EXPLORING THE VALUE OF ONLINE PRODUCT REVIEWS IN FORECASTING SALES: THE CASE OF MOTION PICTURES

CHRYSANTHOS DELLAROCAS, XIAOQUAN (MICHAEL) ZHANG, AND NEVEEN F. AWAD

The growing popularity of online product review forums invites the development of models and metrics that allow firms to harness these new sources of information for decision support. Our work contributes in this direction by proposing a novel family of diffusion models that capture some of the unique aspects of the entertainment industry and testing their performance in the context of very early postrelease motion picture revenue forecasting. We show that the addition of online product review metrics to a benchmark model that includes prerelease marketing, theater availability and professional critic reviews substantially increases its forecasting accuracy; the forecasting accuracy of our best model outperforms that of several previously published models. In addition to its contributions in diffusion theory, our study reconciles some inconsistencies among previous studies with respect to what online review metrics are statistically significant in forecasting entertainment good sales.

JOURNAL OF INTERACTIVE MARKETING VOLUME 21 / NUMBER 4 / AUTUMN 2007
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/dir.20087

CHRYSANTHOS DELLAROCAS
is Associate Professor of Information Systems, R. H. Smith School of Business, University of Maryland in College Park; e-mail: cdell@umd.edu

XIAOQUAN (MICHAEL) ZHANG
is Assistant Professor, Department of Information Systems & Management, Hong Kong University of Science and Technology (HKUST) Business School; e-mail: zhang@ust.hk

NEVEEN F. AWAD
is Assistant Professor in Information Systems and Manufacturing, School of Business Administration, Wayne State University, Detroit, MI; e-mail: nawad@wayne.edu

INTRODUCTION

In the summer of 2004, Steven Spielberg released the movie *The Terminal*, a film with a $75 million production budget, a $35 million marketing budget, and a star cast that included Tom Hanks and Catherine Zeta-Jones. During its opening weekend, *The Terminal* grossed $19 million. This amount was comparable to the opening weekend gross of Tom Hanks's other movies (e.g., *The Road to Perdition*, $22 million; *The Green Mile*, $18 million; *You've Got Mail*, $18 million; *Forest Gump*, $23 million). Nevertheless, *The Terminal* subsequently tanked in theaters, grossing a total of only $77 million. This amount was far less than prerelease predictions had suggested and less than half the average of the cumulative gross of Tom Hanks's other movies ($157 million). Unfavorable word-of-mouth from consumers was cited as the main culprit.

Whereas marketing plays an important role in a movie's opening weekend, consumer word-of-mouth has been frequently cited as the single most important factor that determines the *long-term* success of motion pictures and other experience goods (De Vany & Walls, 1996). Until recently, however, the reliable measurement of consumer word-of-mouth had remained elusive. The situation is changing thanks to the emergence of Internet-mediated communities where consumers exchange their experiences about products and services. In contrast to "offine" word-of-mouth communities where articulated opinions "disappear into thin air," online communities maintain a persistent, and easily accessible, public record of everything that has been posted so far.

Online product reviews represent a potentially valuable tool for firms, who can use them to monitor consumer attitudes toward their products in real time, and adapt their manufacturing, distribution, and marketing strategies accordingly. The development of appropriate models and metrics is a crucial step in harnessing these new sources of information. The recently created Word Of Mouth Marketing Association (www.womma.org) maintains a council on online research and metrics and holds regular conferences on the topic. An industry of professional online content monitoring firms (such as Nielsen Buzzmetrics, Cymphony, MotiveQuest, etc.) has emerged, offering their clients a variety of metrics and reports. Although there exists considerable industry momentum on this topic, so far, few concrete principles have emerged.

Academics have recognized the importance of online product reviews and have already produced a number of important results in this area. Using field data and controlled experiments respectively, Chevalier and Mayzlin (2006) and Senecal and Nantel (2004) independently established the influence of online product reviews on consumer purchase decisions. On the topic of metrics, the results of early studies are intriguing but also somewhat inconsistent with one another: Godes and Mayzlin (2004) have looked at how metrics of Usenet conversations about television shows relate to their Nielsen (viewership) ratings. They find that the dispersion of conversations among different newsgroups has significant explanatory power, but their volume doesn't. Liu (2006) studied the impact of Yahoo! Movies prerelease message board discussions on motion picture box office revenues. He finds that the volume, but not the valence, of online conversations has explanatory power. Duan et al. (2005) looked at the relationship between daily Yahoo! Movies reviews and box office sales. They similarly find that the volume, but not the valence, of movie ratings has explanatory power.

Although these studies have compellingly established the significance of online product reviews as influencer and predictor of product sales, they do not offer concrete models that firms can use in their decision making. Most previous studies explicitly state that their objective is the establishment of statistically significant relationships and not the development of forecasting models. Furthermore, their mutually inconsistent, and somewhat counterintuitive, findings with respect to what metrics have explanatory power invite a closer look.

Our work takes research on online review metrics to its logical next step: We propose concrete models that can be used for decision support in a specific business context (sales forecasting of entertainment goods) and provide quantitative assessments of the information value of online review metrics relative to more traditional metrics. Furthermore, we reconcile some of the inconsistencies among previous studies with respect to what metrics are statistically significant and demonstrate that significance results that are better aligned with theoretical predictions can be obtained by embedding online review metrics in models that more closely match the properties of the markets of interest. Specifically, our study proposes a novel family of revenue forecasting models that are tailored to the entertainment industry and tests their performance in the context of early (opening weekend) postrelease motion picture revenue forecasting. We estimate our models using metrics obtained from user reviews posted on Yahoo! Movies during the opening weekend of a movie, together with a number of more traditional metrics, such as a movie's marketing budget, theatrical availability, professional critic reviews and very early box office revenues.

In contrast to prior studies of online content, which employ linear regression models, our models are based on diffusion theory with some novel elements that capture the unique patterns of entertainment good marketing (heavy prerelease campaigns that usually decline rapidly postrelease) and the time-locality of consumer word-of-mouth (people tend to talk a lot about movies immediately after watching them and less as time goes by). Our models derive high quality forecasts of a movie's future weekly revenues and work well for both wide release (blockbuster) and narrow release (sleeper) movies. Our best model was able to forecast the weekly revenue trajectory of movies in our holdout sample with a mean absolute percentage error (MAPE) of 10% on any given week; this figure represents a substantial improvement over several other models reported in the literature. In terms of the incremental forecasting power of online review metrics relative to more traditional metrics, we find that the addition of online review metrics to a benchmark model that includes prerelease marketing, theater availability, and professional critic reviews reduced the model's MAPE by 38%. Remarkably, in the case of sleeper movies (i.e., movies that are initially released in a small number of theaters and that rely on consumer word-of-mouth for revenue growth), we find that the forecasting accuracy of a model that was based exclusively on online review metrics (i.e., does not use marketing, theater availability, professional critic review, or early box office data) outperforms that of models that also include traditional metrics.

This research contributes on several fronts. First, we contribute to the literature on motion picture revenue forecasting by proposing a novel family of diffusion models that captures some of the unique aspects of entertainment goods (declining postrelease marketing, consumer word-of-mouth whose intensity is correlated with the time of consumption) and whose forecasting accuracy outperforms that of several previously published models. Second, we contribute to research on the business value of consumer-generated content by demonstrating that online review metrics have a broader potential than what has been recognized in the past: In addition to using online review metrics as a proxy of consumer WOM we show that the early volume of online reviews provides an excellent proxy of early box office sales, whereas metrics of online reviewer demographics provide useful indications regarding a product's demand in different customer segments. Last, but not least, our research helps reconcile some of the inconsistencies among previous studies with respect to what online review metrics are significant predictors of future sales. Specifically, we show that, when used in the proper place in the context of a nonlinear diffusion model, the volume, valence, and dispersion of online movie reviews are all statistically significant in predicting future sales in directions that are consistent with what theory would suggest.

The rest of the paper is organized as follows. We begin by discussing related work. We then describe our data set and independent variables. We introduce our forecasting models and estimation technique, present the results of fitting them to our data set and compare their forecasting accuracy to that of older models. Finally, we discuss the managerial implications of this work and suggest potential avenues for future research.

RELATED WORK

Our work relates to two important streams of past research: forecasting models of motion picture revenues and methodologies for measuring consumer word-of-mouth.

Forecasting Models of Motion Picture Revenues

Predicting the success of a motion picture has largely been viewed in the industry as a "wild guess" (Litman & Ahn, 1998). Despite such difficulty, several researchers have proposed models that attempt to forecast motion

TABLE 1 Examples of Econometric Studies of Motion Picture Box Office Performance

FACTORS CONSIDERED	STUDIES
Star power	De Vany and Walls 1999; Ravid 1999
Movie genre and MPAA ratings	Austin and Gordon 1987
Academy awards	Dodds and Holbrook 1988
Media advertising	Faber and O' Guinn 1984
Timing of release	Krider and Weinberg 1996
Distribution strategy	Jones and Ritz 1991
Competition from other movies	Ainslie, Dreze and Zufryden 2003
Professional critic reviews	Eliashberg and Shugan 1997; Reinstein and Snyder 2005; Basuroy, Chaterjee and Ravid 2003
Combination of factors	Litman 1983; Neelamegham and Chintagunta 1999; Elberse and Eliashberg 2003

picture revenues. Such models can be classified along three dimensions.

One classification can be based on the explanatory variables employed. Table 1 provides illustrative examples of the large number of different factors that have been examined.

Another classification can be based on the timing of the forecast. Some of the proposed models are designed to produce forecasts *before* a movie's initial release (Litman, 1983; Zufryden, 1996; De Silva, 1998; Eliashberg et al., 2000), whereas others focus on forecasting later-week revenues *after* a movie's early box office revenues become known (Sawhney & Eliashberg, 1996; Neelamegham & Chintagunta, 1999). The latter category tends to generate more accurate forecasting results because these models have access to more explanatory variables, including early box office receipts, critic reviews, and word-of-mouth effects.

A third classification can be based on the type of model used. The simplest models use linear regression (Eliashberg & Shugan, 1997; Ravid, 1999; Basuroy et al., 2003). The more advanced models are typically based on diffusion theory. Some studies (Jedidi et al., 1998; Swami et al., 1999) attempt to fit an exponential distribution with two parameters to describe the rapidly declining weekly box office revenues of blockbuster movies. However, this pattern does not necessarily conform to sleeper movies that reach their sales peak approximately 3 to 6 weeks after their initial launch. More flexibility can be obtained from the use of Gamma distributions that can describe both pattern types simultaneously. The models proposed by Sawhney and Eliashberg (1996) and Ainslie et al. (2003) are based on generalized Gamma distributions.

Our study proposes a family of diffusion models whose goal is to forecast later-week revenues very soon (i.e., within 2–3 days) after a movie's initial release. As we discuss later, our model family is a novel variant of the well-known Bass diffusion model that captures some of the unique properties of the motion picture industry: the fact that marketing declines rapidly after a movie's release and the fact that most moviegoers talk less about movies they have watched in the distant past than about movies they have watched recently. The novelty of our contribution lies both in the form of our models as well as in examining the extent to which various metrics of online reviews can complement more traditional explanatory variables, such as marketing, theater availability, professional critic reviews and early sales.

Methodologies for Measuring Consumer Word-of-Mouth

Traditional attempts to measure consumer word-of-mouth (WOM) are based on inference, surveys, and controlled experiments. For example, Bass (1969) and those who have extended his model typically use aggregated sales data to infer the model's coefficient of internal influence, which, in turn is *assumed* to relate to WOM. As another example, Reingen et al. (1984) conduct a survey of the members of a sorority in which they compare brand preference congruity between women that lived in the same house and those that did not. They find that those that lived together had more congruent brand preferences than those that did not. The study then *infers* that those that lived together had more opportunities for interaction and thus, that WOM communication was more prevalent.

Surveys remain the most popular method to study WOM, largely because individuals can be asked directly about their communication habits; the error then lies in the self-reporting of behavior. Several well-known studies, such as Bowman and Narayandas (2001), Brown and Reingen (1987), Reingen and Kernan (1986), and Richins (1983), base their analyses on proprietary surveys designed to test a specific hypothesis related to WOM.

Laboratory experiments are another popular method for inferring properties of WOM (see Borgida & Nisbett, 1977; Herr et al., 1991 as two representative examples of a large literature). The issue with experiments is the extent to which properties identified in a controlled setting generalize to larger, real-world settings.

The advent of the Internet introduced a new technique for measuring consumer WOM: directly through online discussion groups, online review forums, and other forms of user-generated online content. Researchers can easily gather large amounts of data from such sources. Nevertheless, sound methodological principles for analyzing such data are still in the process of being established.

Previous research has looked at unstructured online discussion forums and online product reviews and has used *volume, valence,* and *dispersion* when examining consumer postings. The theory behind measuring volume is that the more consumers discuss a product, the higher the chance that other consumers will become aware of it. The theory behind measuring valence, or consumer attitude, is that positive opinions will encourage other consumers to adopt a product whereas negative opinions will discourage them. The theory behind measuring dispersion, or the spread of communication across communities, is that opinions spread quickly within communities, but slowly across them (Granovetter, 1973). Ideas and opinions that exhibit strong dispersion across communities are thus likely to have substantial staying power.

The results of early studies on the explanatory power of the above metrics have been somewhat inconsistent with these theoretical predictions, as well as with one another. Godes and Mayzlin (2004) have looked at how metrics of Usenet conversations about television shows relate to their Nielsen (viewership) ratings. They find that, whereas the dispersion of conversations among different newsgroups has significant explanatory power, the associated volume of postings does not. Liu (2006) studied the impact of Yahoo! Movies prerelease message board discussions on motion picture box office revenues. Somewhat surprisingly, he finds that, whereas the volume of online conversations has explanatory power, their valence does not. Duan et al. (2005) looked at the relationship between daily Yahoo! Movies reviews and box office sales. They similarly find that the volume, but not the valence, of movie ratings has explanatory power.

In this study, we reconcile some of the inconsistencies of previous studies with respect to what online review metrics are significant predictors of future sales. Specifically, we show that, when used in the context of models that more closely capture the properties of the entertainment industry, the volume, valence, and dispersion of online movie reviews are all statistically significant in predicting future sales in directions that are consistent with theoretical predictions.

DATA SET

Data Collection Methodology

Data for this study were collected from Yahoo! Movies (movies.yahoo.com), BoxOfficeMojo (www.boxofficemojo.com) and the Hollywood Reporter (www.hollywoodreporter.com). From Yahoo! Movies, we collected the names of all movies released during 2002. For the purpose of our analysis, we excluded titles that were not released nationwide in the United States and not released in theaters (e.g., DVD releases). For each of the remaining titles, we collected detailed review information, including all professional critic reviews (text and letter ratings, which we converted to a number between 1 and 5), and all user reviews (date and time of review, user id, review text, integer rating between 1 and 5).

We used Boxofficemojo to obtain weekly box office and marketing expense data; we excluded movies for which this data was incomplete. Finally, we used the Hollywood Reporter's Star Power 2002 report to construct a proxy of each movie's *star power*. Based on surveys that are distributed to a panel of industry executives, Hollywood Reporter publishes an annual report that rates each actor's global bankability on a scale from 0–100. Actors rated in the interval 87.50–100 are considered to have *maximum* star power, whereas actors rated in the interval 62.50–87.49 are

TABLE 2 Key Summary Statistics of Our Data Set

VARIABLE	MIN	MEAN	MAX
Box office (aggregate; in millions)	2.5	68.1	403.7
Production Budget (in millions)	2	46.1	140
Marketing Budget (in millions)	2	24.3	50
Exhibition longevity (in weeks)	3	14	51
Screens in opening week	4	2,393	3,615
Volume of total user ratings	67	689	6,295
Volume of first week user ratings	2	312	3,802
Volume of critic ratings	7	13	20
Average aggregate user rating (range 1–5)	1.9	3.4	4.4
Average critic rating (range 1–5)	1.4	3.1	4.6
Total number of movies	80		
Total number of user ratings	55,156		
Total number of critic ratings	1,040		
Total number of unique users	34,893		

considered to have *strong* star power. Most household actor names fall in these two categories. The Hollywood Reporters report has been used by other researchers as a proxy of an actor's star power (Elberse & Eliashberg, 2003; Ainslie et al., 2003).

Our final data set consists of 80 movies with complete production, marketing, weekly box office, critic reviews, and daily user review data. The final movie sample was found to have similar overall profile with the full set of nationally released 2002 movies (in terms of genre, budget, and marketing), ensuring that no bias was introduced by considering only a subset of movies. It consists of 1,188 weekly box office data, 1,040 critic reviews (an average of 13 reviews per movie), and 55,156 user reviews from 34,893 individual users (an average of 689 reviews per movie and 1.5 reviews per user). Table 2 provides some key summary statistics.

Demographics of Online Reviewers

We were able to collect partial rater demographic data by mining the user profiles that are associated with the reviewers' Yahoo IDs. About 85% of reviewers in our data set listed their gender and 34% their age. From that information, we constructed an estimate of the demographic profile of the Yahoo! Movies reviewer population (Table 3). We found that the demographic breakdown of online reviewers is substantially skewed relative to that of U.S. moviegoers. Most notably, a disproportionately high percentage of online reviews were provided by young males under 30.

TABLE 3 Estimated Demographic Profile of Yahoo! Movie Reviewers

AGE	2002 YAHOO! MOVIE RATERS	2001 US MOVIEGOERS*
<18	13%	15%
18–29	58%	35%
30–44	23%	28%
45+	6%	22%
GENDER		
Men	74%	49%
Women	26%	51%

*Source: Newspaper Association of America (NAA)

Relationship between User and Professional Reviews

Because much work has been done on using critic reviews to predict movie revenue (Eliashberg & Shugan, 1997; Reinstein & Snyder, 2005; Basuroy, Chatterjee, & Ravid, 2003), it is natural to ask how well user ratings correlate with critic ratings. Table 4 reports the correlation between critic and user ratings.

TABLE 4 Correlation of Critic and User Ratings

	RATERS		
	ALL*	MALE	FEMALE
First week	0.63	0.61	0.46
Second week	0.58	0.57	0.53
Third week	0.53	0.46	0.45
All weeks	0.59	0.58	0.49

*Includes male raters, female raters and raters who do not specify their gender.

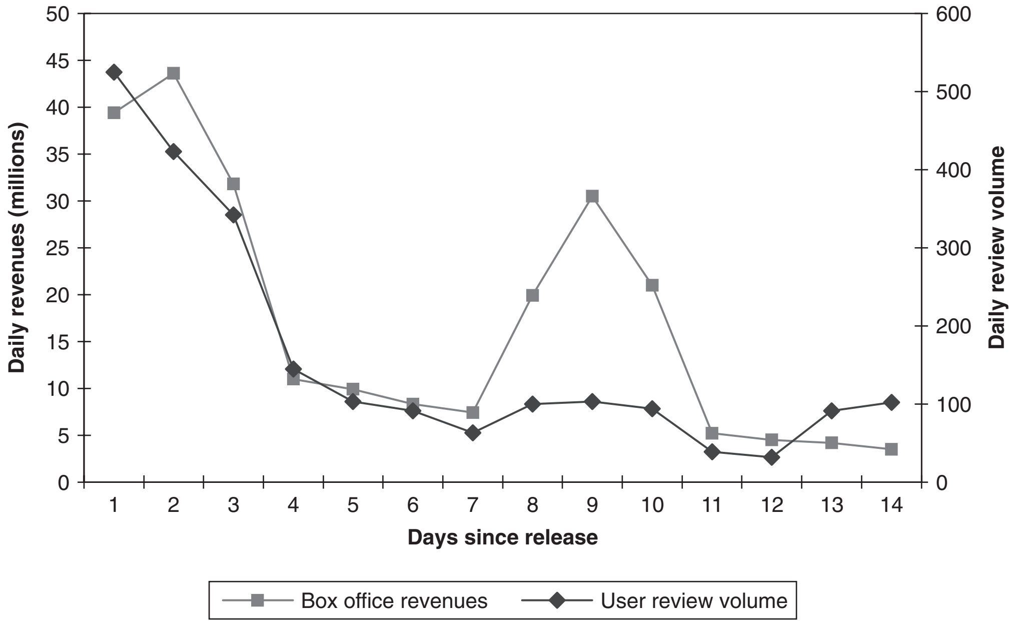

FIGURE 1

Daily Volumes of Sales and Online Reviews During the First Two Weeks of *Spider-Man*.

The relatively low correlation between user and critic ratings emphasizes the importance of examining user reviews as a predictive tool, as the information provided by users appears to be complementary to the information provided by professional movie critics.

Dynamics of Review Volume

Online reviews are (at least in principle) contributed by people who have watched the movies being rated. It is, thus, expected that their early volume will exhibit a strong correlation with the corresponding box office revenues. Figure 1 confirms this for *Spider-Man*. Most movies in our data set exhibit very similar patterns. The correlation between total weekly box office and total weekly volume of reviews for all movies in our data set is 0.80. This suggests that the volume of online reviews could serve as a proxy of sales, an observation that we put into good use later on in our models.

INDEPENDENT VARIABLES

Table 5 lists all independent variables used in subsequent analyses. We divide our independent variables into the following subsets:

Genre and MPAA Ratings. Several papers have included the genre of a film as a control variable (Austin & Gordon, 1987; Litman, 1983; Litman & Ahn, 1998). We collected the genre description from Yahoo! Movies and code a movie's genre using seven dummy variables (*Sci-Fi*, *Thriller*, *Children*, *Romance*, *Comedy*, *Action*, *Drama*). Ravid (1999) found MPAA ratings to be significant variables in his regressions. We code MPAA ratings using five dummy variables (*G*, *PG*, *PG13*, *R*, and *NR*).

Prerelease Marketing and Availability. Several authors have shown that the advertising and theater availability of a film is significantly related to its box office performance (Litman, 1983; Litman & Kohl, 1989; Litman & Ahn, 1998; Ravid, 1999; Elberse & Eliashberg, 2003). Accordingly, we include a movie's prerelease marketing budget (*MKT*) and number of opening weekend screens (*SCR*) to our variable list.

Star Power. The presence of well-known stars has been shown to influence motion picture revenues (De Vany & Walls, 1999; Ravid, 1999). We use a dummy variable (*STAR*) to indicate movies that feature one or more actors that were rated as having *maximum* or *strong* star power in Hollywood Reporter's Star Power 2002 report.

TABLE 5 List of Independent Variables

Production, marketing, star power and distribution strategy	
MKT	Estimated prerelease marketing costs (in millions of $)
SCR	Number of theaters (screens) in opening week
STAR	Dummy variable indicating the presence of one or more well-known star(s)
SLEEPER	Categorical variable indicating if movie is sleeper or wide-release
MPAA Rating (dummy variables)	
G, PG, PG13, R, NR	
Genre (dummy variables)	
SCIFI, THRILLER, COMEDY, ROMANCE, DRAMA, ACTION, KIDS	
Professional Critic Reviews	
CRAVG	Arithmetic mean of professional critics reviews
User Reviews	
AVG	Arithmetic mean of user ratings posted during opening weekend
TOT	Total number of user ratings posted during opening weekend
AENTR	Entropy of age group distribution of opening weekend raters
GENTR	Entropy of gender distribution of opening weekend raters
Box office data	
BOX	Box office revenues during opening weekend (in millions of $)

Release Strategy. Most movies are released using one of two distinct strategies. Wide-release or "blockbuster" movies (such as *Star Wars*) open simultaneously in large numbers of theaters worldwide and are accompanied by intensive prerelease marketing campaigns. Revenues for such movies typically peak during the first weekend and exhibit a steady decline in subsequent weeks. "Sleeper" movies (such as *My Big Fat Greek Wedding*) are initially released in small numbers of theaters with modest marketing campaigns and rely on word-of-mouth for growth. Revenue streams for such movies typically increase for several weeks before they start to decline. Given the different growth patterns of these two movie categories, it is reasonable to expect that release strategy will have an important impact on a movie's revenue trajectory. We use a dummy variable (*SLEEPER*) to distinguish between the two classes of movies in our sample. Consistent with industry practice, we classify a movie as a "sleeper" if its number of opening weekend screens is less than 600.

Professional Critics. An important objective of our study is to compare the relative predictive power of professional critics and user ratings. Accordingly, we include the arithmetic mean (*CRAVG*) of the numerical equivalent (see discussion of Data Set) of all professional critic ratings published by Yahoo for each movie.

User Reviews. Past work on online word-of-mouth has considered the relationship of the *volume, valence,* and *dispersion* of online conversations to product revenues (Godes & Mayzlin, 2004; Duan et al., 2005; Liu, 2006). Our list of independent variables, similarly, includes proxies of these three metrics of online reviews. We use the total number of posted reviews during the first three days of a movie's release (*TOT*) as our measure of volume. We base our measures of valence on the arithmetic mean (*AVG*) of ratings posted during the same period. Finally, we use the *entropy* of the (self-reported) gender (*GENTR*) and age (*AENTR*) distribution of each movie's opening weekend reviewers as our measure of moviegoer dispersion across different gender and age groups.[1]

Early box office revenues. The objective of our model is to forecast a movie's revenue trajectory from early (opening weekend) box office and online review data. Accordingly, we include a movie's opening weekend box office revenues (*BOX*) in our list of independent variables.

MODELS

Because one of our objectives is to assist movie exhibitors in better managing supply (i.e., the number of screens on which a movie is exhibited each week), we are interested in forecasting a movie's revenue trajectory in all future weeks. In common with most models of new product sales growth (Mahajan et al., 1990; Meade, 1984), our model is based on a hazard rate formulation. The hazard rate of product adoption is the instantaneous probability that a representative consumer who has not yet adopted a

[1]Given a population whose members are distributed among a finite number of disjoint classes $i = 1, \ldots, N$ with respective probabilities p_i, entropy, defined as $H = -\Sigma_i p_i log p_i$, represents a measure of population diversity with respect to that classification. Entropy is minimized if all members of the population belong to the same class. On the other hand, entropy is maximized if the population is evenly distributed among all classes.

(durable) product will do so at time t. Assuming that the size of the population is fixed, if $F(t)$ denotes the cumulative fraction of adopters at time t and $F'(t)$ denotes its derivative with respect to time (i.e., the instantaneous rate of adoption at time t), the hazard rate of adoption is defined as:

$$h(t) = \frac{Pr[\text{adopts at time } t]}{Pr[\text{adopts at time } \tau \geq t]} = \frac{F'(t)}{1 - F(t)} \quad (1)$$

If the population size is N and the purchase price is p, the total market size, M, is given by $M = Np$. From equation (1), the evolution of cumulative revenues $R(t) = MF(t)$ is then governed by the following differential equation:

$$R'(t) = (M - R(t))h(t) \quad (2)$$

From a theoretical perspective, hazard rate models have been shown to provide good approximations of the aggregate outcome of a large number of individual-level stochastic product adoption processes (Chatterjee & Eliashberg, 1990). From a practical perspective, most growth curves used in sales forecasting by practitioners can be derived from equation (2) by assuming different functional forms for the hazard rate $h(t)$. For example, a constant hazard rate $h(t) = a$ gives rise to an exponential curve, whereas a monotonically increasing or decreasing hazard rate $h(t) = \alpha t^b$ gives rise to a Weibull distribution. The well-known Bass model (Bass, 1969) also arises as a special case of (2) if we set $h(t) = P + QF(t)$. A common interpretation of the Bass model is that product adoption is driven by two forces: an "external" force, that typically relates to advertising and product availability, and is represented by the coefficient P, and an "internal" force that relates to word-of-mouth, and is represented by the coefficient Q multiplied by the cumulative number of past adopters $F(t)$.

Our model is based on a novel hazard function that bears a relationship to the Bass model but captures more precisely the properties of the movie industry. Just as in the Bass model, we assume that the probability that a nonadopter adopts at time t is driven by an "external" force P that relates to advertising and product availability and an "internal" force Q that relates to word-of-mouth from past adopters. In the movie industry, most advertising and publicity occurs just before a movie's premiere and declines rapidly postrelease. Elberse and Anand (2005) report that the highest median TV advertising spending occurs immediately before a movie's opening weekend; it drops to less than 30% of its peak value in the following week and to less than 10% in later weeks. Most movies, thus, get an initial publicity "jolt" that diminishes in later weeks. We incorporate this in our model by multiplying the "external" force coefficient P by a *discount factor* δ^t $(0 \leq \delta \leq 1)$ that diminishes as the movie moves further away from its initial release. Furthermore, consistent with previous research, we assume that word-of-mouth is localized in time: people talk more about movies immediately after watching them, and less as time goes by. Eliashberg et al. (2000) recognize and explicitly take this phenomenon into consideration in their MOVIEMOD prerelease forecasting model. Elberse and Eliashberg (2003) also implicitly incorporate the "perishability" of word-of-mouth in their model by using a word-of-mouth proxy variable that is based only on previous-period (rather than cumulative) data. We incorporate the time-locality ("perishability") of word-of-mouth in our model by multiplying the internal force Q by a time-discounted integral of past adopters $\int_{\tau=0}^{t} F'(t-\tau)\epsilon^{\tau} d\tau$ $(0 \leq \epsilon \leq 1)$ as opposed to the cumulative sum of past adopters $F(t)$. The resulting hazard function has the following form:

$$h(t) = P\delta^t + Q \int_{\tau=0}^{t} F'(t-\tau)\epsilon^{\tau} d\tau \quad 0 \leq \delta \leq 1,\ 0 \leq \epsilon \leq 1 \quad (3)$$

Substituting into (2) and recognizing that $F(t) = R(t)/M$ we obtain our revenue forecasting equation:

$$R'(t) = (M - R(t))\left(P\delta^t + \frac{Q}{M}\int_{\tau=0}^{t} R'(t-\tau)\epsilon^{\tau} d\tau\right) \quad (4)$$

Observe that equation (4) reduces to the Bass equation when $\delta = \epsilon = 1$. However, our model exhibits *qualitatively different* behavior when the discount factors δ, ϵ are strictly less than one. Specifically, the Bass model assumes that, as time moves forward, the "external" force P remains undiminished whereas the "internal" force $QF(t)$ monotonically increases as the total fraction of adopters grows. This results in a monotonically increasing hazard rate and an adoption curve that is only limited by market saturation. In contrast, our model assumes a decaying external force and an internal force that is multiplied by the number of *recent* adopters only. The resulting hazard rates are, thus, either monotonically declining or inverse U-shaped (first increasing and eventually declining).

Similarly to Elberse and Eliashberg (2003), we adopt the simplifying assumption that the recent adopters who contribute to a movie's internal force of adoption are equal to last week's adopters. Under this simplifying assumption, equation (4) has the following discrete-time formulation that is easier to estimate from weekly box office data:

$$\begin{aligned} Y_{it} &= (M_i - R_{i(t-1)})\left(P_i\delta_i^t + \frac{Q_i}{M_i}Y_{i(t-1)}\right) \\ R_{it} &= Y_{it} + R_{i(t-1)},\ t = 1, 2, \ldots \\ \text{and}\ \ R_{i0} &= Y_{i0} = 0 \end{aligned} \tag{5}$$

where

Y_{it} are the box office revenues of movie i during week t since its initial release

R_{it} are the cumulative revenues of movie i up until (and including) week t

P_i,Q_i are the coefficients of external and internal influence of movie i

M_i is the market potential of movie i

δ_i is movie i's external influence discount factor

Model (5) has 4 movie-specific unknown parameters (P_i, Q_i, M_i, and δ_i). Given estimates of these 4 parameters, it is straightforward to see how, beginning from $R_{i0} = Y_{i0} = 0$, successive application of equations (5) can produce estimates of a movie's weekly box office revenues Y_{it} and cumulative revenues R_{it} for all $t = 1, 2, \ldots$

Given a training set of movies with known weekly box office revenues, production, marketing, and reviews data, the construction of a revenue forecasting model on the basis of equation (5) is based on a hierarchical formulation: On the first level, we estimate equation (5) and, on the second level, we augment the model with a set of movie-specific covariates for parameters P_i, Q_i, M_i, δ_i. The model estimation procedure is discussed later.

To forecast future box office revenues of a new movie j, we reverse the process: First, we feed the new movie's data to the regression equations of our model's second level to derive estimates of P_j, Q_j, M_j, δ_j. Then we apply equations (5) successively to produce forecasts of the new movie's weekly revenues at any desired future point in time.

MOVIE-SPECIFIC COVARIATES

An important step of our modeling technology is the construction of linear regression equations that can be used to estimate a movie's market potential M_i, parameters of external and internal influence P_i,Q_i and discount factor δ_i from appropriate subsets of the available covariates. It is generally agreed that parsimony is a desirable property in forecasting models since overfitting inflates the variance of the prediction error (Box & Jenkins, 1970). Accordingly, we propose a set of regression equations that only include independent variables that are expected to be significant on the basis of our knowledge of the motion pictures domain. Table 6 summarizes the models, indicating which variables are included in which equation. The following paragraphs explain the rationale behind each model.

Market Potential

A movie's market potential M_i captures the (theoretical) maximum revenue that a particular movie can hope to generate if it is played in theaters forever.[2] We hypothesize that a movie's market potential is a function of its genre, MPAA rating, and supply factors such as its initial theater availability (*SCR*) and release strategy (blockbuster or sleeper). These assumptions produce equation M_I (Table 6). For benchmarking purposes, we also construct an alternative equation, labeled M_{II}, that does not make use of theater availability. Equation M_{II} enables revenue

[2]This potential is purely theoretical because most movies are taken out of theaters before they exhaust their full potential and are released again in secondary markets such as DVD sales, rentals, TV broadcasts, and so on.

TABLE 6 Independent Variables Included in Regression Equations that Predict Movie-Specific Parameters P, Q, M, δ

	M_I	M_{II}	P_I	P_{II}	P_{III}	P_{IV}	Q_I	Q_{II}	$logit(\delta_I)$	$logit(\delta_{II})$
Intercept	X	X	X	X	X	X	X	X	X	X
PG	X	X					X	X		
PG13	X	X					X	X		
R	X	X					X	X		
NR	X	X					X	X		
SciFi	X	X					X	X		
Kids	X	X					X	X		
Drama	X	X					X	X		
Comedy	X	X					X	X		
Romance	X	X					X	X		
Action	X	X					X	X		
STAR				X	X					
MKT				X	X					
SCR	X									
SLEEPER	X	X		X	X	X	X	X	X	X
CRITIC				X	X			X		X
TOT				X		X				
AVG						X	X		X	
AENTR				X		X	X			
GENTR				X		X	X			
BOX			X							

forecasting in situations where precise theater counts are not available.

External Influences

Parameter P_i intends to capture the "external" factors that induce a person to watch a movie. Such factors include marketing and publicity, critic reviews and unobservable attributes such as the attractiveness of a movie's plot, the quality of its trailer, and so on. These factors give each movie an initial thrust that is most important in shaping its early revenues and diminishes in later weeks. In fact, for $t = 1$, equation (5) gives $P_i = Y_{i1}/M_i$, that is, the coefficient of external influence is mathematically equal to a movie's opening weekend revenues over the movie's market potential. We therefore expect that, if a movie's opening weekend box office revenues (*BOX*) are available, they will be an excellent predictor of coefficient P_i. This assumption produces equation P_I (Table 6).

To enable forecasting in settings where box office revenues are not available we propose an alternative regression equation, labeled P_{II} , that uses the volume of opening weekend online reviews (*TOT*) as a proxy of opening weekend box office revenues. We have previously established that the volume of online reviews exhibits very high correlation with box office revenues (see Figure 1 and associated discussion). To increase forecasting precision, we also include additional factors that have been shown to influence a movie's initial box office success such as marketing (*MKT*), professional critic reviews (*CRAVG*), and the presence of well-known stars (*STAR*). We also include the age and gender entropy metrics of online reviews (*AENTR*, *GENTR*); we hypothesize that the latter might be significant because they provide indications of how broadly a movie appeals to the general population.

For benchmarking purposes, we construct two more regression equations: equation P_{III} that only includes marketing and professional critic review data (no box office or online review data) and equation P_{IV} that only includes online review data (no box office, marketing, or professional critic review data).

Internal Influences

Parameter Q_i intends to capture the "internal" factors that induce a moviegoer to watch a movie. In our context these factors are primarily related to word-of-mouth from other consumers. The following arguments justify the variables we chose to include in our regression equation for coefficient Q_i:

- Word-of-mouth makes unaware consumers aware of a new movie; the awareness-building function of word-of-mouth is particularly important for sleeper movies. We therefore hypothesize that Q_i will be higher for such movies.
- Word-of-mouth communicates previous moviegoers' assessment of a movie's quality and might encourage or discourage some prospective moviegoers from watching the movie. We, thus, expect that coefficient Q_i will be positively related to the valence of user reviews (*AVG*).
- We expect coefficient Q_i to be positively related to the entropy measures of online reviews (*AENTR*, *GENTR*). These measures capture the age and gender heterogeneity of the population that watched a given movie. Under the assumption that consumers tend to discuss movies with people of similar age and gender, consistent with the arguments brought forth by Godes and Mayzlin (2004), we expect that the more heterogeneous the population, the wider the audience of previously uninformed consumers that an initial set of moviegoers is likely to inform through word-of-mouth.
- MPAA ratings may be significant in terms of predicting what fraction of the population is likely to react to a given word-of-mouth stimulus. For example, positive word-of-mouth about an R-rated movie from a family friend is unlikely to induce parents and children to watch it, whereas positive word-of-mouth about a PG movie from the same source is likely to persuade the entire family to go to the movies.
- Finally, given the skewed demographic distribution of online raters, we control for any systematic differences between the taste of the population of online raters and the population at large by including the genre dummies.

The resulting equation is labeled Q_I. For benchmarking purposes, we construct an alternative equation, labeled Q_{II} , that does not include any online review metrics.

Discount Factor

Discount factor δ_i captures the rate by which a movie's external influences (publicity and marketing) decay after its initial release. From equation (5) a higher δ_i implies slower decay. Because factor δ_i is constrained to lie between 0 and 1, we construct a linear regression model for $logit(\delta_i)$. It is plausible to assume that $logit(\delta_i)$ will be positively related to the quality of a movie (*AVG*) as perceived by consumers: studios tend to sustain advertising and publicity for well-received movies and tend to cut their losses quickly for less-well received movies (Mahajan et al., 1984). It is also plausible to assume that a movie's release strategy (blockbuster or sleeper) may affect the discount factor: whereas marketing of blockbuster movies declines sharply postrelease, sleeper movies start with low marketing that is sustained or even increased in later weeks. The equation that results from these assumptions is labeled δ_I. For benchmarking purposes, we construct an alternative equation, labeled δ_{II}, that uses *CRAVG* instead of *AVG*; the latter can be used in settings where no online review metrics are available.

ESTIMATION AND RESULTS

We estimate model (5) in one step using an MCMC (Markov Chain Monte Carlo) procedure. Specifically, we assume that each movie's observed weekly box office data y_{it} are noisy observations drawn from (5) that incorporate multiplicative log-normal noise, that is, $y_{it} = Y_{it}\varepsilon_{it}$ where $log(\varepsilon_{it}) \sim \text{Normal}(0, \sigma_i^2)$. The first-level model is, thus, of the form:

$$log(y_{it}) \sim \text{Normal}\left(log\left[(M_i - R_{i(t-1)}) \times \left(P_i\delta_i^t + \frac{Q_i}{M_i}Y_{i(t-1)}\right)\right], \sigma_i^2\right),$$
$$\text{for } i = 1, \ldots, N, t = 1, \ldots, T_i \qquad (6)$$

where we use i to index movies and t to index weeks since release (T_i is the total number of weeks that movie i was shown in theaters). The posterior density of M_i, P_i, Q_i, δ_i, σ_i given the vector of weekly box office data observations $\mathbf{y}$ is simply

$$p(M_i, P_i, Q_i, \delta_i, \sigma_i|\mathbf{y}) \propto \prod_{\mathbf{t=1}}^{\mathbf{N}} \prod_{t=1}^{T_i} N\left(log(y_{it})|log\left[(M_i - R_{i(t-1)})\left(P_i\delta_i^t + \frac{Q_i}{M_i}Y_{i(t-1)}\right)\right], \sigma_i^2\right) \tag{7}$$

where $N(\cdot|\mu, \sigma)$ represents the normal density function with mean μ and standard deviation σ. As usual we assume a "flat" prior distribution on M_i, P_i, Q_i, δ_i and σ_i in the simulations.

Our two-level hierarchical model can then be written as a sequence of conditional distributions:

$$\begin{aligned} &y_{it} \mid P_i, Q_i, M_i, \delta_i, \sigma_i^2 \\ &P_i \mid \mathbf{X}_i^P, V_P \\ &Q_i \mid \mathbf{X}_i^Q, V_Q \\ &M_i \mid \mathbf{X}_i^M, V_M \\ &\delta_i \mid \mathbf{X}_i^\delta, V_\delta \\ &\sigma_i^2 \mid \mu_\sigma, V_\sigma \end{aligned} \tag{8}$$

where $\mathbf{X}_i^A$, V_A, $A \in \{P, Q, M, \delta\}$ are the covariates drawn from Table 6 and the corresponding variances, respectively.

We estimate model (8) using MCMC with Gibbs sampling. To generate posterior estimates of the model parameters, we sample each model parameter from its prior distribution conditional on the data and the current values of all other parameters. For each model reported in this paper, we simulated 15,000 MCMC iterations with 5,000 iterations in the burn-in period and 10,000 iterations in the estimation period. We ensure the stability of the parameter estimates by examining the convergence of multiple chains through trace plots and the Gelman-Rubin convergence statistics.

Although the main thrust of the paper is prediction of sales in a holdout sample of movies, it is instructive to take a brief look at the statistical significance and signs of the coefficients for the purpose of validating the theoretical assumptions that led to the construction of our models. We fit 4 different models, each model using different sets of regression equations for estimating P_i, Q_i, M_i, δ_i. The equations used by each model are listed in parentheses next to the model's name below (see Table 6 for the lists of variables included in each regression equation).

Model A (P_I, Q_I, M_I, δ_I) is meant to be used in situations where marketing, theater count, professional critic review, early box office and online review data are available. It is expected to provide the best overall forecasting performance.

Model B (P_{II}, Q_I, M_I, δ_I) should be used in situations where early box office data are not available. It highlights the use of the volume of early online reviews as a proxy of early sales and, thus, allows even earlier revenue forecasting.

Model C (P_{III}, Q_{II}, M_I, δ_{II}) is similar to model B but does not use any online review data. In conjunction with Model B it is meant to serve as a technical benchmark of the incremental forecasting precision that can be achieved through the use of online review metrics relative to more traditional metrics such as marketing budget, theater counts, and professional critic reviews.

Model D (P_{IV}, Q_I, M_{II}, δ_I) relies exclusively on online review data (i.e uses no marketing, theater count, professional critic review or box office data). It is, similarly, meant to provide a technical benchmark on the forecasting accuracy that can be achieved through the exclusive use of publicly available and easy to obtain online review metrics.

Appendix A reports the results of fitting our models to our entire data set. For each model, we list the posterior mean and standard error of all regression equation coefficients. Boldface indicates coefficients that were found to be statistically significant at the 5% level (i.e., coefficients whose 95% posterior estimate confidence intervals do not include zero). Consistently with a number of prior forecasting studies (Ainslie et al., 2003; Lee et al., 2003; Moe & Fader, 2002), we assess each model's goodness of fit by reporting the mean absolute percentage error (MAPE) between the observed and estimated cumulative weekly revenues on any given week. We also report each model's Akaike Information Criterion (AIC).

As expected, Model A, which makes use of early box office revenue and all other covariates, has the lowest MAPE (8.5%). It is followed by Model B (covariates: marketing, theater counts, critic reviews, user reviews; MAPE = 14.5%), Model D (covariates: user reviews; MAPE = 17.7%) and Model C (covariates: marketing, theater counts, critic reviews; MAPE = 20%). The information content of online review metrics (*TOT*, *AVG*, *GENTR*, *AENTR*) can be appreciated by comparing the MAPE of Model B to that of Model C. We see that the removal of online review metrics from Model B increases the model's MAPE by 38%. In contrast, comparison of Models B and D shows that the removal of marketing (*MKT*), theater count (*SCR*) and critic review (*CRAVG*) metrics from Model B increases the model's MAPE by only 22%.

Most coefficient signs and significance levels are consistent with our hypotheses. In summary:

- As hypothesized, our models indicate that a movie's market potential M_i has a significant positive relationship with the number of opening week theaters (SCR) and a significant negative relationship with *SLEEPER* movies.
- In Model A coefficient P_i is almost perfectly predicted from first weekend box office revenues (*BOX*).
- If box office revenues are not available (Models B, C, and D), consistent with our hypotheses we find that P_i has a significant positive relationship with a movie's online review volume (*TOT*), gender entropy of online reviewers (*GENTR*), average valence of critic reviews (*CRAVG*), and the presence of a well-known star (*STAR*). Interestingly, a movie's marketing budget (*MKT*) was only significant in Model C (i.e., the model that does not include online review metrics) and lost significance in Model B that also includes online review metrics. We interpret this finding as suggesting that the information content of online review metrics subsumes whatever information is contained in *MKT*.
- Coefficient Q_i has a significant positive relationship with a movie's online review valence (*AVG*) and gender entropy of online reviewers (*GENTR*); as we hypothesized it is also significantly higher for *SLEEPE*R movies.
- Discount factor δ_i has a significant positive relationship with the average valence of user reviews (*AVG*) and a positive but not significant relationship with *SLEEPE*R movies.

The relationships of MPAA ratings and movie genres with our models do not follow a clean pattern. Signs and significance levels differ across models, perhaps as a result of correlations of these covariates with other aspects of our model. The most consistent pattern across all four models is the negative influence of a "restricted" (R) MPAA rating to a movie's market potential M_i and coefficient Q_i, a fact that is well known to the movie industry.

Our age entropy metric (*AENTR*) did not turn out to be significant in any of our models. This is probably because only 34% of reviewers in our data indicated their age. Our age entropy metric is, thus, most likely too noisy to be significant and will be dropped from all subsequent analyses.

FORECASTING ACCURACY

To test the forecasting accuracy of our models, we follow a k-fold cross-validation procedure (Efron & Tibshirani, 1993), which is an improvement over the simple holdout method. Specifically, we randomly divide our data set into $k = 8$ subsets, each containing 10 movies. We then repeat the holdout method eight times for each of our four models. Each time, one of the eight movie subsets is used as the test set and the other seven subsets are put together to form a training set. The advantage of the k-fold method is that it matters less how the data gets divided. Every data point gets to be in a test set exactly once, and gets to be in a training set $k-1$ times. The variance of the resulting estimate is reduced as k is increased. We compute the average MAPE across all 8 trials for each model and report the results in Table 7. Given the substantial differences in the diffusion dynamics of wide release (blockbuster) vs. narrow release (sleeper) movies we also report the forecasting performance of each model separately for blockbuster and sleeper movies. Appendix B lists the detailed MAPE of our four models for each of the 80 movies in our data set.

In the case of blockbuster movies Model A has, by far, the lowest average MAPE (7%) suggesting that early

TABLE 7 Forecasting Accuracy of Our Models

	MODEL A		MODEL B		MODEL C		MODEL D	
	MAPE	SD	MAPE	SD	MAPE	SD	MAPE	SD
All movies	0.10	0.11	0.24	0.26	0.29	0.27	0.33	0.47
Blockbuster movies	0.07	0.06	0.21	0.26	0.29	0.28	0.33	0.50
Sleeper movies	0.34	0.13	0.41	0.24	0.35	0.13	0.33	0.16

box office revenues are the most important indicator of a widely released movie's future revenue trajectory. This makes intuitive sense, because most widely released movies follow "big bang" dynamics that are heavily influenced by their opening weekend performance. If box office data are not available (Model B) the MAPE triples but still remains a very respectable 21%. Comparing the MAPEs of Model B (21%) and Model C (29%) we see that, consistent with the results we obtained during the estimation phase, removal of online review metrics from Model B leads to an 38% increase in forecasting error. This result is a very powerful testament to the value of online review metrics as a complement of traditional metrics, such as marketing, product availability and professional reviews.

All of our models perform less well in the case of sleeper movies. This is not surprising because the fortunes of sleeper movies are less easy to predict from opening weekend results. Despite this recognized diffculty, our models still achieve very respectable MAPEs that range from 33% to 41%. In the case of sleeper movies, Model D (the model that uses only online review metrics) remarkably exhibits the best performance with an average MAPE of 33%. Because the revenues of sleeper movies are heavily influenced by word-of-mouth, this result provides further support for the value of online review metrics in measuring various aspects of consumer word-of-mouth.

Figure 2 provides illustrative comparisons between the weekly revenues of a blockbuster (*Spider-Man*) and a sleeper (*My Big Fat Greek Wedding*) movie and the corresponding forecasts generated by our four models. In the case of *My Big Fat Greek Wedding*, the sleeper movie surprise success of 2002, it is remarkable to observe that, although all four models underestimate actual revenues, they all correctly predict the general shape of these revenues, including the fact that revenues "pick up" between weeks 20 and 30 and reach their peak between weeks 22 and 24.

Our forecasting accuracy results compare favorably with those reported by Liu (2006). Liu uses a linear regression model that relates Yahoo! Movies prerelease message board discussion metrics (and other traditional metrics such as genre, theaters, critic reviews, etc.) to box office revenues. He uses a data set of 40 movies and a similar cross-validation procedure in which each of the 40 movies is taken in turn as the movie to be predicted, whereas the other 39 movies are used in the calibration process to generate parameter estimates. He reports opening week MAPE of 38% and aggregate revenue MAPE of 47%.

Of the two postrelease motion picture forecasting models that have been reported in academic literature, only the model of Sawhney and Eliashberg (1996) is directly comparable to ours.[3] Sawhney and Eliashberg (1996) developed and tested BOXMOD-I, a model for forecasting the gross revenues of motion pictures based on their early box-office data. They tested how the forecasting accuracy of their model improves as more box-office data becomes available

[3]The model of Neelamegham and Chintagunta (1999) focuses on predicting first-week viewership for movies that are introduced sequentially in different markets (e.g., different countries). They use postrelease data from one market in order to predict the movie's performance in another market. Their objective, thus, is different from ours: our model uses early box office and user review metrics to predict a movie's future performance in the same market.

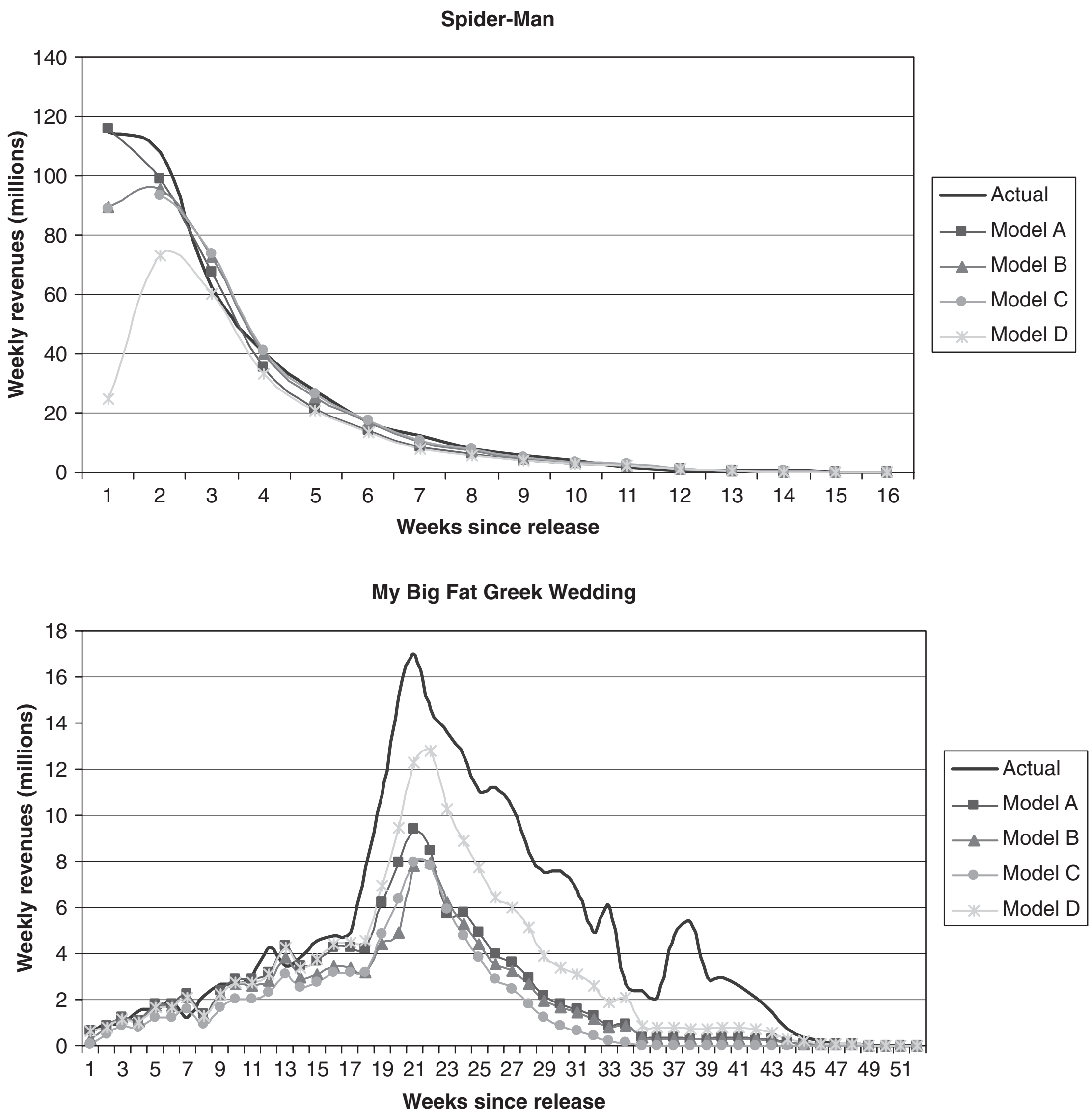

FIGURE 2

Actual Versus Predicted Weekly Revenues for Two Illustrative Movies.

and reported MAPE of 71.1%, 51.6%, 13.2%, 7.2%, and 1.8% when using no box-office data, 1 week of data, 2 weeks of data, 3 weeks of data, and all available box-office data, respectively. Using only 3 days of box-office and user and critic review metrics, our Model A achieves levels of forecasting accuracy (average MAPE of 10%) for which BOXMOD-I requires 3 weeks of box office data. This comparison reinforces our original hypothesis that the use of online review metrics enables reliable forecasts of the impact of a new experience good to be made much faster than with older methodologies. Because BOXMOD-I does not incorporate covariates, our result should be interpreted as evidence for explanatory power of online review metrics rather than as a statement about the power of the underlying behavioral model on which BOXMOD-I is based.

MANAGERIAL IMPLICATIONS AND RESEARCH OPPORTUNITIES

Product review sites are widespread on the Internet and rapidly gaining popularity among consumers. Previous research has established that metrics of online product reviews have an influence on consumer behavior (Chevalier & Mayzlin, 2006; Senecal & Nantel, 2004) and statistically significant relationships with future sales (Godes & Mayzlin, 2004; Duan et al., 2005; Liu, 2006). This paper contributes to a better understanding of how such metrics can add value to firm decision support. Specifically, our study proposes a novel family of revenue forecasting models that are tailored to the entertainment industry and tests their performance in the context of early (opening weekend) postrelease motion picture revenue forecasting. We estimate the models using metrics obtained from user reviews posted on Yahoo! Movies during the opening weekend of a movie, together with a number of more traditional metrics, such as a movie's marketing budget, theatrical availability, professional critic reviews and very early box office revenues. We show that the forecasting accuracy of a model that combines traditional and online review metrics outperforms that of several previously published postrelease motion picture forecasting models. We also show that the addition of online review metrics to a benchmark model that includes prerelease marketing, theater availability, and professional critic reviews substantially increases its forecasting accuracy.

Online movie reviews are available in large numbers within hours of a new movie's theatrical release. Their use, thus, allows the generation of reliable forecasts much sooner than before. Using online review metrics in conjunction with opening weekend marketing, theater count, professional critic, and box office data, our approach can generate forecasts whose accuracy would require 3 weeks of box office data using older techniques.

The ability to derive early postrelease forecasts of a new movie's performance has traditionally been of value to exhibitors (theater owners). Exhibitor chains need to manage the yield from their exhibition capacity, based on their estimates of demand for movies that they are currently exhibiting. Using such estimates they can adapt the exhibition capacity allocated to a new movie, either by dropping the movie from a theater or by shifting it to a smaller (or larger) screening room. They are, thus, very interested in early forecasts of gross box-office revenues in making their exhibition decisions. Today exhibitors usually commit to exhibit a movie for a minimum of 3 to 4 weeks. However, the increasing volatility of second and later-week revenues (Lippman, 2003) plus the availability of rapid forecasting tools, such as the ones we propose in this paper, might lead the industry to adopt more flexible contracts that allow exhibitors to reevaluate their decisions immediately after the opening week. In addition, the ability to generate reliable forecasts so quickly after a movie's premiere can have important implications for motion picture marketing, allowing movie distributors to fine-tune a movie's postrelease marketing campaign in ways similar to those suggested by Mahajan et al. (1984).

In addition to its contributions to the diffusion literature, our study has produced several new empirical insights related to the use of online product reviews in revenue forecasting.

First, we have shown that the early volume of online reviews can be used as a proxy of early sales. This observation allows revenue forecasting to take place before early box office results are published. Furthermore, it has potentially important implications in industries in which sales data are not publicly available, because it implies that firms can use online review data to generate estimates of their competitors' sales.

Second, we have shown that the average valence of user reviews is statistically significant not only as a predictor of a movie's coeffcient of internal influence (that relates to consumer word-of-mouth) but also as a predictor of the rate of decay of a movie's coefficient of external publicity.

Third, we have shown that the gender entropy of online reviewers has statistical significance in terms of predicting both a movie's initial appeal as well as the impact of word of mouth from previous moviegoers on future sales.

We hope that our work has helped reconcile some of the inconsistencies among previous studies with respect to what online review metrics are statistically significant in forecasting the sales of entertainment goods, demonstrating that significance results that are better aligned with theory can be obtained by embedding online review metrics in models that more closely match the properties of the markets of interest. Specifically, we show that, when used in the context of a nonlinear diffusion model that is specifically tailored to the motion picture industry, the volume, valence, and dispersion of online movie reviews all have a positive and statistically significant relationship with future box office sales.

We conclude by pointing out some opportunities for future research. First, in common with the majority of past work in this area, our models do not incorporate the impact of competition from other movies. Such an enhancement is not possible with our current data set, as we don't have data for all movies playing on all weeks. Second, our objective in this paper is to generate future revenue forecasts from a single, early measurement of box office revenues and online reviews. We thus do not have to worry about potential endogeneity issues associated with the interplay between theatrical availability, consumer word of mouth and revenues (see, for example, Elberse & Eliashberg, 2003). In future work, we plan to examine a model that uses measurements of theaters, revenues, and reviews at multiple points in time to obtain more accurate forecasts; in such a model, endogeneity will be a more important factor, and will be dealt with accordingly. Third, although the current study only looks at motion pictures, the novel aspects of our diffusion models (discount factors of external and internal influence) are also potentially applicable in other entertainment good markets that are characterized by heavy prerelease publicity and word-of-mouth whose intensity is correlated with the time of consumption. It would, thus, be interesting to investigate to what extent our models are applicable in the context of other classes of goods such as video games and music.

REFERENCES

Ainslie, A., Dreze, X., & Zufryden, F. (2003). Modeling Movie Lifecycles and Market Share. Marketing Science, 24, 508–517.

Austin, B., & Gordon, T. (1987). Movie Genres: Toward a Conceptualized Model and Standardized Definition. In B. Austin (Ed.), Current Research in Film: Audience, Economics, and Law, Vol. 3. Norwood, NJ: Ablex Publishing Co.

Bass, F. (1969). A new product growth model for consumer durables. Management Science, 15, 215–227.

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003) How Critical are Critical Reviews? The Box Office Effects of Film Critics, Star Power and Budgets. Journal of Marketing, 67, 103–117.

Borgida, E., & Nisbett, R. E. (1977). The Differential Impact of Abstract vs. Concrete Information on Decisions. Journal of Applied Social Psychology, 7, 258–271.

Box, G., & Jenkins, G. (1970). Time Series Analysis: Forecasting and Control. San Francisco, CA: Holden-Day.

Bowman, D., & Narayandas, D. (2001). Managing Customer-Initiated Contacts with Manufacturers: The Impact on Share of Category Requirements and Word-Of-Mouth Behavior. Journal of Marketing Research, 38, 291–297.

Brown, J. J., & Reingen, P. (1987). Social Ties and Word-of-Mouth Referral Behavior. Journal of Consumer Research, 14, 350–362.

Chatterjee, R., & Eliashberg, J. (1990). The Innovation Diffusion Process in a Heterogeneous Population: A Micromodeling Approach. Management Science, 36, 1057–1079.

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. Journal of Marketing Research, 43, 345–354.

De Silva, I. (1998). Consumer Selection of Motion Pictures. In B. R. Litman (Ed.), The Motion Picture Mega-Industry. Needham Heights, MA: Allyn & Bacon Publishing Inc.

De Vany, A., & Walls, D. (1999). Uncertainty in the Movie Industry: Does Star Power Reduce the Terror of the Box Office? Journal of Cultural Economics, 23, 285–318.

Dodds, J., & Holbrook, M. (1988). What's an Oscar Worth? An Empirical Estimation of the Effect of Nominations and Awards on Movie Distribution and Revenues. In B. Austin (Ed.), Current Research in Film: Audience, Economics, and Law, Vol. 4. Norwood, NJ: Ablex Publishing Co.

Duan, W., Gu, B., & Whinston, A. B. (2005). Do Online Reviews Matter? An Empirical Investigation of Panel Data. Working Paper, University of Texas at Austin.

Efron, B., & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Boca Raton, FL: Chapman and Hall.

Elberse, A., & Anand, B. N. (2005). Advertising and Expectations: The Effectiveness of Prerelease Advertising for Motion Pictures. Working Paper 05–060, Harvard Business School.

Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentually Released Products in International Markets: The Case of Motion Pictures. Marketing Science, 22, 329–354.

Eliashberg, J., & Sawhney, M. S. (1994). Modeling Goes to Hollywood: Predicting Individual Differences in Movie Enjoyment. Management Science, 40, 1151–1173.

Eliashberg, J., & Shugan, S. M. (1997). Film Critics: Influencers or Predictors? Journal of Marketing, 61, 68–78.

Eliashberg, J., Jonker, J., Sawhney, M. S., & Wierenga, B. (2000). MOVIEMOD: An Implementable Decision Support

System for Prerelease Market Evaluation of Motion Pictures. Marketing Science, 19, 226–243.

Faber, R., & O'Guinn, T. (1984). Effect of Media Advertising and Other Sources on Movie Selection. Journalism Quarterly, 61, 371–377.

Godes, D., & Mayzlin D. (2004). Using Online Conversations to Study Word of Mouth Communication. Marketing Science, 23, 545–560.

Granovetter, M. (1973). The Strength of Weak Ties. American Journal of Sociology, 78, 1360–1380.

Jedidi, K., Krider, R. E., & Weinberg, C. B. (1999). Clustering at the Movies. Marketing Letters, 9, 393–405.

Jones, J. M., & Ritz, C. J. (1991). Incorporating Distribution into New Product Diffusion Models. International Journal of Research Marketing, 8, 91–112.

Kerr, P. M., Kardes, F. R., & Kim, J. (1991). Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accesibility-Diagnosticity Perspective. Journal of Consumer Research, 17, 454–462.

Krider, R., & Weinberg, C. (1998). Competitive Dynamics and the Introduction of New Products: The Motion Picture Timing Game. Journal of Marketing Research, 35, 1–15.

Lee, J., Boatwright, P., & Kamakura, W. A. (2003). A Bayesian Model for Prelaunch Sales Forecasting of Recorded Music. Management Science, 49, 179–196.

Lippman, J. (2003). Box-Office Records *Mask* Second-Weekend Drops. The Wall Street Journal, June 6, 2003.

Litman, B. R., & Ahn, H. (1998). Predicting Financial Success of Motion Pictures. In B. R. Litman (Ed.), The Motion Picture Mega-Industry. Needham Heights, MA: Allyn & Bacon Publishing Inc.

Litman, B. R., & Kohl, A. (1989). Predicting Financial Success of Motion Pictures: The 80's Experience. The Journal of Media Economics, 2, 35–50.

Litman, B. R. (1983). Predicting Success of Theatrical Movies: An Empirical Study. Journal of Popular Culture, 16, 159–175.

Liu, Y. (2006). Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Receipts. Journal of Marketing, 70, 74–89.

Mahajan, V., Muller, E., & Bass, F. M. (1990). New Product Diffiusion Models in Marketing: A Review and Directions for Research. Journal of Marketing, 54, 1–26.

Mahajan, V., Muller E., & Kerin, R. A. (1984). Introduction Strategy for New Products with Positive and Negative Word-of-Mouth. Management Science, 30, 1389–1404.

Meade, N. (1984). The Use of Growth Curves in Forecasting Market Development-A Review and Appraisal. Journal of Forecasting, 3, 429–451.

Moe, W. W., & Fader, P. S. (2001). Using Advance Purchase Orders to Forecast New Product Sales. Marketing Science, 21, 347–364.

Neelamegham, P., & Chintagunta, P. (1999). A Bayesian Model to Forecast New Product Performance in Domestic and International Markets. Marketing Science, 18, 115–136.

Ravid, S. A. (1999). Information, Blockbusters, and Stars: A Study of the Film Industry. Journal of Business, 72, 463–492.

Reingen, P., Foster, B., Brown, J. J., & Seidman, S. (1984). Brand Congruence in Interpersonal Relations: A Social Network Analysis. Journal of Consumer Research, 11, 1–26.

Reingen, P., & Kernan, P. (1986). Analysis of Referral Networks in Marketing: Methods and Illustration. Journal of Marketing Research, 23, 370–378.

Reinstein, D. A., & Snyder, C. M. (2005). The Influence of Expert Reviews on Consumer Demand for Experience Goods: A Case Study of Movie Critics. Journal of Industrial Economics, 53, 27–51.

Richins, M. L. (1983) Negative Word-of-Mouth by Dissatisfied Consumers: A Pilot Study. Journal of Marketing, 47, 68–78.

Sawhney, M. S., & Eliashberg, J. (1996). A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures. Marketing Science, 15, 113–131.

Senecal, S., & Nantel, J. (2004). The Influence of Online Product Recommendations on Consumers' Online Choices. Journal of Retailing, 80, 159–169.

Swami, S., Eliashberg, J., & Weinberg, C. B. (1999). SilverScreener: A Modeling Approach to Movie Screens Management. Marketing Science, 18, 352–372.

Zufryden, F. S. (1996). Linking Advertising to Box Office Performance of New Film Releases-A Marketing Planning Model. Journal of Advertising Research, 36, 29–41.

APPENDIX A MODEL ESTIMATION RESULTS

MODEL A

	M_I		P_I		Q_I		δ_I	
	MEAN	**SD**	**MEAN**	**SD**	**MEAN**	**SD**	**MEAN**	**SD**
Intercept	−0.241	0.639	0.035	0.057	**0.273**	**0.087**	**−1.517**	**0.188**
PG	0.870	0.468			−0.007	0.030		
PG13	0.436	0.454			−0.025	0.029		
R	**−1.502**	**0.434**			**−0.115**	**0.033**		
NR	0.020	1.003			−0.064	0.033		
SciFi	0.009	1.024			−0.006	0.031		
Kids	−0.022	0.938			**−0.101**	**0.027**		
Drama	**1.202**	**0.510**			**−0.084**	**0.035**		
Comedy	−0.498	1.070			0.006	0.023		
Romance	−0.145	1.131			**−0.257**	**0.056**		
Action	0.213	0.996			−0.030	0.020		
STAR								
MKT								
SCR	**1.200**	**0.699**						
SLEEPER	−1.140	0.763			**0.538**	**0.044**	0.686	0.460
CRITIC								
TOT								
AVG					**0.061**	**0.013**	**0.189**	**0.051**
AENTR					−0.049	0.051		
GENTR					**0.254**	**0.061**		
BOX			**1.011**	**0.012**				
MAPE	0.086							
AIC	731.200							

MODEL B

	M_I		P_{II}		Q_I		δ_I	
	MEAN	**SD**	**MEAN**	**SD**	**MEAN**	**SD**	**MEAN**	**SD**
Intercept	0.066	0.246	**−0.001**	**0.000**	0.159	0.094	**−2.025**	**0.259**
PG	**0.278**	**0.070**			0.018	0.034		
PG13	−0.014	0.067			−0.007	0.034		
R	0.004	0.072			**−0.115**	**0.038**		
NR	0.001	0.073			−0.050	0.039		
SciFi	**−0.814**	**0.102**			0.011	0.034		
Kids	**0.410**	**0.073**			**−0.115**	**0.032**		
Drama	0.021	0.105			−0.066	0.041		
Comedy	**0.206**	**0.074**			0.012	0.024		
Romance	**0.726**	**0.168**			**−0.252**	**0.070**		
Action	0.059	0.056			−0.023	0.024		
STAR			**2.02E–04**	**7.81E–05**				
MKT			6.64E–06	6.19E–06				
SCR	**0.657**	**0.047**						
SLEEPER	**−0.824**	**0.210**			**0.476**	**0.043**	0.471	0.555
CRITIC			**2.69E–04**	**6.95E–05**				
TOT			**5.29E–06**	**1.09E–06**				
AVG					**0.064**	**0.015**	**0.304**	**0.070**
AENTR			−7.77E–05	2.10E–04	0.000	0.055		
GENTR			**2.62E–03**	**5.53E–04**	**0.345**	**0.069**		
BOX								
MAPE	0.145							
AIC	1194.280							

MODEL C

	M_I		P_{III}		Q_{II}		δ_I	
	MEAN	**SD**	**MEAN**	**SD**	**MEAN**	**SD**	**MEAN**	**SD**
Intercept	**0.308**	**0.146**	**−1.15E–04**	**4.25E–05**	**0.475**	**0.059**	**−1.452**	**0.313**
PG	0.019	0.093			0.022	0.042		
PG13	**0.628**	**0.085**			−0.037	0.041		
R	**0.536**	**0.094**			**−0.135**	**0.044**		
NR	**0.564**	**0.090**			−0.081	0.046		
SciFi	**1.790**	**0.095**			**−0.091**	**0.035**		
Kids	**0.987**	**0.080**			**−0.075**	**0.037**		
Drama	0.063	0.138			−0.060	0.048		
Comedy	**0.241**	**0.095**			0.029	0.029		
Romance	**1.213**	**0.236**			**−0.157**	**0.077**		
Action	**0.422**	**0.065**			−0.042	0.026		
STAR			**−3.64E–05**	**1.46E–05**				
MKT			**6.58E–06**	**2.17E–06**				
SCR	**1.106**	**0.070**						
SLEEPER	**−0.669**	**0.188**			**0.336**	**0.034**	0.739	0.764
CRITIC			**8.18E–05**	**1.73E–05**	**0.050**	**0.014**	0.083	0.096
TOT								
AVG								
AENTR								
GENTR								
BOX								
MAPE	0.200							
AIC	1626.680							

MODEL D

	M_{II}		P_{IV}		Q_I		δ_I	
	MEAN	**SD**	**MEAN**	**SD**	**MEAN**	**SD**	**MEAN**	**SD**
Intercept	**1.115**	**0.161**	**−0.005**	**0.001**	**0.257**	**0.113**	**−2.252**	**0.345**
PG	**0.926**	**0.112**			−0.001	0.041		
PG13	0.150	0.118			−0.011	0.042		
R	−0.087	0.121			**−0.110**	**0.044**		
NR	0.192	0.122			−0.054	0.046		
SciFi	**−1.260**	**0.179**			0.049	0.043		
Kids	0.189	0.091			**−0.105**	**0.036**		
Drama	**−0.327**	**0.122**			−0.069	0.048		
Comedy	0.189	0.079			−0.011	0.028		
Romance	0.088	0.180			**−0.266**	**0.084**		
Action	0.285	0.071			−0.003	0.029		
STAR			**0.001**	**0.000**				
MKT								
SCR								
SLEEPER	**−2.181**	**0.162**			**0.462**	**0.047**	0.401	0.635
CRITIC								
TOT			**1.32E–05**	**2.64E–06**				
AVG			**0.001**	**0.000**	**0.050**	**0.018**	**0.356**	**0.094**
AENTR			0.000	0.000	−0.028	0.067		
GENTR			**0.003**	**0.001**	**0.356**	**0.078**		
BOX								
MAPE	0.177							
AIC	1542.880							

APPENDIX B DETAILED CROSS VALIDATION RESULTS

		MEAN ABSOLUTE PREDICTION ERROR			
	MOVIE TITLE	**MODEL A**	**MODEL B**	**MODEL C**	**MODEL D**
1	abandon	0.19	0.61	0.56	0.09
2	about a boy	0.08	0.34	0.38	0.11
3	adventures of pluto nash	0.08	0.96	1.48	0.28
4	analyze that	0.05	0.05	0.17	0.16
5	bad company	0.05	0.06	0.34	0.09
6	ballistic: ecks vs. sever	0.03	0.32	0.26	0.41
7	banger sisters	0.04	0.06	0.08	0.31
8	barbershop	0.04	0.29	0.37	0.10
9	big trouble	0.22	0.58	1.28	0.55
10	blood work	0.16	0.22	0.17	0.28
11	blue crush	0.03	0.25	0.36	0.07
12	bourne identity	0.08	0.21	0.24	0.14
13	brown sugar	0.20	0.49	0.56	0.35
14	catch me if you can	0.08	0.11	0.09	0.18
15	changing lanes	0.12	0.19	0.13	0.18
16	chicago	0.39	0.81	0.35	0.46
17	city by the sea	0.04	0.10	0.34	0.36
18	clockstoppers	0.02	0.11	0.13	0.14
19	confessions of a dangerous mind	0.34	0.48	0.61	0.46
20	count of monte cristo	0.08	0.18	0.30	0.03
21	country bears	0.06	0.11	0.07	0.43
22	divine secrets of the ya-ya sisterhood	0.08	0.15	0.24	0.10
23	eight legged freaks	0.16	0.06	0.81	0.18
24	emperor's club	0.08	0.25	0.38	0.16
25	empire	0.04	0.51	0.55	0.37
26	enough	0.02	0.05	0.18	0.12
27	formula 51	0.15	0.70	0.25	2.86
28	four feathers	0.05	0.08	0.08	0.11
29	frailty	0.02	0.06	0.10	0.74
30	frida	0.28	0.20	0.23	0.21
31	full frontal	0.25	0.14	0.26	0.15
32	ghost ship	0.02	0.09	0.22	0.23
33	good girl	0.16	0.17	0.35	0.19
34	halloween: resurrection	0.02	0.38	0.51	0.16
35	high crimes	0.08	0.03	0.05	0.03
36	ice age	0.02	0.14	0.31	0.08
37	jackass: the movie	0.02	0.06	0.41	0.07
38	jonah: a veggietales movie	0.10	0.18	0.40	1.53
39	k-19: the widowmaker	0.08	0.08	0.52	0.11
40	kangaroo jack	0.23	0.29	0.42	0.17
41	life of david gale	0.03	0.12	0.31	0.78
42	like mike	0.02	0.15	0.25	0.56
43	master of disguise	0.07	0.25	0.32	0.16
44	men in black ii	0.01	0.09	0.03	0.25
45	minority report	0.06	0.03	0.07	0.06
46	moonlight mile	0.22	0.32	0.30	0.31
47	mr. deeds	0.01	0.14	0.19	0.19

48	murder by numbers	0.02	0.07	0.07	0.15
49	my big fat greek wedding	0.41	0.69	0.52	0.22
50	new guy	0.03	0.04	0.06	0.14
51	one hour photo	0.49	0.36	0.23	0.36
52	panic room	0.02	0.03	0.10	0.04
53	phone booth	0.08	0.22	0.20	0.27
54	pianist	0.56	0.54	0.32	0.62
55	red dragon	0.08	0.09	0.15	0.20
56	reign of fire	0.04	0.05	0.11	0.03
57	return to never land	0.12	0.07	0.14	0.34
58	ring	0.25	0.19	0.41	0.33
59	rules of attraction	0.03	0.47	0.10	1.10
60	scooby-doo	0.06	0.22	0.13	0.34
61	scorpion king	0.12	0.05	0.07	0.19
62	serving sara	0.03	0.04	0.06	0.08
63	signs	0.05	0.19	0.21	0.16
64	solaris	0.15	1.03	1.17	1.05
65	spider-man	0.06	0.11	0.11	0.43
66	spirit: stallion of the cimarron	0.06	0.08	0.14	0.31
67	spy kids 2: the island of lost dreams	0.14	0.07	0.06	0.21
68	star wars: episode ii - attack of the clones	0.04	1.60	0.80	2.79
69	stuart little 2	0.08	0.07	0.14	0.15
70	sum of all fears	0.08	0.16	0.12	0.21
71	sweet home alabama	0.04	0.10	0.16	0.17
72	swimfan	0.13	0.22	0.05	0.07
73	transporter	0.06	0.04	0.18	0.06
74	trapped	0.06	0.23	0.29	0.25
75	treasure planet	0.13	0.16	0.78	0.11
76	tuxedo	0.03	0.09	0.03	0.08
77	unfaithful	0.13	0.13	0.20	0.23
78	we were soldiers	0.05	0.04	0.08	0.03
79	white oleander	0.04	0.14	0.22	0.15
80	windtalkers	0.03	0.12	0.16	0.19