a 1% change in each of the demographic characteristics. For example, considering the interaction effect between income and the effects of blog volume, we look at a change in the income level by 1% (in that particular interaction) and compute the impact on box office performance. The biggest effect we find is for gender. A higher proportion of women lowers box office performance with respect to blog valence and advertising much more than any other demographic variable interacted with volume, valence, or advertising. This is followed by the role of income and the proportion of Caucasians in driving advertising effects (prerelease), albeit in opposite directions (see Table 3); the effect of the proportion of Caucasians on advertising (postrelease) and on blog valence (prerelease); the effect of income on advertising and blog valence (postrelease); and the effect of the proportion of young consumers on advertising (prerelease) and blog valence and blog volume (pre- and postrelease). This reveals a nuanced pattern of demographic effects pre- and postrelease as well as across the three drivers of box office performance.

Recovering the Mean Effects of *BLOGVOL*, *BLOGPROP*, and *ADVERTISING*

Next, we turn to the results from estimating the parameters of Equation (2). These results are presented in Table 4. Recall that the standard errors here are inflated relative to ordinary least squares because of (i) instrumenting and (ii) sampling error in the estimates of α_i. Column 1 refers to the results from the release day (*PRE*) analysis, whereas those in column 3 are for the postrelease (*POST*) analysis. Unlike the results in Table 3, the mean effects across markets of consumer- and firm-generated media do depend on when the analysis is conducted. In particular, we find for the *PRE* period that on average across markets, the key drivers of box office performance are blog volume and advertising. Second, the blog sentiment effect, although directionally positive, is not precisely estimated. However, this does not preclude certain markets from having statistically significant effects, given our results from Table 3.[6] This finding is intuitively plausible since the valence at this stage is based largely on the speculation of the bloggers rather than on actual movie quality. On the other hand, the impact of blog volume suggests the importance of "buzz" before a movie's release. Further, advertising also provides information and has a significant impact. Recall that the correlation among the three

Table 4 Recovering the Mean Effects of Advertising and Blogs

	Release day ($t = PRE$)	Postrelease ($t = POST$)	Postrelease ($t = POST$)
BLOGVOL_PRE	0.269*		0.112
	(0.143)		(0.126)
BLOGPROP_PRE	0.162		0.071
	(0.114)		(0.180)
ADVERTISING	0.389**	0.465**	0.529***
	(0.168)	(0.197)	(0.191)
BLOGVOL_POST		0.051	−0.010
		(0.164)	(0.142)
BLOGPROP_POST		0.495**	0.349**
		(0.226)	(0.158)
RATINGVOL[a]			0.084
			(0.093)
RATINGVAL[a]			0.221***
			(0.074)
STARS	−0.105	−0.359	−0.310
	(0.125)	(0.296)	(0.231)
CRITICS	0.011**	0.017**	0.016*
	(0.005)	(0.009)	(0.009)
PG	−0.203	0.278	0.228
	(0.196)	(0.434)	(0.371)
PG13	0.149	0.130	0.061
	(0.171)	(0.295)	(0.314)
ACTION	−0.285	0.221	0.129
	(0.264)	(0.509)	(0.421)
COMEDY	−0.372	0.528	0.507
	(0.334)	(0.561)	(0.422)
DRAMA	−0.427*	0.494	0.228
	(0.245)	(0.512)	(0.429)
HOLIDAY	0.335*	0.286*	0.256*
	(0.192)	(0.157)	(0.142)
SEASON 1	−0.186	−0.359	−0.048
	(0.237)	(0.638)	(0.486)
SEASON 2	−0.663**	−0.512	−0.401
	(0.264)	(0.684)	(0.515)
SEASON 3	−0.191	−0.119	−0.218
	(0.245)	(0.647)	(0.479)
R^2	0.904	0.917	0.949

Note. Blog, advertising, and critics variables are all in logged form.
[a]Based on user reviews from Yahoo! Movies.
$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

[6] One could also argue that the lack of significance is due to the larger standard errors from instrumenting and from correcting for sampling error. At the same time, as our postrelease analysis will show, the valence measure does have a statistically significant effect there.

variables is not very high (ranging between 0.22 and 0.31), implying that multicollinearity is unlikely to play a role in explaining our results.

We report two sets of results for the *POST* period. The first includes only postrelease blog volume and blog valence and not the information in the prerelease period. Further, we use total advertising to date and do not include the user ratings information. The results deviate from those above in two key respects. First, the effect of buzz or blog volume is much smaller than in the *PRE* case (0.051 versus 0.269); further, the effect is no longer precisely estimated. More importantly, however, the effect of blog valence is now much larger than in the previous case (0.495 versus 0.162) and is precisely estimated. These results seem to suggest that once a movie is released, consumers care about the content of the movie and how that is viewed by consumers rather than whether or not more consumers expressed opinions about the movie.

The above finding is supported by the analysis where we also include prerelease user-generated content and postrelease volume and valence of user ratings (column 4). We find that prerelease user-generated content has no statistically significant impact on postrelease box office performance. Further postrelease, the valence of blogs as well as user ratings appears to influence box office behavior (see also Chintagunta et al. 2010). Advertising continues to be a significant driver of box office performance in these data. An important note about advertising elasticities—a majority of previous studies (e.g., Sethuraman and Tellis 1991) finds that advertising elasticities tend to be small; this is not the case in the movie business. Specifically, movies are short life-cycle products. For these types of products it does appear that advertising plays a major role in driving sales.

Overall Market Level Effects of Advertising and Blogs: Identifying "Firm-Driven" Markets vs. "Consumer-Driven" Markets

Having obtained the estimates of the mean effects (across markets) of the blogs and advertising, we next compute the elasticity for each market. In Figure 2 we provide the distribution of elasticities for prerelease blog volume (Figure 2(a)); postrelease blog valence (Figure 2(b)); and advertising (Figure 2(c)) across the various geographic markets. Two points to note with these figures: (i) We use only the magnitudes of those elasticities that are statistically significantly different from zero at the 10% level of significance; all others are set to zero. (ii) For advertising, we use the elasticities from the prerelease analysis. Figure 2 shows that there is a considerable heterogeneity in the elasticities for the valence and advertising measures—the coefficients of variation for these variables are 0.36 and 0.31,

Figure 2(a) Distribution of Prerelease Blog Volume Elasticities

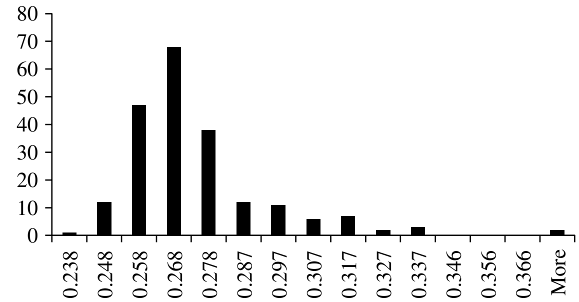


Figure 2(b) Distribution of Postrelease Blog Valence Elasticities

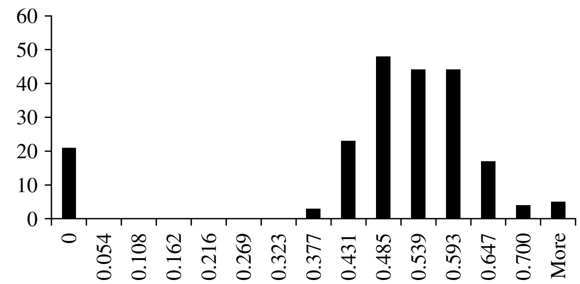
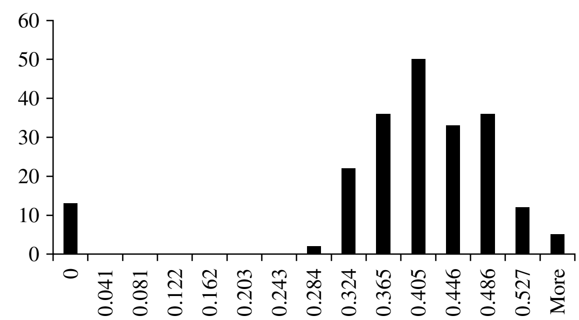


Figure 2(c) Distribution of Prerelease Advertising Elasticities



respectively. There appears to be less heterogeneity in the blog volume effect with a coefficient of variation of 0.08.

Another noteworthy point about these elasticities is the correlation among the three measures across markets. In particular, we find that the prerelease blog volume elasticity has a correlation of −0.28 with postrelease blog valence; that is, markets that are very sensitive to buzz prerelease tend to be somewhat less sensitive to positively valenced comments postrelease. This finding can be explained by how demographic characteristics of markets are differentially related to volume and valence sensitivities (Tables 3(b) and 3(c)). In particular, differences in the sensitivities of the young and the various racial groups across these blog measures seem to drive the negative correlation. On the other hand, we find that advertising sensitivity is positively correlated with blog valence sensitivity (correlation = 0.42). The similarities in the *BLOGPROP_POST* and *ADVERTISING* rows in Table 3 seem to account for this correlation.

Table 5 Top 20 Consumer-Driven and Firm-Driven Markets—Overall Market Elasticities

	Consumer driven (blogs); Gross = DMA rank by box office gross								
	Blog volume			Blog valence			Firm driven (advertising)		
Rank	Market	Elasticity	Gross	Market	Elasticity	Gross	Market	Elasticity	Gross
1	DENVER	0.3765	17	BATON ROUGE	0.7543	91	CASPER-RIVERTON	0.5672	196
2	CEDAR RAPIDS-WTRLO-IWC & DUB	0.3692	102	CHAMPAIGN & SPRNGFLD-DECATUR	0.7466	77	CHARLOTTESVILLE	0.5466	170
3	HELENA	0.3287	201	MERIDIAN	0.7195	191	MARQUETTE	0.5427	172
4	FRESNO-VISALIA	0.3282	42	GREENWOOD-GREENVILLE	0.7095	206	RAPID CITY	0.5321	186
5	GRAND JUNCTION-MONTROSE	0.3275	194	ALEXANDRIA, LA	0.7029	175	GREAT FALLS	0.5307	192
6	CHICAGO	0.3201	3	MARQUETTE	0.6627	172	WHEELING-STEUBENVILLE	0.5058	168
7	BAKERSFIELD	0.3182	96	PEORIA-BLOOMINGTON	0.6615	93	MANKATO	0.5055	197
8	LAFAYETTE, LA	0.3158	131	RALEIGH-DURHAM (FAYETVLLE)	0.6603	34	SIOUX FALLS(MITCHELL)	0.5049	145
9	PALM SPRINGS	0.3129	86	LINCOLN & HASTINGS-KRNY	0.6513	135	CHATTANOOGA	0.4975	107
10	ALPENA	0.3119	209	MANKATO	0.6368	197	PADUCAH-CAPE GIRARD-HARSBG	0.4947	89
11	HARLINGEN-WSLCO-BRNSVL-MCA	0.3109	54	CHARLOTTESVILLE	0.6262	170	CORPUS CHRISTI	0.4942	104
12	MONTEREY-SALINAS	0.3093	67	BIRMINGHAM (ANN AND TUSC)	0.6246	55	ALBANY-SCHENECTADY-TROY	0.4939	56
13	YUMA-EL CENTRO	0.309	154	ST. LOUIS	0.6232	27	JONESBORO	0.4939	183
14	LOS ANGELES	0.3078	1	CORPUS CHRISTI	0.6222	104	EUREKA	0.4934	167
15	OMAHA	0.3033	63	SYRACUSE	0.6141	76	TYLER-LONGVIEW (LFKN & NCGD)	0.4903	111
16	YAKIMA-PASCO-RCHLND-KNNWCK	0.3017	121	EUREKA	0.6138	167	JACKSON, MS	0.4902	113
17	VICTORIA	0.3015	182	PADUCAH-CAPE GIRARD-HARSBG	0.6133	89	SALT LAKE CITY	0.4869	28
18	SPRINGFIELD, MO	0.2994	99	LANSING	0.6123	75	ROCHESTER, NY	0.4861	62
19	SANTABARBRA-SANMAR-SANLUOB	0.2989	65	GREEN BAY-APPLETON	0.6119	108	NEW ORLEANS	0.4861	48
20	SIOUX CITY	0.2981	179	BOWLING GREEN	0.6101	165	LEXINGTON	0.4855	83

In Table 5 we provide the top 20 markets on each of the three elasticities—blog volume (prerelease), blog valence (postrelease), and advertising. We note that several large markets by box office revenues—Los Angeles, Chicago, and Denver—rank among the top in terms of how sensitive they are to movie buzz. These same markets are, however, much less sensitive to blog valence and advertising. On the other hand, markets that are sensitive to postrelease blog valence tend to include smaller towns, with the possible exception of cities like St. Louis. We also find a few medium-sized cities showing a high responsiveness to advertising—New Orleans; Salt Lake City; Lexington, New York; and Rochester, New York.

Two noteworthy features of Table 5 are the following. First, it provides a basis by which studios can target release markets, especially for the more limited release ones. Specifically, if the firm would like to generate prerelease buzz by having some special events around the movie, markets such as Los Angeles and Chicago seem more appropriate because these markets also tend to respond the most to such prerelease activity. On the other hand, if the studio wants to target its advertising resources more judiciously, then focusing on spot TV markets in the most responsive geographic regions in Table 5 might be most appropriate. A second feature of Table 5 is that there does not appear to be any natural geographic or regional grouping of the markets. This argues for a more targeted approach than a broad-based or regional approach to marketing the movies.

Figure 3 shows the regional differences in responsiveness to blogs and advertising. We find that the markets most responsive to blogs and advertising and the markets highly responsive to only blogs are spread more or less evenly across the different regions. Moreover, the markets highly responsive to only advertising seems to be concentrated in the Midwest region. The figure also indicates that the markets that are least responsive to both blogs and advertising seem to be clustered in the East Coast region. Overall, there seems to be more variation in market responsiveness to blogs and advertising in the Midwest and East Coast regions. This suggests the need for a more microlevel advertising and movie release strategy in these regions compared with the rest of the country.

We also conducted additional analyses that shed light on the heterogeneity of market responsiveness depending on the genre of the movie. We find that not all genres of movies do equally well in the top geographic markets (based on average opening day gross

Figure 3 Regional Differences in Reponsiveness to Blogs and Advertising

Notes. Both blogs and advertising—refers to prerelease blog volume, postrelease blog valence, and advertising. Only blogs—refers to prerelease blog volume and postrelease blog valence. Prerelease blogs—refers to prerelease blog volume. Postrelease blogs—refers to postrelease blog valence.

across movies). For example, the New York market is most responsive to a movie of the comedy genre and least responsive to a movie of the thriller genre. These insights imply that it is important to take the genre of a movie into account when choosing the initial set of markets as well.

Implications for Studios/Distributors

How can our analysis aid studios in picking an initial set of geographic markets based on their responsiveness to advertising and blogs? The importance of market selection is critical for limited releases. However, as mentioned earlier, even the widest releases are not released in all markets on the opening day (as we see from Table 1, across 208 markets and 75 movies, there are only 9,000+observations out of a possible number exceeding 15,000 in the *PRE* time period). Using the market-level elasticities, we next came up with groups of markets that range from those geographic areas whose advertising, prerelease blog volume, and postrelease blog valence elasticities all exceed the median elasticity for that variable across markets to those areas where all three elasticities fall below the median. Given the correlations in elasticities across markets discussed above, this classification scheme yielded 28 markets with above median elasticities on all three variables; 13 markets with above median blog elasticities on both blog metrics; 20 markets with high blog volume and advertising elasticities; 44 markets with high blog valence and advertising elasticities (the 20 and 44 relate to the negative and positive correlations in the corresponding elasticities); and 43, 18, and 10 markets, respectively, with above median blog volume, valence, and advertising elasticities only. Finally, we have 33 markets that are below the median level on all three elasticities. Table 6 provides the top and bottom groups of markets. If a studio is going for a limited release of a well-previewed movie, it might want to choose markets that have high elasticities on all three dimensions, as in Table 6. However, if the movie is not as well previewed, then focusing on the top advertising elasticities markets in Table 5 might be a preferred approach. Regardless of the specific strategy the studio would like to adopt, our results should provide some guidance for studios' movie launch decisions.

Next, we consider studios' release strategies for movies that did not receive the widest release (i.e., were released in 170 markets or fewer) to see if they could have performed better if movie studios/distributors had access to the information on market-level elasticities available from our analysis. Because prior to release the two dimensions on which studios can base their decisions are advertising and blog volume elasticities, we focus on these two measures. Say a movie was released in 100 markets. We then look at the proportion of the top 100 markets ranked by advertising elasticities and by prerelease blog volume elasticities accounted for by the 100 markets where the movie was actually released. This number can range anywhere from 0 (since there are more than 200 markets in the data) to 1. The closer the number is to 1, the better the release strategy *based purely on the market-level elasticities we compute*. Based on this metric, we find that across movies the proportion ranges from 8% to 82% for advertising elasticities and from 3% to 80% for blog volume elasticities. For example, the movie *Little Black Book* was released in 54% of the most advertising responsive markets and 41% of the most blog volume responsive markets. We summarize this information in Table 7(a) by looking at all movies released in fewer than 170 markets at the studio level by the top studios. For the movies in our data, we find that Sony seems to have the best coverage and Warner Bros. the worst. This implies that the latter studio may have the most to gain from rethinking its launch markets. There is also some variation in the release strategies across genres, but it is not that marked as in the case of studios. Movies of the thriller genre have the largest market coverage, and movies of the drama genre have the smallest market coverage (Table 7(b)). We emphasize that this is one metric to assess studios' market release strategies.

Table 6 Top and Bottom Markets by Responsiveness

	Top markets	Bottom markets
1	CASPER-RIVERTON	BALTIMORE
2	CHICO-REDDING	BLUEFIELD-BECKLEY-OAK HILL
3	COLUMBIA-JEFFERSON CITY	BUFFALO
4	DOTHAN	CHARLESTON, SC
5	EL PASO (LAS CRUCES)	CINCINNATI
6	EUGENE	COLUMBUS, GA
7	EUREKA	COLUMBUS, OH
8	EVANSVILLE	DETROIT
9	GREAT FALLS	ERIE
10	GREEN BAY-APPLETON	FAIRBANKS
11	JONESBORO	FLINT-SAGINAW-BAY CITY
12	JUNEAU	GRAND RAPIDS-KALMZOO-B.CRK
13	MADISON	GREENVILLE-N.BERN-WASHNGTN
14	MANKATO	GREENVLL-SPART-ASHEVLL-AND
15	MERIDIAN	HARTFORD & NEW HAVEN
16	MISSOULA	HOUSTON
17	MONROE-EL DORADO	KNOXVILLE
18	MYRTLE BEACH-FLORENCE	MINOT-BISMARCK-DICKINSON
19	NORTH PLATTE	ORLANDO-DAYTONA BCH-MELBRN
20	SALT LAKE CITY	OTTUMWA-KIRKSVILLE
21	SIOUX FALLS (MITCHELL)	ROCKFORD
22	SPOKANE	SALISBURY
23	TUCSON (SIERRA VISTA)	SAN ANGELO
24	TULSA	SAVANNAH
25	TWIN FALLS	SEATTLE-TACOMA
26	WACO-TEMPLE-BRYAN	SPRINGFIELD-HOLYOKE
27	WAUSAU-RHINELANDER	TALLAHASSEE-THOMASVILLE
28	WILMINGTON	TAMPA-ST. PETE (SARASOTA)
29		TRAVERSE CITY-CADILLAC
30		TYLER-LONGVIEW(LFKN & NCGD)
31		WASHINGTON, DC (HAGRSTWN)
32		WATERTOWN
33		WEST PALM BEACH-FT. PIERCE

Table 7(a) Studios' Coverage of Most Responsive Markets in Their Limited Release Strategies

Studio	Top blog market coverage (mean across all movies) (%)	Top advertising market coverage (mean across all movies) (%)
Sony	62	70
Universal	42	51
Fox	39	49
Paramount	37	43
Warner Bros.	28	41
All studios	44	53

Table 7(b) Studios' Coverage of Most Responsive Markets in Their Limited Release Strategies by Genre Across Studios

Studio	Top blog market coverage (mean across all movies) (%)	Top advertising market coverage (mean across all movies) (%)
Thriller	51	58
Action	49	58
Comedy	42	51
Drama	36	46
All Genres	44	53

Robustness Checks

We undertook several robustness checks.

(i) *Market share as a dependent variable*: We find that although our point estimates are affected (as we would expect), the qualitative nature of our results for blog volume, blog valence, and advertising remain unchanged.

(ii) *Directly estimating market-level effects of blogs and advertising*: Our market-level effects of blogs and advertising have been obtained by projecting the market-level effects onto demographics and other factors. We estimated a market-specific effect for blog volume, blog valence, and advertising and then projected these estimates onto demographic and other factors for the release date analysis. Demographics explain anywhere from 50% to 67% of the market specific effects for the three variables; further, the nature of our results remains unchanged.

(iii) *Using different functional forms for the regression equation* (e.g., linear, log-log, semilog): Although not reported here, we find that our results are relatively unaffected by alternative functional forms for the regressions in Equations (1) and (2).

(iv) *Using different operationalizations of the variables* (e.g., ratios, differences, percentage differences for the valence variable): Once again, the qualitative nature of our results was unaffected by alternative specifications for our variables.

(v) *Using time-series data*: The dependent measure is the box office performance of the movie in each market in each week. We have five observations (opening week and each of four subsequent weeks) for each movie market. The analysis is carried out in two steps, where in the first step we estimate *movie-week* fixed effects in addition to other movie-market-week drivers such as the number of theaters. In the second step we regress these fixed effects on the weekly levels of blog volume, valence, and advertising in the preceding one-month period, movie fixed effects, and age of the movie. Because no time-varying instruments are available to us, this analysis does not account for endogeneity. Further, because user reviews are not available prior to the movie's release, for the opening box office we set the values for these variables to zero and for each of the subsequent weeks we compute the numbers for the previous corresponding week. Consistent with our postrelease analysis above, we find blog valence and advertising to be significant drivers of box office performance.

(vi) *Too few blogs in 2004?* We gathered blog data for movies released in 2009. The numbers from the two years (2004 and 2009) are quite comparable for many of the movies, although the higher grossing movies in 2009 had more blogs than those in 2004. Blog valence, however, is not affected. This leads us to believe that the basic nature of our results is unlikely to be affected. However, because we do not have box office data from 2009, we cannot verify this.

Discussion and Conclusions

Our analysis of pre- and postrelease consumer-generated media (prerelease blogs and postrelease blogs and user ratings) and firm-generated media (advertising) on the box office performance of movies at release and one month subsequent to release reveals considerable heterogeneity in these effects across instruments and geographic markets. Release day performance is impacted most by prerelease blog volume and advertising, whereas postrelease performance is influenced by blog valence, user rating valence, and advertising. Across markets, there is more variance in advertising and blog valence (postrelease) elasticities than there is in blog volume (prerelease) elasticities. Further, the former two elasticities are positively correlated across markets and are negatively correlated with blog volume elasticities. We identify the top 20 markets in terms of their elasticities to each of these three instruments. Then we classify markets in terms of their sensitivities across these three instruments to identify the most sensitive markets that studios can target with their limited release strategies. Finally, we characterize the extent to which studios could have improved their limited release strategies by identifying the overlap between the actual release markets and the most responsive ones.

One limitation of the current study is that we treat all blogs as having an equal impact on performance. In reality, some blogs may be more popular and hence more credible than others. The reach of such blogs may be higher as well. It would be of interest to characterize blogs based on levels of persuasion, information content, etc., in addition to volume and valence, which we currently include. From a methodological perspective, one might also consider using recent Bayesian approaches to instrumental variables estimation to estimate the model's parameters (Conley et al. 2012) and obtain more efficient estimates. From a data perspective, supplementing our analysis with actual data on the exposure to blogs could strengthen the insights obtained. Another data limitation is that our advertising and blog measures as noted before are at the national level. Local advertising and blog effects can easily be incorporated into the analysis if the appropriate data are available. Currently, local advertising is part of the error term in Equation (1) and would be problematic if correlated with market-specific observable factors included in that regression, in our case, the number of theaters in which the movie is released in the local market. Since we instrument for the number of theaters in the estimation of Equation (1) and because that instrument is unlikely to be correlated with local advertising, our estimation results from Equation (1) are still valid. Our blog measures include blogs generated both locally and elsewhere. Our current analysis assumes that these are indistinguishable to blog readers, but if they are distinguishable, then our assumption is that the effects at the local market level are the same for both of these types of blogs. Relaxing these assumptions would require breaking down the blog data to the local level. Finally, in a recent paper, Godes and Mayzlin (2009) explore the effects of firm-generated WOM communication. Extending the analysis presented here to account for such WOM could be a useful research prospect.

Acknowledgments
The authors thank Jehoshua Eliashberg and the participants of the Emergence and Impact of User-Generated Content conference organized by the Wharton Interactive Media Initiative (WIMI), December 2009. The authors also thank the Marketing Science Institute (MSI) and the Kilts Center for Marketing at the University of Chicago for financial support. Comments from reviewers, an associate editor, and editor-in-chief Gérard P. Cachon are gratefully acknowledged. This research forms a part of the first author's doctoral dissertation.

References
Ainslie A, Drèze X, Zufryden F (2005) Modeling movie life cycles and market share. *Marketing Sci.* 24(3):508–517.

Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341.

Aronson E (1972) *The Social Animal* (W.H. Freeman and Company, San Francisco).

Banning SA, Sweetser KD (2007) How much do they think it affects them and whom do they believe?: Comparing the third-person effect and credibility of blogs and traditional media. *Comm. Quart.* 55(4):451–466.

Basuroy S, Chatterjee S, Ravid SA (2003) How critical are critical reviews? The box office effects of film critics, star-power, and budgets. *J. Marketing* 67(4):103–117.

Chen, Y, Wang Q, Xie J (2011) Online social interactions: A natural experiment on word of mouth versus observational learning. *J. Marketing Res.* 48(2):238–254.

Chevalier J, Mayzlin D (2006) The effect of word of mouth online: Online book reviews. *J. Marketing Res.* 43(3):345–354.

Chintagunta PK, Gopinath S, Venkataraman S (2010) The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Sci.* 29(5):944–957.

Conley T, Hansen C, Rossi PE (2012) Plausibly exogenous. *Rev. Econom. Statist.* 94(1):260–272.

Copeland H (2004) Blog reader survey. *Blogads* (May 21), http://www.blogads.com/survey/blog_reader_survey.html.

Dellarocas C, Zhang XM, Awad N (2007) Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Marketing* 21(4):23–45.

Dewan S, Ramaprasad J (2012) Music blogging, online sampling, and the long tail. *Inform. Systems Res.* 23(3, part 2 of 2): 1056–1067.

Dhar V, Chang EA (2009) Does chatter matter? The impact of user-generated content on music sales. *J. Interactive Marketing* 23(4):300–307.

Duan W, Gu B, Whinston AB (2008) The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *J. Retailing* 84(2):233–242.

Dubé J-P, Manchanda P (2005) Differences in dynamic brand competition across markets: An empirical analysis. *Marketing Sci.* 24(1):81–95.

Edelman (2007) A corporate guide to the global blogosphere. Accessed July 24, 2013, http://edelmaneditions.com/wp-content/uploads/2010/12/edelman-corporate-guide-to-global-blogosphere-2007.pdf.

Elberse A, Eliashberg J (2003) Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Sci.* 22(3):329–354.

Emerson R (2011) Women use social media more than men: Study. *Huffington Post* (November 23), http://www.huffingtonpost.com/2011/09/23/women-use-social-media-more_n_978498.html.

Forrester Research (2006) NACTAS Q1 2006 devices and access online survey. Accessed July 24, 2013, http://www.forrester.com/NACTAS+Q1+2006+Devices+Access+Online+Survey/-/E-SUS471.

Gleser GC, Gottschalk LA, John W (1959) The relationship of sex and intelligence to choice of words: A normative study of verbal behavior. *J. Clinical Psych.* 15(2):182–191.

Godes D, Mayzlin D (2004) Using online conversations to study word-of-mouth communication. *Marketing Sci.* 23(4):545–560.

Godes D, Mayzlin D (2009) Firm created word-of-mouth communication: Evidence from a field test. *Marketing Sci.* 28(4):721–729.

Goldfarb A, Prince J (2008) Internet adoption and usage patterns are different: Implications for the digital divide. *Inform. Econom. Policy* 20(1):2–15.

Hoffman DL, Novak TP (2000) The growing digital divide: Implications for an open research agenda. Brynjolfsson E, Kahin B, eds. *Understanding the Digital Economy: Data, Tools, Research* (MIT Press, Cambridge, MA), 245–260.

Holbrook M (1978) Beyond attitude structure: Toward the informational determinants of attitude. *J. Marketing Res.* 15(4):545–556.

Jank W, Kannan PK (2005) Understanding geographical markets of online firms using spatial models of customer choice. *Marketing Sci.* 24(4):623–634.

Johnson TJ, Kaye BK (2004) Wag the blog: How reliance on traditional media and the Internet influence credibility perceptions to weblogs among blog users. *Journalism & Mass Comm. Quart.* 81(3):622–642.

Lenhart A, Fox S (2006) Bloggers: A portrait of the internet's new storytellers. Report, Pew Internet and American Life Project, Washington, DC.

Liu Y (2006) Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* 70(3):74–89.

Liu Y, Huang X, An A, Yu X (2007) ARSA—A sentiment aware model for predicting sales performance using blogs. *Proc. 30th Annual Internat. ACM SIGIR Conf. Res. Development in Inform. Retrieval, Amsterdam*.

Mishne G, Glance N (2006) Predicting movie sales from blogger sentiment. Presentation, Spring 2006 Symposium of the American Association for Artificial Intelligence, May 22–26, Edinburgh, UK.

Moe W, Trusov M (2011) The value of social dynamics in online product ratings forums. *J. Marketing Res.* 48(3):444–456.

Nair H, Manchanda P, Bhatia T (2010) Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *J. Marketing Res.* 47(5):883–895.

Onishi H, Manchanda P (2012) Marketing activity, blogging and sales. *Internat. J. Res. Marketing* 29(3):221–234.

Raine L (2004) The state of blogging. Report, Pew Internet and American Life Project, Washington, DC.

Sadikov E, Parameswaran A, Venetis P (2009) Blogs as predictors of movie success. *Proc. Third Internat. ICWSM Conf.* (Association for the Advancement of Artificial Intelligence, Palo Alto, CA), 304–307.

Sethuraman R, Tellis GJ (1991) An analysis of the tradeoff between advertising and pricing. *J. Marketing Res.* 31(2):160–174.

Shavitt S, Lowrey P, Haefner J (1998) Public attitudes towards advertising: More favorable than you might think. *J. Advertising Res.* 38(4):7–22.

Sinai T, Waldfogel J (2004) Geography and the Internet: Is the Internet a substitute or a complement for cities? *J. Urban Econom.* 56(1):1–24.

Stephen AT, Galak J (2012) The effects of traditional and social earned media on sales: A study of a microlending marketplace. *J. Marketing Res.* 49(5):624–639.

Stock JH, Watson MW (2010) *Introduction to Econometrics* (Addison-Wesley, Boston).

Trusov M, Bucklin RE, Pauwels K (2009) Estimating the dynamic effects of online word-of-mouth on member growth of a social network site. *J. Marketing* 73(5):90–102.

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