Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning

Chrysanthos Dellarocas
MIT Sloan School of Management
dell@mit.edu

Neveen Farag Awad
University of Michigan Business School
nfarag@umich.edu

Xiaoquan (Michael) Zhang
MIT Sloan School of Management
zxq@MIT.EDU

May 10, 2004

Preliminary and Incomplete

Abstract[¶]

Despite the widespread popularity of online opinion forums among consumers, the business value that such systems bring to organizations has, so far, remained an unanswered question. This paper addresses this question by studying the value of online movie ratings in forecasting motion picture revenues. First, we conduct a survey where a nationally representative sample of subjects who do not rate movies online is asked to rate a number of recent movies. Their ratings exhibit high correlation with online ratings for the same movies. We thus provide evidence for the claim that online ratings can be considered as a useful proxy for word-of-mouth about movies. Inspired by the Bass model of product diffusion, we then develop a motion picture revenue-forecasting model that incorporates the impact of both publicity and word of mouth on a movie's revenue trajectory. Using our model, we derive notably accurate predictions of a movie's total revenues from statistics of user reviews posted on Yahoo! Movies during the first week of a new movie's release. The results of our work provide encouraging evidence for the value of publicly available online forum information to firms for real-time forecasting and competitive analysis.

[¶] This is a preliminary draft of a work in progress. It is being distributed to seminar participants for comments and discussion. This material is based upon work partially supported by the National Science Foundation under CAREER Grant No. 9984147. We are grateful to Erik Brynjolfsson for valuable suggestions.

1. Introduction

One of the most intriguing social phenomena brought forth by advances in information and communication technologies is the vast amplification of the power of word-of-mouth. With the help of the Internet, wireless networking, and mobile telephony, today's citizens and consumers are forming a bewildering array of technology-mediated communities where they exchange opinions and experiences on companies, products, services, and even world events.

Word-of-mouth is arguably a phenomenon as old as society itself. Nevertheless, the advent of the Internet has added two important new dimensions to this timeless concept:

- *Unprecedented scalability and speed of diffusion*. Information technologies enable opinions of a single individual to instantly reach thousands, or even millions of consumers. This escalation in audience is changing the dynamics of many industries in which word of mouth has traditionally played an important role. For example, the entertainment industry has found that the rapid spread of word of mouth is shrinking the lifecycles of its products (movies) and causing it to rethink its pre- and post-launch marketing strategies (Muñoz, 2003). In fact, movies are seeing much more rapid change in revenues between the opening weekend and second weekend, suggesting that public opinion is spreading faster[1].

- *Persistence and Measurability*. In offline settings word-of-mouth disappears into thin air. In online settings traces of word-of-mouth can be found in many publicly available Internet forums, such as review sites, discussion groups, chat rooms, and web logs. This public data provides organizations with the ability to quickly and accurately measure word-of-mouth *as it happens* by mining information available on Internet forums.

Rapid measurement is the first prerequisite for the fast *reactions* that are needed in this new playing field. Nevertheless, the information value of online forums to organizations is currently

---

[1] Rick Sands, the chief operating officer at Miramax, summarized this trend by stating that "In the old days . . .you could buy your gross for the weekend and overcome bad word of mouth, because it took time to filter out into the general audience. Those days are over. Today, there is no fooling the public" (Muñoz, 2003).

not well understood. There is controversy related to the reliability of online reviews as well as to how well these reflect the opinions of the population of consumers. Anecdotal evidence suggests that some of this information may be biased and is sometimes provided anonymously by the companies themselves (White 1999; Harmon 2004). Finally, even though the impact of online reviews on consumer behavior has been the focus of recent research (Chevalier and Mayzlin 2003; Senecal and Nantel 2003), there is very little work on how such information can be used by *firms* to gain business advantage.

This paper addresses firm usage of publicly available online word of mouth data by studying the information value of online movie reviews in forecasting motion picture revenues. We focused on the motion picture industry because word of mouth plays an important role and because online movie reviews are readily available.

Our study provides affirmative answers to two important questions:

- *How well do online reviews represent the opinions of the population at large?* We conducted a survey where a nationally representative sample of subjects who do not rate movies online was asked to rate a number of recent movies. Their ratings exhibit very high correlation with online ratings for the same movies. This result provides evidence for the claim that online ratings can be considered as a useful proxy for word-of-mouth about products.
- *Are online reviews a useful tool for forecasting future revenues?* Inspired by the Bass model of product diffusion, we develop a notably accurate revenue-forecasting model that is based on statistics of online movie reviews posted during the first week of a new movie's release.

In contrast with some previous motion picture revenue forecasting literature (Eliashberg and Shugan 1997), the thesis of this paper is not that online movie reviews *influence* future revenues, but rather, that online movie reviews constitute a *measurable proxy* for word of mouth that can be exploited by studios for revenue forecasting and planning. To our knowledge, we are the first to provide positive evidence for this question.

The rest of the paper is organized as follows. Section 2 discusses related work. Section 3 discusses our data set. Section 4 introduces our forecasting model. Section 5 discusses the results of fitting the model to our data set. Finally, Section 6 concludes and discusses the broader methodological implications of this work for revenue forecasting and competitive analysis.

2. Related Work

Our work relates and contributes to three important streams of past research: forecasting models of motion picture revenues, diffusion models of new product adoption, and methodologies for measuring 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 and Ahn, 1998). Despite such difficulty, several researchers have attempted to develop predictive models forecasting movie revenue (see Litman, 1998, for a review of such models). Such models can be classified along two main methodological dimensions: (i) Quantitative/Econometric Models that focus on factors that predict or influence motion picture revenue (Litman, 1983; Litman and Kohl, 1989; Sochay, 1994; Litman and Ahn, 1998; Neelamegham and Chintagunta, 1999; Ravid, 1999; Elberse and Eliashberg 2002); and (ii) Behavioral Models that focus on factors involved in individual decision making towards selecting a movie to watch (Eliashberg and Sawney, 1994; Sawney and Eliashberg, 1996; Zufryden, 1996; De Silva 1998; Eliashberg et al. 2000).

Research shows that a quarter of a motion picture’s total revenue comes from the first two weeks (Litman, 1997). Therefore, first-week box-office receipts can be used to predict total box office receipts of a particular movie (Sawhey and Eliashberg, 1996). Eliashberg and Shugan (1997) have showed that professional movie reviews are good predictors of late and cumulative box office receipts, but are not significantly associated with early box office receipts.

Notably missing from most studies is a consideration of the impact of word-of-mouth. Our study thus extends previous quantitative work on drivers of motion picture revenues by incorporating measurable proxies of word-of-mouth in forecasting box-office revenues and by examining the relative predictive power of such variables compared to more established variables such as critics' reviews and marketing expenditures.

Diffusion models of new product adoption. The impact of word of mouth on product sales was first examined by Bass (1969). The Bass model is a classic model of new product diffusion that incorporates the impact of mass media and interpersonal communication. The model has been shown capable of predicting the growth pattern of a wide range of new products with minimal data. The Bass model has spawned a huge literature of theoretical and empirical work. Many extensions to the model have been proposed. For excellent literature surveys see Mahajan et al. (1990; 2000). We contribute to the diffusion literature by proposing a novel extension of the original Bass model that includes a time discounting factor for word of mouth. Time discounting captures the fact that a consumer's intensity of interpersonal communication about a product is highest immediately following the time of adoption and tends to die out over time.

Methodologies for measuring word-of-mouth. Traditional attempts to measure word of mouth are based on two principal techniques: inference and surveys. Bass (1969) used aggregated sales data to infer the coefficient of internal influence. Reingen et. al. (1984) infers that dense interpersonal communication occurs with women who live in the same residence. Surveys have been used more often, largely because individuals can specifically be asked about their communication habits (e.g. Bowman and Narayandas, 2001); the error then lies in the self-reporting of their behavior.

The advent of the Internet introduced a third technique for measuring word of mouth: directly through Usenet groups and feedback forums. Researchers can gather large amounts of data from online feedback forums. Previous research has used *volume* and *dispersion* when examining online word of mouth (Godes and Mayzlin, 2002). The theory behind measuring dispersion, or the spread of communication across communities, is that word of mouth spreads quickly within communities, but slowly across them (Granovetter, 1973). The theory behind volume is that the more consumers discuss a product, the higher the chance that other consumers will become aware of it. In this study we extend previous attempts to measure the impact of online word-of-mouth by experimenting with additional measures such as the *valence* (e.g. the average rating of a movie), *density* (e.g. the fraction of consumers who feel compelled to provide online feedback) and *time evolution* of online feedback.

3. Online Movie Rating Sites

Few types of products capture people's imagination as much as motion pictures. It is, therefore, not surprising that a variety of online communities have sprung up on the Web covering various aspects of movies and the motion picture industry at large. Some of these communities are among the most popular Web destinations overall. Figure 1 shows the top 10 video/movie-related sites in terms of user traffic for the week ending March 28, 2004.

Of particular interest to this paper are sites that solicit and publish user ratings of movies. This section provides a brief introduction to the two sites we have used in the context of this study.

1.	Netflix (www.netflix.com)
2.	Internet Movie Database (www.imdb.com)
3.	Yahoo! Movies (movies.yahoo.com)
4.	UnderGroundOnline Film/TV (www.ugo.com)
5.	Moviefone (www.moviefone.com)
6.	Rotten Tomatoes (www.rottentomatoes.com)
7.	Sony Pictures Movies (www.sonypictures.com)
8.	Movies.com (movies.go.com)
9.	Universal Studios (www.universalstudios.com)
10.	fandango.com (www.fandango.com)

Figure 1: Top Internet video/movie destinations for week ending March 28, 2004. (Source: Nielsen/NetRatings)

3.1 Yahoo! Movies

Yahoo!Movies is part of the Yahoo portal. It offers a wealth of information about current and older movies including synopses, cast and credits, trailers, "news and gossip", box office data, theaters and show times (including the ability to buy tickets from affiliated websites), professional critic reviews, and user reviews.

Critic reviews are taken from major daily newspapers and specialized websites (such as filmcritic.com). Yahoo aggregates between 10 and 20 reviews for most movies. The system publishes the text of each critic's review together with a letter grade that ranges from A+ ("Oscar-worthy") to F ("all-time worst"). If the original critic's review does not include a rating, Yahoo! Movies assigns a grade based on an assessment of the review. An "average grade" that summarizes all reviews is also published.

Yahoo! Movies changed the format of user reviews in the summer of 2003, shortly after the data collection for this study was completed. We will therefore briefly describe both the old format (used in this study) and the new format.

Yahoo! Movies user reviews before Summer 2003

"Old format" Yahoo! Movies user reviews consisted of an optional text review together with an integer numerical rating that ranged from 5 (best) to 1 (worst). The system also allowed users to enter a text review or comment without specifying a numerical rating. Only registered Yahoo users could enter reviews; all published reviews listed the Yahoo user id of the author. However, since user registration on Yahoo makes entry of personal information optional, Yahoo user reviews are essentially anonymous. Nevertheless, a significant percentage of users supply at least partial demographic information (their gender and, less frequently, their age group).

Yahoo! Movies calculated the average numerical rating (with one decimal point of accuracy) and total number of user reviews submitted for each movie and prominently displayed this information at the top of the page devoted to that movie. In addition it listed all submitted reviews in reverse chronological order (40 reviews per page) and allowed users to browse the detailed text of each review.

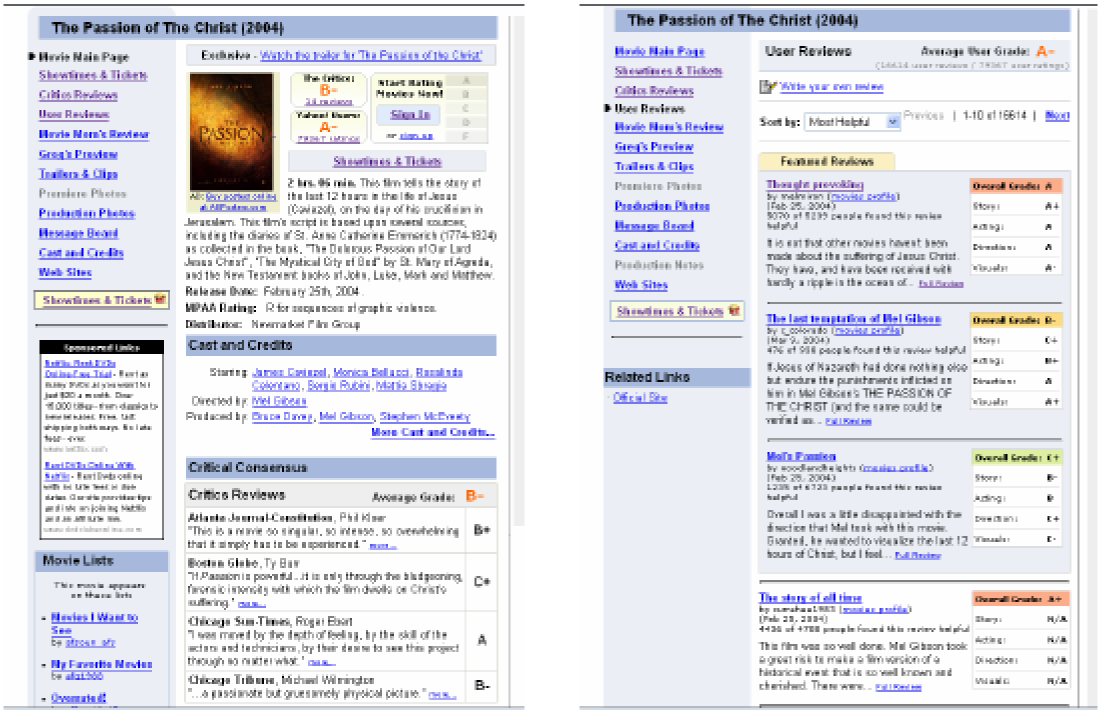


Figure 2: Example Yahoo! Movies movie information and user ratings pages.

Yahoo! Movies user reviews after Summer 2003

In the summer of 2003 Yahoo! Movies made some changes to the format of user reviews. The following are the major changes relative to "old format" user reviews:

- Instead of soliciting a single "aggregate" numerical rating per movie, Yahoo! Movies now allows users to enter separate ratings for Story, Acting, Direction, Visuals, as well as an Overall rating.
- The rating scale has changed to a 13-point letter grade ranging from A+ to F.
- Users can rate the usefulness of a review they have read by answering "Yes" or "No" to the question "Was this review helpful?" displayed at the bottom of the text review.
- By default, reviews are now sorted according to their usefulness rating (number of people who found the review useful).

As before, only registered Yahoo users can enter movie ratings. Yahoo still allows users to browse individual user reviews. Reviews can be sorted "by usefulness", chronologically, or by grade.

3.2 Internet Movie Database (IMDB)

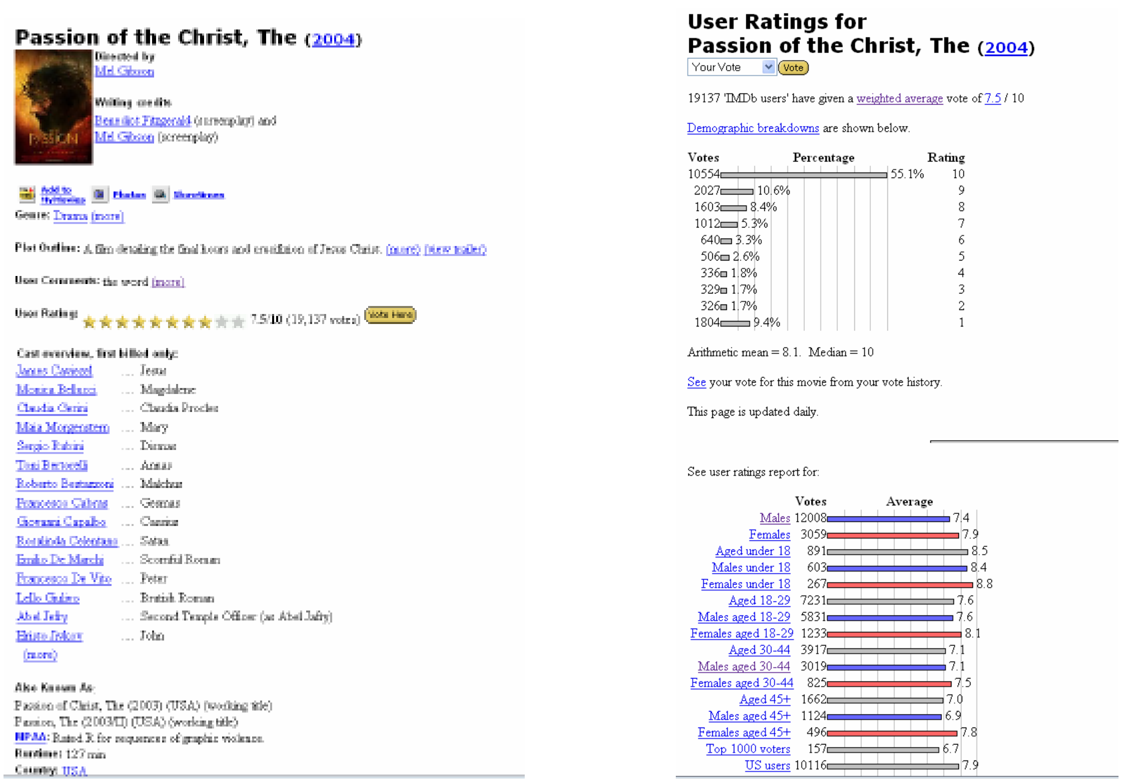


Figure 3: Example IMDB movie information and user ratings pages.

IMDB is a part of the Amazon group of companies. The site has a similar structure and offers a similarly rich set of information about current and older movies as Yahoo! Movies (synopses, cast and credits, trailers, news, box office data, theaters and show, links to critics reviews and newsgroup reviews, and user reviews). IMDB does not summarize critic reviews but provides links to the original text of those reviews.

IMDB solicits numerical user ratings on a scale from 1 ("awful") to 10 ("excellent"). IMDB stands out among movie sites in terms of the detailed statistical information it provides about user ratings. In addition to a weighted average rating[2] and total number of votes submitted for a movie, IMDB displays the histogram of votes as well as averages and histograms of votes submitted by various demographic groups within the population of voters (males, females, various age groups, US voters, non-US voters, top 1000 voters). Unlike Yahoo, IMDB does not make the history of user votes publicly available. It simply displays averages and histograms of the cumulative statistics of votes submitted so far.

4. Do online reviews represent the opinions of the population at large?

The reliability of online reviews in terms of reflecting the opinions of the population at large has been often put into question. A frequent argument is that the Internet user population is skewed relative to the national average (high proportion of higher-income males, low proportion of females and low-income groups). To shed some light on this question we conducted a survey where a nationally representative sample of subjects who do not rate movies online was asked to rate a number of recent movies. The responses were then compared to online ratings for the same movies collected from IMDB. The correlation between the two sets of ratings was remarkably high.

Methodology

An initial structured questionnaire was developed based on a review of the existing literature, as well as the structure of the online ratings scale on IMDB. The survey asked respondents to rate 25 recently released movies on the same ten-point scale used by IMDB; it also asked a series of questions regarding Internet usage, and frequency of use of online forums, and demographic

[2] The following explanation appears on IMDB's site: "IMDb publishes weighted vote averages rather than raw data averages. Various filters are applied to the raw data in order to eliminate and reduce attempts at 'vote stuffing' by individuals more interested in changing the current rating of a movie than giving their true opinion of it. The exact methods we use will not be disclosed. This should ensure that the policy remains effective. The result is a more accurate vote average."

information, including age, gender, and geographic location. The survey was executed during the summer of 2003. The survey was developed and pre-tested for content, flow, scope, and purpose on a group of undergraduate business students. The respondents were asked to comment on questions, to demarcate ambiguities, and to introduce any factors that may have been omitted. The questionnaire was refined based on initial feedback. The final questionnaire had fourteen questions related to various constructs, and twenty-five rating questions (the ratings responses can be found in Appendix A).

The survey was implemented in a web-based survey system, and emailed out to over 3000 randomly selected respondents from a nationally representative pool that MarketTools, Inc. owns. We received a total of 2007 respondents, 37 of which had significant missing data, leaving 1970 of useable responses. Partial summary information is shown in Tables 1A-1C.

Survey Results

The goal of the survey was to assess the validity of using online movie ratings as a proxy for word of mouth about movies in general. To that end, we tested the correlation between online ratings gathered through IMDB, and "offline" ratings, gathered through the survey. The correlation coefficient for Average Online Rating and Average Offline Rating was 0.8390, indicating that the online word of mouth is highly correlated with offline word of mouth, and is therefore a useful proxy variable.

We then divided the sample according to age, and gender, and again checked for correlations. In the sample of females only, the correlation was even higher, at 0.8592. In the case of Men, the correlation was slightly lower, but still significant, at 0.8258. The correlations across various consumer demographics are shown in Table 2.

Table 1A

	Yes	**No**
After Seeing Movies do you usually rate them online?	147 (7.0%)	1823 (93.0%)
Do you usually verbally recommend movies to other people?	1645 (84.0%)	324 (16.0 %)
Are you Male or Female	**Male:** 861 (44.0%)	**Female:** 1109 (56.0%)

Table 1B

	None	**1**	**2**	**3**	**4**	**5**	**6-10**	**>10**
How Many Times did you go to the movies in the past month?	729 (37%)	344 (17%)	303 (15%)	215 (11%)	152 (8%)	85 (4%)	107 (5%)	34 (2%)

Table 1C

Did the following factors affect which movie(s) you chose to watch?	**No**	**Somewhat**	**Yes**
Professional reviews (e.g. Roger Ebert)?	1316 (67%)	520 (26%)	133 (7%)
Television advertisement?	601 (31%)	893 (45%)	475 (24%)
Internet Advertisement?	1467 (75%)	411 (21%)	91 (5%)
Recommendation by Another Person?	574 (29%)	613 (31%)	782 (40%)
Online Ratings (e.g. Yahoo! Movies, Internet Movie Database, etc)?	1407 (71%)	438 (22%)	124 (6%)

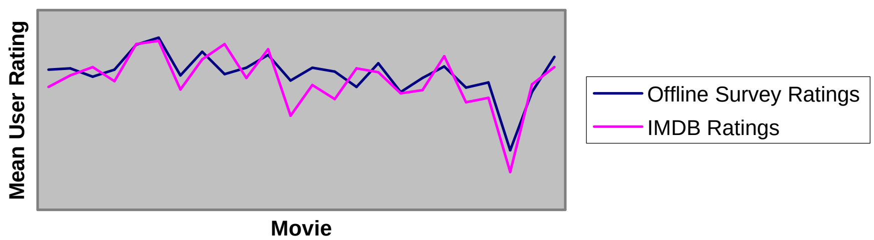
Table 1: Partial summary of survey results.

	Correlation Coefficient
Men	0.8258
Women	0.8592
18-29	0.7722
30-44	0.8680
45+	0.8475

Table 2: Correlation between Online and Offline Word of Mouth

Examining the online and offline data graphically illustrates that the two samples follow a similar patter. Figure 4 charts average user ratings at IMDB versus average offline user ratings.

Figure 4: Offline versus Online Word of Mouth



Yahoo! Movies Data and IMDB Data

We compared the offline ratings with online ratings from the International Movie Database, IMDB, because IMDB has a larger user base than Yahoo! Movies and better user demographics. However, due to confidentiality issues with the International Movie Database, we are unable to use their data for our forecast model. Thus, in the forecasting model, we use ratings from Yahoo! Movies. The obvious question, therefore, is to what degree to Yahoo! Movies ratings data and IMDB ratings data correlate? Interestingly, when we examine the 25 movies used in our sample, the correlation between Yahoo! Movie's user ratings and IMDB's user rating is 0.8590. In addition, the correlation between offline ratings from our survey on online ratings from Yahoo! Movies is 0.8947, which is higher than the correlation with IMDB ratings (0.8390). This higher correlation of offline ratings and Yahoo! Movies ratings lend support to our use of Yahoo! Movie ratings in our forecasting model.

Online Ratings Data and Critics Scores

Since much work has been done on using critics scores to predict movie revenue, the natural question becomes are movie critics ratings correlated with online word of mouth. Table 3 shows the correlation scores between critic scores and word of mouth scores from three sources: Yahoo! Movies, IMDB, and offline word of mouth collected from the survey. The scores are relatively low, with only the correlation between Critics and IMDB surpassing the threshold of 0.70 (Nunally 1967, 1978). The low correlation between word of mouth scores emphasizes the importance of examining word of mouth as a predictive tool, as the information provided by word of mouth is substantially different from the information provided by expert movie critic reviews.

Table 3: Critics Scores versus Word of Mouth

	Correlation
Critics and Yahoo! Movies	0.5270
Critics and IMDB	0.7463
Critics and Offline Word of Mouth	0.5123

5. Using online reviews as a tool for forecasting motion picture revenues

This section demonstrates that online ratings posted during the first week of a new movie's release can form the basis for remarkably accurate forecasting of that movie's future box office revenues. Specifically, we introduce a motion picture revenue forecasting model inspired by the Bass equation of product diffusion and apply the model to a rich data set of 2002 movie production, box office and user ratings that we collected from online sites.

5.1 Data Set

Data for this study were collected from Yahoo! Movies (http://movies.yahoo.com), BoxOfficeMojo.com (http://www.boxofficemojo.com), and The-Numbers.com (http://www.the-numbers.com). From Yahoo! Movies, we collected the names of all the movies of year 2002; this gave us a total of 2942 titles. For the purpose of our analysis, we excluded all titles that fall in one or more of the following categories: (a) not released in the United States, (b) not a theatrical release (VHS, DVD, etc), (c) not a nation-wide release. We were left with 474 movies, the total number of nationally released movies in 2002. For each of these titles, we collected detailed rating information from Yahoo! Movies, including all the critic reviews (score and review text) and all user reviews (date and time of review, rating, review text). We were also able to get somewhat noisy demographic information (gender, age) about each of the individual reviewers from the information attached to the associated Yahoo IDs.

We used Boxofficemojo.com and The-numbers.com to obtain weekly box office, budget and marketing expenses data. This information was missing for several movies from the publicly accessible parts of those sites. We obtained a data set of 128 movies with complete production, weekly box office, and daily user review data. We further trimmed this data set to only include movies for which at least 10 user ratings were posted during the first week of release. Our final data set consists of 80 movies, 1188 weekly box office data, 1040 critic reviews (an average of 13 reviews per movie), and 55156 user reviews from 34893 individual users. Of particular interest to our model are user reviews posted during the first week of a movie's release. Our data set contains an average of 312 first week user reviews per movie (minimum 12, maximum 3802).

5.2 Forecasting Model

Our revenue-forecasting model assumes that the evolution of a movie's revenues follows a modified Bass equation:

$$\dot{R}(t) = (N - R(t))(p\mathbf{d}^{t} + q\int_{k=0}^{t} \dot{R}(t-k)\mathbf{d}^{k}\,dk)\,,\ 0 < \mathbf{d} < 1 \qquad (1)$$

In equation (1), *t* represents elapsed time since a movie's nationwide release, $R(t)$ is the *cumulative* revenue generated until time *t*, $\dot{R}(t)$ is the rate of revenue increase at time *t*, and *N* is a measure of total potential revenue, i.e. the product of the total potential audience times the price of a movie ticket. In the diffusion literature (Majahan et al. 1990) the parameter *p* is traditionally known as the coefficient of external influence. In this case, *p* relates to the intensity of a movie's marketing campaign. For this reason it will be referred to in the rest of the article as the *coefficient of publicity*. The parameter *q* is traditionally known as the coefficient of internal influence. In our setting, it captures the effect of word of mouth from past moviegoers on subsequent movie revenues, and will, therefore, be referred to as the *coefficient of word of mouth*.

The difference of equation (1) relative to the traditional Bass model is the addition of the time discounting factor d^t. The introduction of time discounting attempts to model the following industry-specific facts: (i) Most movie marketing campaigns occur before or during the first weeks of a movie's release, and, therefore, have a diminishing effect in later weeks. (ii) Word of mouth is "localized in time"; people who watch a movie on week *k* typically talk about it the most in the immediately following days. Therefore, the contribution of past adopters on current adoption must be discounted by the amount of time that separates the current date from the time of adoption.

Given a training set of movies for which we have available production, ratings, and weekly box office revenue data, our forecasting model can be calibrated in two steps:

1. Using nonlinear least squares estimation, equation (1) is fitted to the weekly revenue vector of each movie in our training set to derive a pair of coefficient estimators $\hat{p}_i, \hat{q}_i$ per movie.

2. Using linear regression, predictive models are developed that estimate parameters $\hat{p}_i, \hat{q}_i$ using linear combinations of movie ratings statistics and other production data.

To forecast future box office revenues of a new movie we reverse the process:

1. The linear regression models derived by Step 2 above are applied to movie production and ratings data to derive estimates of coefficients p and q for that movie.

2. The estimated coefficients p and q are substituted into equation (1). Forecasts of the movie's revenues in future weeks are then derived using numerical integration.

5.3 Results

To test the predictive power of our model we divided our data set into two randomly generated subsets of 40 movies each. In the rest of the paper we will refer to these subsets as Data Set 1 and Data Set 2. We calibrated the model using each subset and used it to predict the total revenues of movies in the other subset.

Calibration Step 1: Nonlinear parameter estimation

We wrote a C program that numerically integrates equation (1) and performs nonlinear least-square parameter estimation based on the NL2SOL algorithm (Dennis et al. 1981). We expressed revenues in millions and time in weeks, such that N=1000 corresponds to \$1 billion in revenues, and t=1 refers to the end of week 1. We picked $N = 1000$ and $d = 0.3$ to run the model, however it should be noted that neither N nor the discount factor $0 < d < 1$ were very critical in achieving a good fit of equation (1) to our data set. The overall fit of the two-parameter model (1) to the weekly revenue vectors of each movie was excellent with an average $R^2 = 0.976$, which says that once we know p and q of a particular movie, we can predict its future weekly box office performance trajectory with an error less than 3%

The fact that the modified Bass equation (1) fits our movie revenue data so well is not surprising. It is well documented that the Bass equation works well with a wide range of phenomena (Bass et. al. 1994). What is remarkable, however, is that the estimated coefficients $\hat{p}_i, \hat{q}_i$ can be predicted with notable accuracy through simple linear regression models that use only first week online ratings and other publicly available movie information[3].

Calibration Step 2: Linear regression models to estimate p and q

We analyzed the relationship of coefficients $\hat{p}_i, \hat{q}_i$ to a large number of statistics derived from first week online ratings, box office, and production data (Table 4) and developed linear regression models for estimating each coefficient. To highlight the predictive power of online ratings relative to more established variables used in past research (such as first week box office revenues, critics reviews, budget and marketing expenditures) we developed two separate predictive models for each of the parameters $\hat{p}_i, \hat{q}_i$:

- Model A only uses independent variables that correspond to online ratings statistics
- Model B was built by considering independent variables corresponding to both online ratings and to first week box office and production data.

In selecting our models, our objective was to construct models with small numbers of independent variables and high levels of predictive power, thus balancing parsimony and precision. We followed a variable selection procedure similar to the traditional stepwise selection method: in each step, we included a significant variable (at the 5% level) that brought the highest increase in adjusted R^2 and checked if inclusion of that variable caused a blow-up of the variance inflation factor (VIF), which is a commonly used measure of multicollinearity. We stopped adding variables when the adjusted R^2 did not increase, when additional variables were no longer significant, or when adding new variables resulted in VIF higher than 8 for any of the variables. The resulting models are summarized in Table 5.

---

[3] We experimented with several variations of equation (1), including a simple Bass model without time discounting. The use of time discounting was not essential in fitting the Bass equation to the weekly revenue vectors of our data set. However, time discounting was crucial in order to obtain coefficient estimates $\hat{p}_i, \hat{q}_i$ that had meaningful relationships with (and therefore could be estimated by) first-week online ratings and box office data.

Yahoo! Movies Rating Statistics

Average User Rating (*week1*)	Arithmetic average of all nonzero Yahoo user ratings posted during movie's first week.
Adjusted Average User Rating (*w1adj*)	Weighted average of nonzero user ratings posted by males, females, and users with no declared gender (see text)
Entropy Coefficient (*ent1*)	Entropy of all nonzero Yahoo user ratings posted during a movie's first week.
Density of User Ratings (*dens1*)	$\frac{\text{Total First Week User Ratings}}{\text{First Week Box Office Revenues}}$ A measure of the fraction of moviegoers who posted Yahoo ratings for a given movie during the first week
Ratings Evolution Coefficient (*evol1*)	$\left(\frac{\text{Total First Week Ratings}}{\text{Total First Weekend Ratings}}\right)^2\left(\frac{\text{First Week Average Rating}}{\text{First Weekend Average Rating}}\right)^{10}$ A measure of how the initial "excitement" that surrounds the release of a new movie evolves during the rest of the first week (see text).
Total User Ratings (*tot1*)	Total number of Yahoo user ratings posted during movie's first week
Average Critics Rating (*critics*)	Arithmetic average of Yahoo's "letter grade" assessment of critics reviews.

Box Office and Production Data

Box Office Revenues (*box1*)	Total first week box office revenues for movie
Inverse Revenues per Theater Coefficient (*invpth1*)	$\left(\frac{\text{First Week Box Office Revenues}}{\text{Number of theaters where movie is shown}}\right)^{-\frac{1}{2}}$ Inverse of a measure of a movie's first week box office success relative to the producers' initial expectations.
Budget (*bdgt*)	Movie production budget
Marketing Costs (*mktg*)	Estimated movie marketing costs

Note: The functional forms of variables *evol1* and *invpth1* were determined by trial and error to maximize predictive power.

Table 4: List of independent variables considered in model selection.

Dependent variable: *p* (coefficient of publicity)

Independent Variables	Standardized parameter estimates					
	Data Set 1		Data Set 2		Complete Data	
	Model A	Model B	Model A	Model B	Model A	Model B
Total User Ratings (*tot1*)	0.92675 (16.10)***		0.80886 (8.91)***		0.84328 (14.44)***	
Density of User Ratings (*dens1*)	-0.29483 (-5.12)***		-0.34891 (-3.84)***		-0.32591 (-5.58)***	
Box Office Revenues (*box1*)		0.99349 (53.04)***		0.97010 (53.85)***		0.96764 (54.57)***
Marketing Costs (*mktg*)				0.03903 (2.17)**		0.03886 (2.19)**
Adj-R^2	0.8758	0.9867	0.6772	0.9928	0.7357	0.9851

t statistic in parentheses. ** 5% significance. *** 1% significance

Dependent variable: *q* (coefficient of word-of-mouth)

Independent Variables	Standardized parameter estimates					
	Data Set 1		Data Set 2		Complete Data	
	Model A	Model B	Model A	Model B	Model A	Model B
Adjusted Average User Rating (*w1adj*)	0.47601 (3.88)***	0.22151 (1.96)**	0.70433 (6.20)***	0.37570 (3.35)***	0.60896 (7.20)***	0.27998 (3.46)***
Ratings Evolution Coefficient (*evol1*)	0.36336 (3.07)***	0.20221 (2.13)**			0.21776 (2.58)***	0.13414 (2.00)**
Total User Ratings (*tot1*)	0.25174 (2.14)**					
Inverse Revenues per Theater (*invpth1*)		-0.60070 (-5.21)***		-0.55092 (-4.91)***		-0.57654 (-7.07)***
Adj-R^2	0.5350	0.7040	0.4832	0.6754	0.4774	0.6807

t statistic in parentheses. ** 5% significance. *** 1% significance

Table 5: Parsimonious linear regression models for estimating coefficients p and q

Coefficient of publicity

The coefficient of publicity *p* can be predicted with good accuracy using pure online ratings data. The use of box office and marketing data allows the construction of even more accurate estimators.

Model A. Using online ratings data alone, our two-variable model was able to explain 88% and 68% of the variation in *p* in Data Sets 1 and 2, respectively. The two independent variables are: 1) *first week's number of ratings* (*tot1*), with a positive coefficient; and 2) *first week's density of ratings* (*dens1*), with a negative coefficient. The number of ratings of a given movie is highly correlated to its first week box office revenues. These, in turn, are highly correlated to the publicity surrounding a movie's release. The observation of a strong positive relationship between *tot1* and *p* is therefore not surprising. On the other hand, the negative relationship between *dens1* and *p* is more perplexing. The density of ratings is a proxy for the fraction of moviegoers who feel compelled to post online ratings. Our data seems to suggest that this fraction is inversely proportional to the publicity surrounding a movie: proportionally fewer people post ratings for movies that are surrounded by high publicity.

A possible explanation for this phenomenon can be based on a "crowding-out" argument from the theory of public goods (Bernheim 1986). Movie ratings can be considered as a public good, in that their posting costs effort to the user but benefits the entire community. In several public good settings (e.g. charities), individuals have been observed to contribute less when there are substantial third-party sources of contributions (e.g. from the government). A variation of this argument can be used to hypothesize that users have a lower propensity to post ratings and reviews for popular movies for which they know that a lot of information exists from alternative sources. Another possible explanation is that the population of online movie raters is a fixed subset of the population, which is still rather small. Therefore, when the number of raters is normalized by the total first week box office revenues, this density measurement decreases as box office revenue increases. It then follows that if box office revenue and publicity are positively correlated, density and publicity are negatively correlated. More research is needed to ascertain the exact reason behind the relationship between *dens1* and *p*; nevertheless this finding provides an interesting glimpse into the complexity of the social dynamics of online communities.

Model B. When box office and production data is included in the list of candidate variables, none of the previous two variables remain significant. In particular, a movie's first week's total revenue (*box1*) is overwhelmingly influential to *p*; this variable together with marketing costs (*mktg*) explains 98.6% of the variation in *p* in Data Set 1 (99.3% in Data Set 2).

Given the set of phenomena (scope of marketing campaign, breadth of initial release etc.) that the coefficient of publicity is intended to capture, it is not surprising that box office and marketing data (that are direct measures of these phenomena) can lead to more accurate coefficient estimates than online ratings. Nevertheless, the respectable model fit (Adj-R^2 between 68% and 88%) that was achieved exclusively through the use of online ratings data provides evidence for the value of online ratings as a proxy of sales and marketing efforts. This could prove useful in competitive analyses of industries where sales and marketing data are not publicly available.

Coefficient of Word of Mouth

Reconfirming our expectations, the coefficient of word-of-mouth *q* is well explained by first week online ratings statistics. Adding box office measures only marginally improved the model.

Model A. Using online-ratings alone, our three-variable model was able to explain 53% and 48% of the variation in *q* in Data Sets 1 and 2 respectively. The following is a discussion of the three independent variables, in descending order of significance.

1. The first week's *adjusted average user rating* (*w1adj*) is a weighted average of ratings submitted by males, females and users of undeclared gender. The weights were determined experimentally to maximize model fit:

$$w1adj = 0.53 \times (FemaleRatings) + 0.05 \times (MaleRatings) + 0.42 \times (NoGenderRatings)$$

The adjusted average *w1adj* was more informative than the raw arithmetic average of ratings (*week1*). This is not surprising, given that the demographic breakdown of Yahoo! Movie users (75% male, Under 35) in our data set is skewed compared to the national average. An interesting observation is that ratings submitted by females carried substantially more weight than ratings submitted by males in terms of predicting the coefficient of word of mouth. This finding persisted in ratings submitted in later weeks and merits further attention.

2. By analyzing the evolution of Yahoo user ratings over time we discovered that they exhibit a systematic upward bias during the first weekend of a new movie's release (Figure 5). We hypothesize that this bias is due to the fact that a significant fraction of people who choose to watch a new movie during its first weekend are "self-selected" (i.e. have a special interest for the movie's genre, are devoted fans of the movie's stars, etc.) and thus have a higher propensity to like it than the average moviegoer. The *evolution coefficient* (*evol1*) is a measure of how fast this initial "excitement" about a movie (proxied by the product of the number of ratings times the average valence of ratings) declines during the remainder of the first week. The higher the value of *evol1*, the lower the decline of "excitement" about the movie relative to the first weekend. This measure turned out to be the second most significant predictor of the coefficient of word of mouth in our set of variables. The significance of the evolution coefficient in predicting a movie's coefficient of word of mouth demonstrates that valuable information can be extracted from studying the dynamics of online feedback communities.

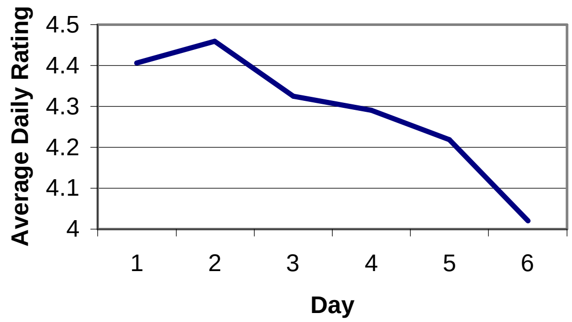


Figure 5: Average daily ratings posted on Yahoo! Movies during the first weekend of a new movie's release tend to be higher than average daily ratings posted on later days. This figure depicts the evolution of average daily ratings posted for "Spider-Man" during the first 6 days of its nationwide release. Most movies in our data set exhibit similar patterns.

3. The *total number of user ratings* (*tot1*) submitted during the first week turned out to be significant for predicting the coefficient of word of mouth, although less so than for predicting the coefficient of publicity. We believe that the presence of this variable in both models indicates that the number of first week moviegoers relates to both a movie's publicity campaign as well as to the strength of initial word of mouth about the movie.

Model B. The addition of box office and production data to the set of candidate variables increased the ability of our model to explain the variation of q in Data Sets 1 and 2 to 70% and 67.5% respectively. A movie's adjusted average rating (*w1adj*) and evolution coefficient (*evol1*) remained significant in this model as well. Among all possible measures of box office, there was only one variable significant in explaining q's variation: (*invpth1*) defined as the inverse of the square of *first week's revenue per theater*. We consider the first week's revenue per theater to be a measure of the financial success of a movie relative to the studio's pre-release expectations (these expectations determine the number of theaters on which the movie is initially released). Studios use sophisticated pre-release models that consider a variety of factors in making this decision (Litman, 1983; Litman and Kohli, 1989; Sochay, 1994; Zufreyden, 1996; De Silva, 1998; Eliashberg et al. 2000). Word of mouth is probably the most important factor that studios *cannot* accurately predict beforehand. Our hypothesis, therefore, is that the variation in revenue per theater is correlated with the strength of word of mouth about the movie. This hypothesis is confirmed by our data that shows a strong negative relationship between the *inverse* of this measure and the coefficient of word of mouth.

We were somewhat surprised to find that the predictive power of the *adjusted average user rating* (*w1adj*) was greater than that of *average critic reviews* (*critics*) on both data sets. Furthermore, the *critics* variable did not pass our variable selection criteria on any of our models (e.g. did not substantially increase the adjusted R^2 of any model and/or exhibited substantial multicollinearity with some other model variable). The tentative conclusion is that, when large numbers (~312 per movie) of online ratings are properly weighted and assessed, they can provide more information than a small number (~13 per movie) of expert reviews. Although specific to our context, this finding supports the viewpoint that online forums are emerging as a valid alternative source of information, replacing our societies' traditional reliance on the "wisdom of the specialist" by the "knowledge of the many".

A number of other variables that were considered but did not pass our variable selection criteria merit a brief mention: 1) the *entropy* (degree of dispersion or disagreement) of user ratings exhibited negative correlation with the coefficient of word of mouth (more disagreement correlated with lower intensity of word of mouth) but the effect was small compared to that of other variables we considered, 2) a movie's *budget* and *seasonality* (a categorical variable that specified whether the movie was released on a holiday weekend) were statistically insignificant for either coefficient.

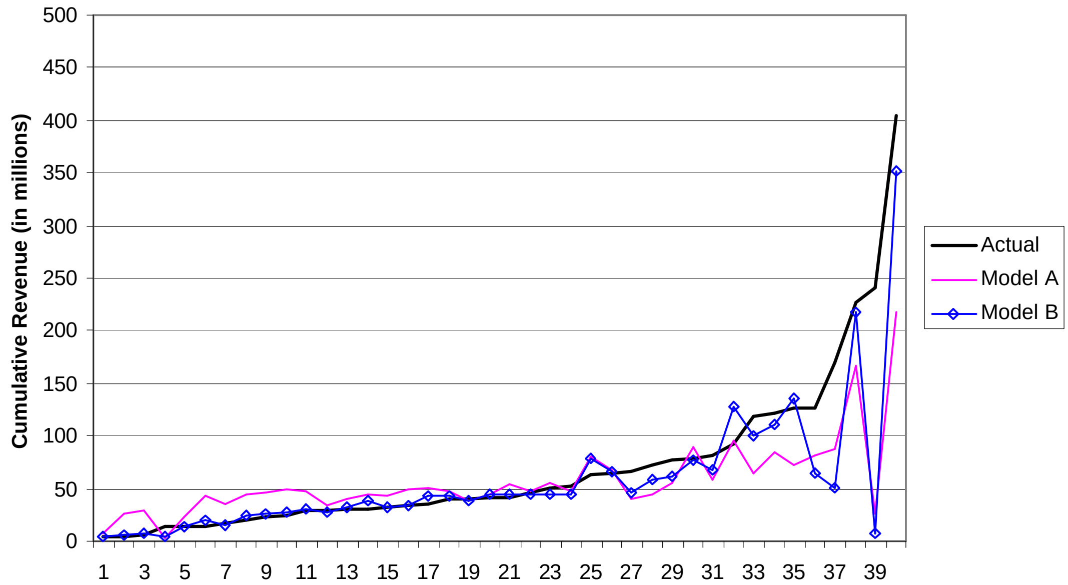
Forecasting Accuracy

The ultimate objective of our model is to help studios forecast a movie's future box office revenues from first week's rating and box office data. To test the predictive accuracy of our approach, we calibrated two pairs of models (each pair consisting of Model A and Model B), one using Data Set 1 and another using Data Set 2. We then used each model to derive estimates of coefficients p and q for movies in the *other* data set. Finally, we substituted those coefficients into equation (1) and performed numerical integration to derive forecasts of each movie's future week revenues.

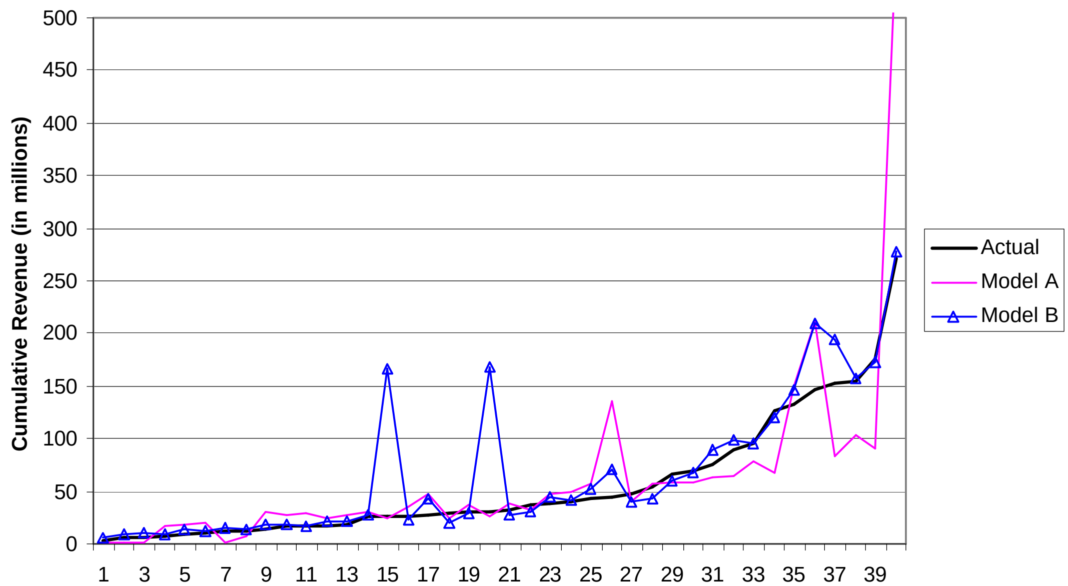
Figures 6A and 6B plot the predicted vs. actual *last-week* cumulative box office revenues for movies in each data set. Individual movies in each data set have been sorted in terms of increasing actual revenues. The X-axis of the plots, therefore, corresponds to the index of each of the 40 movies in each data set. From observing Figure 2 we can see that, with the exception of a small number of movies for which our models "get it very wrong" the models' predictions are remarkably accurate.

Figure 7 plots the relative absolute errors ($|\text{Predicted - Actual}|/\text{Actual}$) sorted in ascending order for each model and each data set. We see that the distribution of each model's errors is consistent across the two data sets. From Figure 3 we see that Model A achieves absolute prediction errors lower than 25% approximately 45% of the time and errors lower than 50% approximately 70% of the time. The more accurate Model B achieves absolute prediction errors lower than 25% approximately 70% of the time and errors lower than 50% approximately 90% of the time.

Movies with unusually high prediction errors correspond to "sleeper" movies such as "My Big Fat Greek Wedding" and "Chicago". Such movies are characterized by relatively low marketing campaigns and a slower box office revenue buildup, primarily fueled by word of mouth. Given that the parameters of our model were calibrated using a set of movies that contained a majority of blockbuster movies, it is not surprising that the model fails to predict accurately the revenue trajectory of sleeper movies. Nevertheless, in theory at least, the diffusion equation should be valid for sleeper movies as well. In future work we will calibrate the model using data sets that contain larger numbers of sleeper movies to test this point.



(A) Data Set 1



(B) Data Set 2

Figure 6: Actual vs. predicted last week cumulative revenues. The X axis corresponds to the index of each movie within its respective data set. Movies in each data set have been ordered in terms of increasing actual revenues.

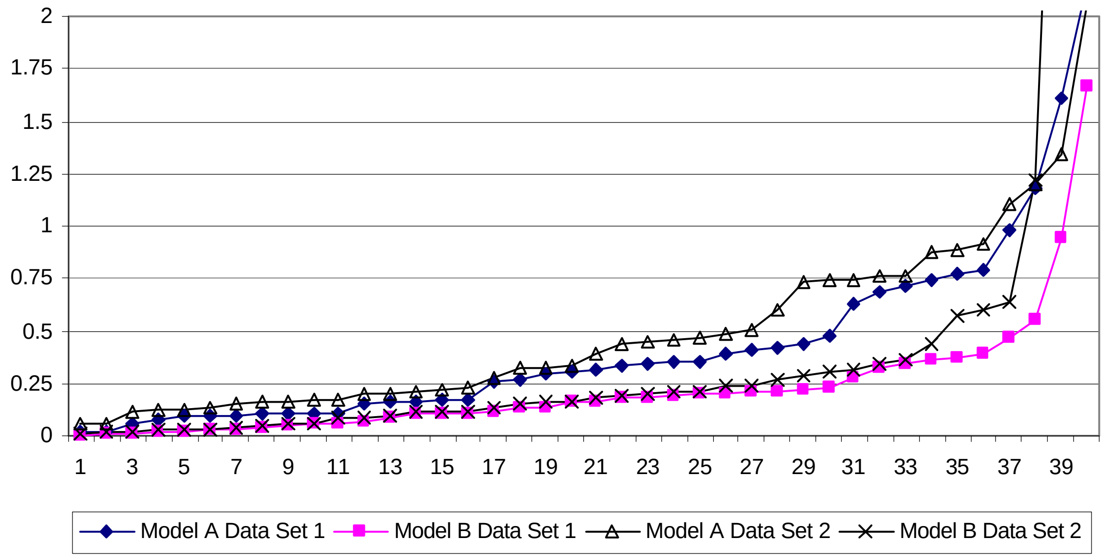


Figure 7: Relative absolute errors (RAE) for each model and each data set. Movies have been ordered in ascending order of RAE.

6. Conclusion

Online review sites are widespread on the Internet and rapidly gaining in popularity among consumers. Nevertheless, the business value of such information systems to organizations has, to date, not been established. This paper contributes in this direction by studying the value of online movie reviews in forecasting motion picture revenues. Our study provides affirmative answers to two important questions. First, we conduct a survey where a nationally representative sample of subjects who do not rate movies online is asked to rate a number of recent movies. Their ratings exhibit high correlation with online ratings for the same movies. We thus provide evidence for the claim that online ratings can be considered as a useful proxy for word-of-mouth about products. Second, inspired by the Bass model of product diffusion, we develop a simple and notably accurate motion picture revenue-forecasting model, based on statistics of online movie reviews posted by users on Yahoo! Movies during the first week of a new movie's release.

In addition to providing positive evidence for the value of online ratings in forecasting and planning, our study has produced several interesting insights related to online communities. First, we found that weighted averages of user ratings were more informative in predicting future revenues than averages of professional critic reviews. Even though we cannot claim generality, this finding supports the viewpoint that online forums are emerging as a valid alternative source of information to mainstream media, replacing our societies' traditional reliance on the "wisdom of the specialist" by the "knowledge of the many". Second, in addition to serving as a proxy of word-of-mouth, we have shown that statistics of online ratings can also be used to estimate aspects of a company's sales and marketing efforts. Finally, we found that the density and time evolution of online feedback contain valuable information that merits further research.

Apart from helping a company forecast demand and plan its own actions, we believe that the techniques introduced in this paper have the potential to play an important role in competitive analysis. In a lot of product categories, sales and marketing budgets are secret information, and therefore competitive analysis has hitherto been a difficult task. The vast amounts of consumer ratings that are publicly available on the Internet have the potential to fundamentally change this.

Our future work will explore enhancements that further improve the predictive power of our model. Initial experiments indicate that consideration of competition from other simultaneously released movies and calibration of separate models for different movie genres can improve forecast accuracy. Our most important long run goal, however, is to apply the techniques explored in this paper to other industries as well, in order to better understand the impact of online word of mouth on product diffusion and the competitive implications of this phenomenon for firms and online community operators.

References

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

Bass, F., T.V. Krishnan, D.C. Jain (1994). Why the Bass model fits without decision variables. *Marketing Science* 13 (2): 203–223.

Bernheim, B. D. (1986) On the Voluntary and Involuntary Provision of Public Goods. American Economic Review, 76 (4): 789-793.

Chevalier, Judith A., and, Mayzlin, Dina (2003). The Effect of Word of Mouth on Sales: Online Book Reviews. *Yale SOM Working Paper No's. ES-28 & MK-15*.

Dennis, J. E., Gay, D. M. and Welsch, Roy E. (1981) Algorithm 573: NL2SOL—An Adaptive Nonlinear Least-Squares Algorithm. ACM Transactions on Mathematical Software (TOMS) 7 (3) (September 1981): 369 - 383

De Silva, I. (1998). Consumer Selection of Motion Pictures. *The Motion Picture Mega-Industry*. B. Litman. Boston, MA, Allyn & Bacon Publishing, Inc.

Elberse, A. and Eliashberg, J. (2002). *The Drivers of Motion Picture Performance: The Need to Consider Dynamics, Endogeneity and Simultaneity*. Business and Economic Scholars Workshop in Motion Picture Industry Studies, Florida Atlantic University.

Eliashberg, Jehoshua, and, Sawhney, M.S. (1994). Modeling Goes to Hollywood: Predicting Individual Differences in Movie Enjoyment. *Management Science* 40(9): 1151-1173.

Eliashberg, Jehoshua, and, Shugan, Steven M. (1997). Film critics: Influencers or predictors? *Journal of Marketing* 61(2): 68-78.

Eliashberg, Jehoshua, Jedid-Jah Jonker, Mohanbir S. Sawhney, and Berend Wierenga (2000), MOVIEMOD: An Implementable Decision Support System for Pre-Release Market Evaluation of Motion Pictures. *Marketing Science,* Vol. 19, No. 3, pp. 226-243.

Godes, David, and, Mayzlin, Dina (2002). Using Online Conversations to Study Word of Mouth Communication. *Yale SOM Working Paper No. MK-13; Harvard NOM Working Paper No. 02-32; HBS Marketing Research Paper No. 02-01*.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology* 78(6): 1360-1380.

Harmon, Amy (2004). Amazon Glitch Unmasks War of Reviewers. *The New York Times*. New York, February 14.

Litman, B.R. and Ahn, H, (1998). Predicting Financial Success of Motion Pictures. *The Motion Picture Mega-Industry*. B. R. Litman. Boston, MA, Allyn&Bacon Publishing, Inc.

Litman, B.R., and Kohl, A. (1989). Predicting financial sucess of motion pictures: The 80's experience. *The Journal of Media Economics* 2(1): 35-50.

Litman, Barry R. (1983). Predicting Success of Theatrical Movies: An Empirical Study. *Journal of Popular Culture* 16(Spring): 159-175.

Mahajan, Vijay, Muller, Eitan, and Kerin, Roger A. (1984). Introduction Strategy for New Products With Positive and Negative Word-of-Mouth. *Management Science* 30(December): 1389-1404.

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

Mahajan, V., Muller, E., and Wind, Y. (2000). *New product diffusion models*. Norwell, MA, Kluwer Academic Publishers.

Muñoz, Lorenza (2003). High-Tech Word of Mouth Maims Movies in a Flash. *Los Angeles Times*. Los Angeles, August 17.

Neelamegham, P., and Chintagunta, P. (1999). A Bayesian Model to Forecast New Product Performance in Domestic and International Markets. *Marketing Science* 18(2): 115-136.

Ravid, S.A. (1999). Information, Blockbusters, and Stars: A Study of the Film Industry. *Journal of Business* 72(4): 463-492.

Sawhey, M.S., and Eliashberg, J. (1996). A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures. *Marketing Science* 15(2): 113-131.

Senecal, S. and Nantel, J. (2003) The Influence of Online Product Recommendations on Consumers' Online Choices. Working Paper.

Sochay, S. (1994). Predicting the performance of motion pictures. *The Journal of Media Economics* 7(4): 1-20.

White, E. (1999). Chatting a Singer Up the Pop Charts. *The Wall Street Journal* October 5.

Zufryden, F.S. (1996). Linking advertising to box office performance of new film releases - A marketing planning model. *Journal of Advertising Research*. 36 (4): 29-41.

Appendix A: Survey Movie Ratings Responses

1. S.W.A.T	0% 8	0% 9	0% 8	1% 13	1% 22	1% 29	2% 39	2% 42	2% 37	2% 42	87% 1720
2. The Recruit	0% 5	0% 8	1% 16	1% 15	3% 51	3% 59	5% 105	5% 92	3% 56	3% 64	76% 1498
3. Freaky Friday	1% 13	1% 10	1% 14	1% 26	1% 26	2% 31	2% 45	3% 61	2% 45	2% 46	84% 1652
4. American Wedding	1% 10	1% 12	1% 11	1% 17	1% 23	2% 42	3% 52	3% 56	2% 40	3% 62	83% 1644
5. Pirates of the Caribbean: The Curse of the Black Pearl	0% 5	0% 7	1% 11	1% 13	1% 20	1% 28	3% 67	7% 135	6% 111	11% 207	69% 1365
6. Finding Nemo	0% 5	0% 6	0% 4	1% 10	1% 21	2% 30	3% 68	5% 99	7% 140	17% 328	64% 1258
7. Anger Management	1% 10	1% 15	1% 20	1% 19	2% 39	3% 57	4% 70	4% 83	2% 44	3% 53	79% 1559
8. Seabiscuit	0% 9	0% 8	0% 5	1% 13	1% 17	1% 25	2% 32	3% 55	4% 78	6% 109	82% 1618
9. Whale Rider	0% 9	0% 4	0% 4	0% 8	1% 15	1% 12	1% 11	1% 19	1% 24	1% 22	93% 1841
10. Bruce Almighty	1% 17	1% 15	1% 10	1% 20	2% 45	3% 63	5% 96	6% 111	4% 77	4% 79	73% 1436
11. X2: X-Men United	0% 8	0% 3	1% 16	1% 11	1% 28	3% 55	5% 101	7% 139	5% 98	6% 126	70% 1384
12. Spy Kids 3-D: Game Over	1% 11	1% 12	1% 14	1% 14	2% 30	2% 37	2% 43	2% 48	1% 27	2% 30	86% 1703
13. Bad Boys II	1% 15	1% 12	1% 12	1% 11	1% 28	1% 28	2% 46	3% 52	3% 50	3% 66	84% 1649
14. Lara Croft Tomb Raider: The Cradle of Life	1% 20	0% 8	1% 15	1% 13	1% 27	2% 44	3% 57	3% 52	2% 41	3% 56	83% 1636
15. Swimming Pool	1% 11	0% 6	0% 3	0% 5	0% 9	1% 12	1% 18	1% 18	1% 10	1% 10	95% 1867
16. Terminator 3: Rise of the Machines	1% 13	0% 6	1% 19	1% 16	2% 37	3% 58	4% 79	5% 93	4% 76	5% 94	75% 1478
17. The Hulk	2% 36	1% 16	2% 35	2% 30	3% 60	4% 71	3% 66	3% 52	2% 37	2% 40	78% 1526
18. Daddy Day Care	1% 16	0% 8	1% 15	1% 18	2% 34	2% 44	2% 43	3% 62	2% 31	2% 48	84% 1650
19. Bend it Like Beckham	1% 11	0% 6	0% 4	0% 7	1% 10	1% 16	1% 25	2% 32	2% 42	2% 34	91% 1782
20. Dreamcatcher	1%	0%	1%	0%	1%	2%	1%	2%	1%	1%	89%

	16	8	12	9	26	33	29	36	22	20	1758
21. 2 Fast 2 Furious	1% 21	1% 16	1% 16	1% 20	2% 30	1% 29	3% 56	2% 41	2% 34	2% 41	85% 1665
22. Gigli	5% 89	1% 21	1% 17	0% 7	0% 6	1% 15	0% 9	0% 2	0% 5	0% 6	91% 1792
23. Freddy Vs. Jason	1% 25	1% 17	0% 8	0% 5	1% 14	1% 18	1% 17	1% 19	1% 22	1% 28	91% 1796
24. The Matrix Reloaded	1% 12	1% 11	1% 19	1% 22	2% 43	3% 62	5% 104	5% 106	6% 115	9% 170	66% 1305