MARKETING SCIENCE
Vol. 24, No. 3, Summer 2005, pp. 508–517
ISSN 0732-2399 | EISSN 1526-548X | 05 | 2403 | 0508

DOI 10.1287/mksc.1040.0106


# Modeling Movie Life Cycles and Market Share

Andrew Ainslie
UCLA Anderson School of Management, 110 Westwood Plaza, Room B412, University of California, Los Angeles, Los Angeles, California 90095, andrew.ainslie@anderson.ucla.edu

Xavier Drèze
Wharton School, University of Pennsylvania, 3730 Walnut Street, Suite 700, Philadelphia, Pennsylvania 19104, xdreze@wharton.upenn.edu

Fred Zufryden
Marshall School of Business, Marketing Department, Accounting Building 301G, University of Southern California, Los Angeles, California 90089-0443, fred.zufryden@marshall.usc.edu

We examine box-office sales in the context of a market share model. This is accomplished by developing a combination of a sliding-window logit model and a gamma diffusion pattern in a hierarchical Bayes framework. We show that accounting for the full choice set available every week not only increases the fit of weekly movie sales but also leads to parameter estimates that depict a richer picture of the movie industry. We show that movie studios appear to have a good understanding of the products they produce, knowing when to support them and when not to. We also show that the effect of the number of opening week screens is overestimated in traditional models. Our research indicates that actors have a direct and directors an indirect effect on consumers' movie choice. Releasing a movie contemporaneously with other movies of the same genre adversely affects box-office performance all around. Releasing a movie against movies of the same Motion Picture Association of America (MPAA) rating hurts its sales in the beginning, but there is a displacement effect, which leads to a less severe sales loss in the long run.

*Key words*: entertainment marketing; motion picture distribution and exhibition; movie choice; new product research
*History*: This paper was received November 21, 2003, and was with the authors 4 months for 2 revisions; processed by Pradeep Chintagunta.

## 1. Introduction

The American public has long had a fascination with movies. With 449 new movies released, $9.5 billion in domestic box-office sales, and 1.6 billion tickets sold, 2003 was a record year (MPAA 2004). Movie production costs for the studios that are members of the MPAA[1] leveled around $60 million per movie in the last couple of years, but marketing costs are still rising. These costs add an average of $39 million per movie. Recent films such as *The Matrix Reloaded* and *Spider-Man II* have production costs of more than $100 million (www.IMDb.com). The expenditures on production represent a significant investment that the studios seek to recoup, first through domestic and international box-office release, and then through the sales of prerecorded tapes and DVDs (the Home Video market). Cable and broadcast TV can also be sizable sources of revenue. The U.S. box-office revenue typically establishes the "value" of the movie for the other markets.

Given the size of the bets made by each studio when they produce and release movies, their marketing budgets, and the frequency with which new movies are released (nine every week on average for 2003, MPAA 2004), one would assume that studios have become proficient at predicting movie success. Thus, it might be expected that studios would use tools similar to those of Gillette or 3M to predict sales before launching one of their products. However, surprisingly, Hollywood has not put much stock in sales prediction models, arguing that movies are artistic creations that cannot be modeled. The movie industry believes more in instinct and analysis by anecdote (Red Herring 1998).

Studios do, however, recognize the impact of competition. The release dates of many movies are pushed back, or brought forward, to avoid coming out simultaneously with competing movies that may be stronger players (Eonline 2002). This behavior has been shown to be optimal from a theoretical standpoint by Krider and Weinberg (1998) in the presence of a strong seasonal pattern. However, there has been little published empirical research incorporating the

[1] MPAA member companies account for almost half of the movies released in a given year.

individual effects of competing movies in the modeling of box-office receipts. This is likely because of the nature of the business (e.g., a rapid diffusion process, over a very short period, as hundreds of movies enter and exit the market every year), which makes such modeling efforts difficult.

Sawhney and Eliashberg (1996) developed the BOXMOD model, where they decompose the consumer's movie selection in two steps: (1) the consumer makes the decision to see a movie and (2) the consumer acts on this decision. By modeling the time to decide and the time to act as exponential decays, a three-parameter model is derived from a general gamma. They then perform a metaanalysis on the three parameters to study factors that drive movie sales (e.g., MPAA rating and movie genre). An important contribution of their work is the distinction between blockbusters and sleepers. The classification as a sleeper or blockbuster movie is made on the basis of the diffusion pattern and not on the actual box-office earned (i.e., some sleepers actually earn more than some blockbusters). Blockbuster-type movies have an exponential-decaying sales pattern, with the opening week grossing the largest sales. Sleeper-type movies build sales gradually and generally peak 3 to 6 weeks after launch.

Shugan (1998) looks at box-office performance based on the team that participated in the creation of the movie (i.e., writers, directors, actors). His goal is to help studios predict box-office sales early on during the production process—a time at which the finalized product is not available, but the track record of the production team is known. Based on the past box-office performance of the movies in which the production team was involved, Shugan (1998) predicts opening day box office with an $R^2$ of 0.59 and total box office with an $R^2$ of 0.34.

Some researchers have concentrated on postlaunch profitability. Swami et al. (1999) study the allocation of multiplex screens to movies so as to maximize distributor profits. Neelameghan and Chintagunta (1999) and Elberse and Eliashberg (2003) study the international diffusion of movies. An important contribution of the latter work is making the screen allocation (i.e., breadth of diffusion) endogenous.

There has also been some theoretical work on the competitive aspect of movie release. Krider and Weinberg (1998) investigate film release strategies based on the assumption of exogeneity of the highly seasonal nature of the movie business (sales peak during the summer months and during the holiday season) coupled with the shortness of movies' life cycles. In particular, they suggest that strong movies should compete head to head during peak weeks, while weak movies should delay their release if they are facing strong competition. Radas and Shugan (1998) also study the timing game. They propose an ingenious approach for handling seasonality through acceleration and deceleration of time.

Nevertheless, to date, with the exception of Swami et al. (1999), published studies still model box-office performance of movies in isolation of the other concurrent movie releases. To address this issue, and to further our understanding of the movie business, we propose to study box-office performance in a setting where moviegoers choose which movie to see among *all* the movies present at the box office at any point in time. Framing the problem in market share terms has important ramifications for the model specifications. We do not limit ourselves to only studying the large movies, but also consider smaller ones. Thus, the model must be flexible enough to fit both sleeper- and blockbuster-type movies. Indeed, using Sawhney and Eliashberg's (1996) taxonomy, 36% of the movies in our database are sleepers. We also incorporate the effects of seasonality and handle heterogeneity in the path followed by movies' market shares.

We develop our model in the next section of this paper. We then calibrate the model using a comprehensive movie database. This empirical analysis is followed by a discussion of findings and managerial implications.

## 2. A Flexible Bayesian Model for Predicting Movie Sales

We combine a random effects logit model with a gamma diffusion pattern, adapting each part to fit our goal. To account for the short life of movies, the logit model incorporates an indicator variable ($I_{it}$) for each movie that is set to one during weeks for which a particular movie appears in theaters, and 0 otherwise (i.e., before the movie is released or after it has been pulled out of the theaters). We describe this as a *sliding-window* logit model, in that it is essentially a logit model that allows for a different product set in each period. Finally, we include an outside good to account for those consumers who choose not to go to the movies at any given time. As we will show, an outside good allows us to incorporate seasonal effects directly into the market share model. We assume an extreme value error term on the attractiveness of each movie (i.e., $U_{it} = V_{it} + \varepsilon_{it}$ with errors distributed i.i.d.). This leads to the following logit formulation for the market share of a movie:

$$M_{it} = \frac{e^{V_{it}} I_{it}}{e^{V_{Ot}} + \sum_j e^{V_{jt}} I_{jt}}. \tag{1}$$

For the outside good it is

$$M_{ot} = \frac{e^{V_{Ot}}}{e^{V_{Ot}} + \sum_j e^{V_{jt}} I_{jt}}, \tag{2}$$

where
$M_{it}$ = expected market share of movie $i$ in week $t$,
$M_{ot}$ = expected market share of the outside good in week $t$,
$V_{it}$ = deterministic component of the market attractiveness of movie $i$ in week $t$,
$I_{it}$ = indicates whether movie $i$ is screened in week $t$ (1 if it is, and 0 otherwise),
$V_{Ot}$ = deterministic component of the market attractiveness of the outside good in week $t$.

To model the deterministic component of market attractiveness of each movie ($V_{it}$), we take inspiration from BOXMOD but strive for clear interpretability of the parameters. BOXMOD is a variant of the general gamma model (Sawhney and Eliashberg 1996, McGill and Gibbon 1965). In BOXMOD, the parameters ($\lambda$, the time to decide and $\gamma$, the time to act), although appealing from a consumer behavior standpoint, are of limited practical use to managers and difficult to interpret when the dependent variable is market attractiveness rather than sales. To remedy this while retaining the proven ability of gamma functions to describe the diffusion patterns of movies, we begin with a gamma distribution. When used in diffusion settings, the gamma is generally parameterized as

$$N_i \frac{1}{\beta_i^{\alpha_i}\Gamma(\alpha_i)} t^{\alpha_i - 1} e^{-t/\beta_i}$$

(where $N_i$ is total demand for the movie).

We reparameterize this as follows:

$$\eta_i t^{\gamma_i/\beta_i} e^{(1-t)/\beta_i},$$

where

$$\eta_i = N_i \frac{1}{\beta_i^{\alpha_i}\Gamma(\alpha_i) e^{1/\beta_i}} \quad \text{and}$$
$$\gamma_i = (\alpha_i - 1)\beta_i.$$

We then incorporate this as the deterministic component of the movie attractiveness formulation

$$U_{it} = V_{it} + \varepsilon_{it}, \qquad V_{it} = \ln\left(\eta_i w_{it}^{\gamma_i/\beta_i} e^{(1-w_{it})/\beta_i}\right), \tag{3}$$

where $w_{it}$ represents the number of weeks in release of movie $i$ in time period $t$ (i.e., $w_{it} = 1$ during the first week of release). In this representation, $\eta_i$ is the expected attractiveness of the movie in its opening week, $\gamma_i$ indicates when peak attractiveness occurs (i.e., the modal value of this distribution), and $\beta_i$ is a speed parameter representing how fast attractiveness builds and decays.

To model the demand for the outside good, we take advantage of the highly seasonal pattern of total weekly box-office sales. We define the weekly demand for the outside good as:[2]

$$S_{Ot} = \bar{S} - \sum_{i=1}^{I} S_{it} I_{it},$$

where
$S_{it}$ = box-office sales for movie $i$ in week $t$,
$\bar{S}$ = an arbitrary large maximum potential demand for movies.

We then set an autoregressive model (with $K$ 52-week lags) to estimate the attractiveness of the outside good in a way that accounts for potential seasonality

$$V_{Ot} = \alpha + \sum_{k=1}^{K} \phi_k \ln\left(\frac{S_{Ot-(k*52)}}{\bar{S}}\right). \tag{4}$$

Our final model is then

$$M_{it} = \frac{\eta_{i,\text{ms}} w_{it}^{\gamma_{i,\text{ms}}/\beta_{i,\text{ms}}} e^{(1-w_{it})/\beta_{i,\text{ms}}} I_{it}}{e^{V_{Ot}} + \sum_j \eta_{j,\text{ms}} w_{jt}^{\gamma_{j,\text{ms}}/\beta_{j,\text{ms}}} e^{(1-w_{jt})/\beta_{j,\text{ms}}} I_{jt}}, \tag{5}$$

where we use the subscript ms to refer to the market share model.

To link the three movie-level parameters ($\eta_{i,\text{ms}}$, $\gamma_{i,\text{ms}}$, and $\beta_{i,\text{ms}}$) to movie characteristics, we set up a hierarchical regression on the three parameters using a mix of continuous ($X_i$) and categorical variables (Genre and Studio, see §3 for a list of the covariates). To allow for the fact that some of the categorical variables have a large number of levels (e.g., there are 22 different studios), we use the massively categorical methodology described in Steenburgh et al. (2003) to handle the hierarchical regression. In our estimation, we normalize $\eta_{i,\text{ms}}$ so that it represents the expected market share of movie $i$ in its opening week (see the technical appendix at http://mktsci.pubs.informs.org). Thus, we constrain $\eta_{i,\text{ms}}$ to lie between 0 and 1, and $\beta_{i,\text{ms}}$ to be positive in the regression structure by applying appropriate transformations to them. Hence, our hierarchical regression has the following structure:

$$\begin{gathered}
\eta^*_{i,\text{ms}} = \ln\left(\eta_{i,\text{ms}}/(1-\eta_{i,\text{ms}})\right), \\
\gamma^*_{i,\text{ms}} = \gamma_{i,\text{ms}}, \\
\beta^*_{i,\text{ms}} = \ln(\beta_{i,\text{ms}}), \\
\begin{bmatrix} \eta^*_{i,\text{ms}} \\ \gamma^*_{i,\text{ms}} \\ \beta^*_{i,\text{ms}} \end{bmatrix} = X_i \Delta_{\text{ms}} + \theta^{\text{Studio}}_{i,\text{ms}} + \theta^{\text{Genre}}_{i,\text{ms}} + \Sigma_{\Delta_{\text{ms}}}.
\end{gathered} \tag{6}$$

In this formulation, the three parameters have a direct interpretation: Recall that $\eta_{i,\text{ms}}$ is the attractiveness of the movie in its opening week; $\gamma_{i,\text{ms}}$ is

[2] See the Technical Appendix available on the *Marketing Science* website (http://mktsci.pubs.informs.org) for complete details about the operationalization of the outside good.

the location of peak attractiveness; and $\beta_{i,\text{ms}}$ is the build/decay rate. We also have a direct interpretation of the effect of the regressor variables on these three parameters. Further, we can test for the presence of competitive effects by comparing the hyperparameters ($\Delta_{\text{ms}}$, $\theta_{\text{ms}}^{\text{Studio}}$, $\theta_{\text{ms}}^{\text{Genre}}$) obtained in our market share model (6) with the hyperparameters obtained from the following model (henceforth called the demand model), where weekly box-offices sales ($S_{it}$) are regressed for each movie independently of other movies. Note that in Equation (7) and elsewhere, the subscript d denotes that the parameters refer to the demand specification of the model as opposed to the market share model

$$S_{it} = \eta_{i,\text{d}} w_{it}^{\gamma_{i,\text{d}}/\beta_{i,\text{d}}} e^{(1-w_{it})/\beta_{i,\text{d}}} + \varepsilon_{it\text{d}} \tag{7}$$

with

$$\begin{aligned} \eta_{i,\text{d}}^* &= \ln(\eta_{i,\text{d}}), \\ \gamma_{i,\text{d}}^* &= \gamma_{i,\text{d}}, \\ \beta_{i,\text{d}}^* &= \ln(\beta_{i,\text{d}}), \\ \begin{bmatrix} \eta_{i,\text{d}}^* \\ \gamma_{i,\text{d}}^* \\ \beta_{i,\text{d}}^* \end{bmatrix} &= X_i \Delta_{\text{d}} + \theta_{i,\text{d}}^{\text{Studio}} + \theta_{i,\text{d}}^{\text{Genre}} + \Sigma_{\Delta_{\text{d}}}, \\ \varepsilon_{it\text{d}} &\sim \text{i.i.d.}\, N(0, \sigma_{\text{d}}^2). \end{aligned} \tag{8}$$

We can compare parameters across models because the inclusion of an outside good in the logit specification makes $S_{it}$ and $M_{it}$ perfectly correlated. Hence, the difference (across models) *in relative order* of the hyperparameters (within model) can be attributed to the coexistence of other movies at the box office.

## 3. Empirical Analysis

Data were collected for all movies released domestically between March 31, 1995, and June 25, 1998. During this period, 825 movies were released in the United States. We collected weekly box-office sales ($S_{it}$, in dollars) for each movie as well as a set of descriptive variables (obtained from EDI Nielsen, LNA, and IMDb.com). Our covariates are similar to the ones used in past research (e.g., Zufryden 1996, Neelameghan and Chintagunta 1999, Lehmann and Weinberg 2000, Elberse and Eliashberg 2003); we make use of media expenditures, screens,[3] critics rating, actor and director *Star Power*,[4] sequel, movie genre, and distributor. As mentioned previously, we model genre and distributor differently from previous researchers in that we make use of the massively categorical methodology by Steenburgh et al. (2003) to estimate the impact of each genre and distributors.

To both mitigate the problem of the "independence of irrelevant alternatives" property inherent in logit models and directly incorporate the effect of competition from closer substitutes, we constructed two competitive variables, *N*Genre and *N*MPAA, which reflect the number of movies of the same genre and the number of movies with the same MPAA rating that were launched at the same time as the movie of consideration. A negative coefficient on these variables would allow for the impact of close substitutes on the utility of a movie, thus reducing the relative attractiveness of two or more movies of the same genre/rating that run concurrently. While arguably ad hoc, this is a more parsimonious method than using a nested logit model, because we would need two layers of nesting, and 65 times as many hierarchical coefficients (13 genres $*$ 5 MPAA ratings).

Finally, to help differentiate between potential blockbusters and sleepers, we included the reel length of the movie (runtime, in minutes) and a dummy variable that indicates whether a movie is a rerelease (i.e., is in a second or third run). Indeed, "art" movies tend to be longer than blockbusters and rereleases do not have the same appeal as new ones. We did not include a categorical variable for MPAA rating as its inclusion in the analysis did not yield any significant parameters. We should also note that media expenditures were missing for 32% of the movies in our data set. We addressed this limitation by using the Expectation Maximization algorithm (Dempster et al. 1977) to "fit" covariates to the missing data. We also used total box office for all movies for the four years preceding our data set as lagged variables for seasonality.

### 3.1. Estimation Method

Both the market share (5) and the demand (7) models were estimated using a Markov Chain Monte Carlo (MCMC) algorithm, as briefly described in the Appendix and more fully in the Technical Appendix to this paper available at http://mktsci.pubs.informs.org. The first seven weeks of the data set were ignored, as the box-office figures for these weeks were incomplete (i.e., movies released the week prior to March 31, 1995, would still be in the theater at the beginning of our data set but were not known to us). This left us with 162 weeks of usable data. We should also note that the number of moviegoers in a given week is extremely large (in excess of $10^7$ on a busy weekend). Hence, instead of modeling this directly as a logit model, we use a Poisson transform approximation (Baker 1995; Technical Appendix). Finally, the

[3] We only include the number of opening weekend screens in our model. We do not take into account how the number screens evolve over time.

[4] Earlier paper (e.g., Sawhney and Eliashberg 1996) traditionally used a dummy variable to indicate the presence or absence of a major star in the movie. We make use of the more recent (1998) Hollywood Reporter Star and Director Power indices that provided ratings on a 1 to 100 scale.

number of lags used to estimate the demand of the outside good was set to three,[5] and the maximum demand for movies ($\bar{S}$) was set to 1.2*, the largest weekly value of the data set.[6] When fitting the lags, care was taken to ensure that holiday weeks matched from one year to another.

The MCMC algorithm ran for a burn-in period of 30,000 iterations and an estimation period of 60,000 iterations. We checked the stability of the parameters by comparing parameter estimates from the first 20,000 postburn-in draws with those obtained in the last 20,000 draws. Because the very small movies showed particularly noisy sales patterns with very few weeks in the theater (often less than three), we restricted our analysis to the movies that constituted the top 95% of box-office sales (404 movies).[7]

To test the soundness of our model, we also fitted Bayesian versions of both the Bass model and BOXMOD (i.e., allowing for a hierarchical regression structure across movies using the same regressors as the ones used for our model). In addition, to ensure that the results were not spurious (e.g., because of overfitting of the data), we performed an out-of-sample analysis, where we dropped the last 52 weeks of the data set and re-estimated the model using only weeks 1 to 110. We then used the estimates produced by this estimation, as well as the covariates for the movies present in week 111, to predict sales by movie for week 111. This procedure was repeated week by week for the last 52 weeks of the data set (e.g., to predict week 112, we used all data up to week 111). The fit statistics for both the 162-week models and the out-of-sample models, reported in Table 1, provide support for our model.

### 3.2. Model Fit

In this section, we demonstrate the effect of both modeling market share, and of using our adapted gamma to describe the evolution of market share. First, we consider the performance of BOXMOD, Bass, and our adapted gamma within the context of the *demand model*. In terms of total box office, we find that our adapted gamma performs marginally better than Bass and BOXMOD (mean absolute prediction error (MAPE) of 6.03% versus 6.23% and 7.21%, respectively, as shown in Table 1). However, in terms of weekly estimates, our demand model significantly outperforms the other two models—especially for the opening week (MAPE of 38.7% versus 181.2% and 135.8%). Moreover, when we use our adapted gamma within the *market share model* specification, the fit was further improved; providing a significant reduction in MAPE of 40%–50% when compared to Bass and BOXMOD (3.66% versus 6.23% and 7.21%). Thus, in terms of all statistics, our adapted gamma on its own shows a substantially better model fit, and using a market share model further improves the results.

**Table 1 Model Fit Comparison—MAPE for All Four Models**

| | Total BO (%) | Week-by-week BO (%) | Opening week (%) | Out of sample: One week ahead BO (%) |
|---|---|---|---|---|
| BOXMOD | 7.21 | 104.29 | 181.23 | 514.98 |
| Bass | 6.23 | 63.03 | 135.79 | 99.59 |
| Gamma demand | 6.03 | 47.64 | 38.74 | 75.15 |
| Gamma market share | 3.66 | 40.32 | 33.26 | 73.62 |

The benefits of using a market share model can be illustrated using a couple of examples. Figure 1 shows the fit of both the demand and the market share model for *Star Kid* (a typical *Blockbuster* diffusion pattern). Although the demand model fits the data reasonably well, it tends to underestimate the decay in the first week and overestimate the decay in the subsequent week. In contrast, the market share model offers a close fit, indicating that the change in decay rate can be explained by the impact of the other movies released in theaters at the same time as *Star Kid*. Figure 2 shows similar effects for a *Sleeper* movie (*Sliding Doors*). Here, we can see that the trough in week 4 and the peak in week 5 are not random variations, but rather changes in sales that can be explained when one takes all of the competing movies into account throughout the life of the movie. The market shares are driven by the strength of both the movie and of other movies available to the moviegoer over time.

Detailed fit statistics for the hierarchical regressions are shown in Table 2. To better understand the role of the various regressors, we computed conditional $R^2$ values for the categorical variables and for screens, as well as for the continuous variables as a whole (i.e., including screens).[8] We report on the interpretation of these fit statistics in the next section. Parameter estimates are shown in Table 3. Note that the studios are sorted in decreasing order of $\eta_{ms}$ intercept shifts. Further, we do not report the parameter estimates for the Genre categorical variable as it has very few significant parameters. A complete set of results is available from the authors for the interested reader.

[5] There were negligible differences in fit between 3 and 4 lags. Hence, we chose the 3-lag model for reasons of parsimony.

[6] We tried different $\bar{S}$ but found the estimates to be insensitive to the actual value used.

[7] As a test of reliability, we kept the full data set and verified that the results from the two data sets were not markedly different. This was largely found to be the case. Hence, we only report the results for the 95% data set herein.

[8] The conditional $R^2$ shows the improvement in $R^2$ when the variable, or a set of variables, is added to the regression (i.e., conditional on the other variables being already accounted for). It is akin to a type III sums-of-square in traditional regression.

**Figure 1 Model Fit for a Blockbuster-Type Movie: *Star Kid***

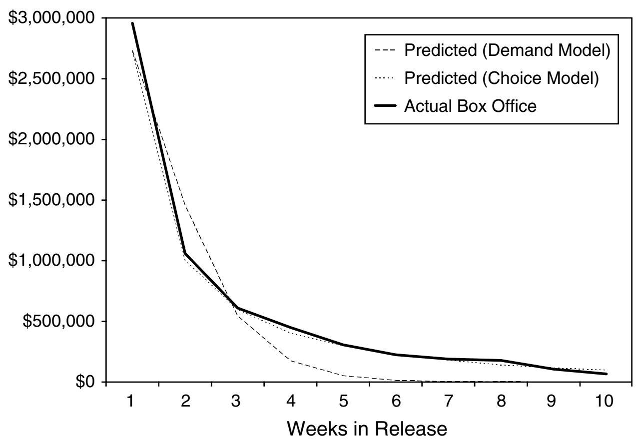



**Table 2 Measures of Hierarchical Variable Fit**

| | Demand | | | Market share | | |
|---|---|---|---|---|---|---|
| | First week $\eta_d$ | Peak $\gamma_d$ | Speed $\beta_d$ | First week $\eta_{ms}$ | Peak $\gamma_{ms}$ | Speed $\beta_{ms}$ |
| Hierarchical $R^2$ | 0.792 | 0.671 | 0.623 | 0.614 | 0.414 | 0.211 |
| Conditional $R^2$ | | | | | | |
| All continuous (including screens) | 0.489* | 0.417 | 0.543 | 0.223 | 0.163 | 0.124 |
| Genre | 0.013 | 0.021 | 0.210 | 0.005 | 0.009 | 0.017 |
| Studio | 0.016 | 0.018 | 0.050 | 0.096 | 0.033 | 0.084 |
| Screens | 0.333 | 0.309 | 0.051 | 0.147 | 0.084 | 0.005 |
| MAPE (Total BO) | 6.03% | | | 3.66% | | |

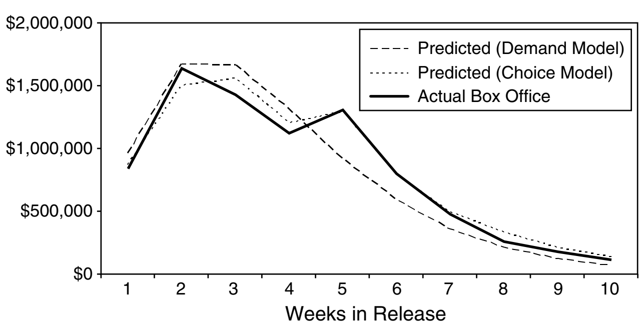
*0.489 indicates that removing the continuous variables from the regression would lead to an $R^2$ of $0.792 - 0.489 = 0.303$.

### 3.3. Estimation Results and Interpretation

First, let us look at the hierarchical $R^2$ and conditional $R^2$ (Table 2). We first note that, although the hierarchical regressions fit better in the demand model than in the market share model, the MAPE values (both in and out of sample) are better for the market share model. This indicates that the demand model might attribute effects to the covariates that should, instead, be attributed to competitive pressures or seasonality.

Second, we see an interesting pattern in the conditional $R^2$ associated with the continuous variables. Looking at the opening week sales parameter ($\eta_d$ and $\eta_{ms}$), for instance, the continuous variables uniquely contribute to 62% of the explained variance (0.489/0.792) in the demand model, while in the market share model, they only contribute to 36% of the explained variance (0.223/0.614). In particular, the Screens variable uniquely accounts for 42% of the explained variance in the demand model and only 24% in the market share model. In contrast, the Studio categorical variable would be viewed as superfluous in the demand model, but important in the market share model. Similar patterns can be found in the hierarchical regressions for the other two parameters. It can also be found looking at the continuous variables. Only 18 of the 33 coefficients on the continuous variables are significant in the demand model while 25 are in the market share model.

**Figure 2 Model Fit for a Sleeper-Type Movie: *Sliding Doors***




To better understand why screen loses half of its predictive power when incorporated in the choice model, we need to consider the timing game played by studios. As mentioned earlier, studios release big movies during "big weekends." Further, studios avoid going head to head against big movies. This means that when they are launched, the big movies have little competition, and thus can garner a large proportion of the available screens. This creates a correlation between screens and total movie demand as big movies are launched during big weeks with a large number of screens. For our data, the correlation is 0.14 ($p = 0.0038$). Thus, in the demand model, Screens accounts not only for the true effect of screens, but also for some seasonality effect, thereby inflating its importance. When used in the choice model, the effect of seasonality is already accounted for by the outside good, and thus the effect of Screens is much closer to its true effect. This shows that looking at movie releases independently from the other movies available to consumers may be misleading.

Opening week screens and media spending are factors that strongly affect moviegoers, and that movie studios can still influence close to the release date. One possible argument is that a studio could try to recoup its losses by hyping a bad movie and releasing it in as many theaters as possible. However, this is a short-term strategy that might backfire in the long run if unhappy theater operators refuse to give wide distribution to subsequent movies from the hyping studio. In our model, we see positive coefficients for Media on all three parameters. Thus, an increase in media spending is positive on all fronts. It leads to higher opening week sales, longer legs, and slower sales decay. Nelson (1975) argues that advertising should be a credible source of information, in that it is only worth advertising if the product is truly of

**Table 3 Parameter Estimates**

| | Demand model | | | Market share model | | |
|---|---|---|---|---|---|---|
| | First week $\eta_d$ | Peak $\gamma_d$ | Speed $\beta_d$ | First week $\eta_{ms}$ | Peak $\gamma_{ms}$ | Speed $\beta_{ms}$ |
| Continuous | | | | | | |
| Media | *0.098*[*] | 0.054 | *0.270* | *0.150* | *0.199* | *0.127* |
| Screens | *0.865* | *−0.822* | *0.149* | *0.629* | *−0.857* | *0.104* |
| TVGen | 0.026 | *0.070* | *0.133* | *0.098* | *0.281* | 0.013 |
| Actor *Star Power* | −0.004 | 0.045 | 0.041 | 0.011 | *−0.139* | *0.074* |
| Director *Star Power* | *0.084* | −0.028 | −0.003 | 0.015 | *0.111* | *−0.045* |
| Runtime | 0.014 | *0.071* | 0.040 | −0.045 | 0.061 | *0.056* |
| NGenre | −0.018 | 0.003 | *−0.051* | *−0.082* | *−0.076* | 0.008 |
| NMPAA | *−0.029* | *0.056* | −0.013 | *−0.160* | *0.082* | *−0.042* |
| Sequel | *0.038* | −0.009 | 0.010 | *0.063* | 0.017 | −0.009 |
| Rerelease | −0.010 | *−0.081* | *−0.160* | *−0.098* | *−0.333* | *−0.047* |
| Intercept | *15.534* | *1.114* | *0.281* | *−4.100* | *0.601* | *0.684* |
| Studio | | | | | | |
| TRIMARK | 0.197 | −0.303 | −0.043 | 1.221 | −0.214 | 0.071 |
| TWENTIETH CENTURY FOX | *0.214* | −0.087 | −0.128 | *1.046* | −0.457 | −0.077 |
| NEW LINE | *0.245* | −0.179 | −0.045 | *0.960* | −0.360 | −0.008 |
| UNIVERSAL | 0.157 | *−0.277* | −0.116 | *0.958* | *−0.492* | −0.068 |
| MGM/UA | −0.022 | −0.076 | *−0.288* | *0.950* | −0.254 | *−0.358* |
| PARAMOUNT | 0.154 | −0.138 | −0.122 | *0.887* | 0.130 | *−0.341* |
| DREAMWORKS SKG | 0.129 | 0.147 | −0.022 | *0.887* | 0.052 | −0.031 |
| WARNER BROTHERS | 0.112 | −0.031 | *−0.198* | *0.883* | −0.178 | *−0.239* |
| POLYGRAM | 0.257 | −0.297 | −0.003 | *0.825* | −0.134 | 0.033 |
| BUENA VISTA | *0.172* | *−0.258* | −0.141 | *0.781* | −0.289 | *−0.198* |
| SONY PICTURES | 0.058 | −0.056 | −0.118 | *0.684* | 0.140 | *−0.382* |
| ORION | 0.000 | −0.139 | −0.198 | 0.681 | −0.796 | 0.298 |
| SAVOY | 0.264 | −0.298 | −0.041 | 0.680 | −0.493 | 0.322 |
| MIRAMAX | 0.013 | −0.066 | −0.025 | 0.169 | −0.071 | *0.232* |
| MACGILLIVRAY FREEMAN FILM | −0.040 | −0.016 | 0.169 | 0.125 | −0.043 | −0.014 |
| ARTISAN ENTERTAINMENT | −0.262 | 0.212 | −0.073 | −0.105 | −0.046 | 0.098 |
| GOLDWYN ENTERTAINMENT | −0.154 | −0.008 | 0.114 | −0.412 | 0.163 | 0.359 |
| FOX SEARCHLIGHT | −0.215 | 0.334 | 0.229 | −0.437 | 0.591 | 0.340 |
| GRAMERCY | *−0.368* | 0.455 | 0.006 | −0.477 | 0.280 | 0.010 |
| SONY CLASSICS | −0.241 | −0.058 | 0.112 | *−2.117* | 0.415 | 0.050 |
| FINE LINE | −0.284 | 0.570 | *0.580* | *−3.269* | *1.266* | 0.107 |
| OCTOBER FILMS | −0.387 | 0.569 | 0.349 | *−4.918* | 0.790 | −0.206 |

[*]Underlined coefficients are significant at the 0.05 level.

good quality; otherwise a poor experience will reduce future period sales, thereby making the cost of advertising too high. Unsurprisingly, heavy media expenditure leads to higher opening week sales—but contrary to the "hyping" argument above, we find that media expenditure also leads to improvement in sales over the remaining period. If one computes a media arc-elasticity by simulating the impact of a 10% increase in media spending for any one movie (holding the other movies constant), we see an increase in opening week sales of 3.07%, (std. dev. 1.18%) and an increase in total box office of 6.61% (std. dev. 2.60%). Hence, the long-term effects of advertising are larger than the short-term effects. This is in line with Nelson's (1975) argument—the studios understand which movies to advertise heavily and which not to. However, the impact occurs here not because of repeat purchase but because of favorable word of mouth.

In the case of Studio, one would view the movie studios as undifferentiated entities in the demand model as opposed to their differential impacts in a model accounting for seasonality and competition. One must be careful when interpreting the Studio intercepts (see bottom of Table 3). In traditional logit analysis the brand intercepts are often interpreted as brand preference indicators. A high coefficient for a brand intercept; say, Tide, indicates that ceteris paribus consumers prefer Tide. This interpretation hardly applies in the movie studio context as moviegoers seldom make a movie choice based on the studio that released the movie. We see the Studio intercepts more as indicators of the internal competencies of each studio and their ability to leverage them. With this point of view, we make the following two observations.

First, the larger movie studios often have two distinct divisions: one for the *mainstream* releases and one for the smaller *art*-type movies. For instance, Fox Filmed Entertainment releases movies both under its 20th Century Fox label (mainstream release) and its Fox Searchlight label (independent and foreign movies). Similarly, we find New Line and Fine Line and MGM/UA and Goldwyn Entertainment. Looking at the Studio intercepts in the bottom of Table 3, we see that these large studios have entirely different parameters for their mainstream and their art labels. The mainstream labels have large positive first week intercepts (i.e., large opening weeks), negative peak intercepts (i.e., blockbuster type), and negative decay rate intercepts (i.e., quick decay rate), while the art labels have intercepts of opposite signs (i.e., sleepers with small opening weeks and more gradual change in sales from week to week). This suggests clear and successful strategic decisionmaking by the studios.

Second, we can compare the studio intercepts in the market share and the demand model for the opening week parameter ($\eta_{ms}$ and $\eta_{d}$, respectively). This will indicate whether the studios are successful in their release timing. Indeed, take the hypothetical situation of two studios (A and B) that have identical opening week parameters in the market share model (i.e., $\eta_{msA} = \eta_{msB}$), but have different parameters in the demand model (say $\eta_{dA} > \eta_{dB}$). This would imply that, holding everything else constant, studios A and B produce movies of the same quality, but that studio A is better at timing its releases than studio B, because it generates more revenues from movies with the same market attractiveness. This allows us to apply the theoretical findings of Krider and Weinberg (1998). They show that in a highly seasonal world, studios should release their big movies in direct competition to their competitors' big movies and release their small movies in less intense weeks. With this framework in mind, we can see that among the big studios, 20th Century Fox and New Line seem to time their release more effectively. They have positive and significant intercepts in both the market share and the demand model. In contrast, MGM/UA and Universal seem to be too ambitious and release their movies against bigger ones and lose out (they have positive and significant market share intercepts, but their demand intercepts are not). A clear distinction can be made between New Line and Universal in that they have almost identical first week parameters in the market share model, but New Line's demand parameter is significantly larger than zero while Universal's is not. This suggests that New Line is more adept at launching its movies.

At the other end of the spectrum, October Films seems to behave extremely opportunistically, releasing its movies in the least competitive weeks (it has an extremely low market share intercept, but its demand intercept is not significantly lower than the other studios in its peer group). Although these findings may not prove that Krider and Weinberg (1998) are correct in an absolute sense, they indicate that Hollywood behaves in accordance with their model.

An interesting pattern emerges when comparing the Star Power coefficients across the demand and the market share model. For actors, the coefficients in the demand model are not significant. This probably does not mean that actors do not matter, but more likely that studios support their actors in proportion to their star power, leading to no marginal impact of the star. Indeed, the correlations between the Actor Star Power index and the media and screen allocation variables are 0.41 and 0.25, respectively. We see much lower correlations for the Director's Star Power index (0.32 and 0.00), which indicates that studios support directors less than actors. The reasons for this difference in support may be found in the results of the market share model. Indeed, we see a negative peak coefficient and a positive decay coefficient for actors (i.e., more blockbuster-type diffusion patterns), while we see the opposite for directors (i.e., more sleeper-type diffusion). Actors have a direct effect on movie attendance, leading viewers to watch the movie earlier in its release, while directors have a more indirect effect on consumers: good directors make good movies, good movies have positive word of mouth, positive word of mouth delays peak sales. This is corroborated by the larger correlation between Director Star Power and movie rating (0.38) than between Actor Star Power and rating (0.21).

Finally, consider the *N*Genre and *N*MPAA variables. In the market share model, we find that releasing a movie in a week when other movies of the same genre or the same MPAA rating are released negatively affects initial sales. This is to be expected. However, looking at the peak parameters paints an interesting picture. *N*Genre has a negative peak coefficient while *N*MPAA has a positive coefficient. Thus, releasing a movie against other movies of the same genre adversely affects box-office performance, both in the short and the long term; however, releasing a movie against movies of the same MPAA rating hurts its sales in the beginning, but there is a displacement effect such that in the long run, the sales loss is less severe.

## 4. Conclusion

The goal of this paper is to study box-office sales in a market share context rather than studying movies in isolation of each other. To do so, we propose a *sliding-window* logit model combined with a *gamma* diffusion pattern, with parameters modified to enhance interpretability. This is implemented in connection with a hierarchical Bayes framework, with massively categorical variables, to both pool information across movies and extract information from the large number of studios involved with movie release.

Using this approach, we show that properly accounting for the set of movies available at the box office at any given time not only provides a better fit of the data, but also leads to a better understanding of the drivers of movie market share. We find that studios seem to be proficient at choosing which movies or actors to push. We find that advertising spending follows Nelson's (1975) concept of advertising as information. We also show that the large studios are correct in their product segmentation of mainstream versus artistic movies, and that many studios behave in accordance with Krider and Weinberg's view (1998) of release timing. We demonstrate that the impact of screens on movie sales may be lower than previously thought, as screens act as a proxy for seasonality in models that do not incorporate competition.

We demonstrate that actors have a direct effect on consumer choice by leading viewers to see a movie earlier in its release, while directors have a more indirect effect on consumers. Finally, our model shows that releasing a movie against other movies of the same genre hurts sales all around; releasing a movie against movies of the same MPAA rating hurts in the beginning, but there is a displacement effect, whereby the long run loss of sales is less severe.

A limitation of our paper is that we have not incorporated endogeneity in a systematic manner (see, for example, Elberse and Eliashberg 2003, Desiraju et al. 2005 for a different diffusion process). This may induce bias in some of our parameters. Our focus is on the issues of modeling market share, heterogeneity, and interpretable diffusion models. Further, although we model seasonal effects, we do not decompose them into endogenous and exogenous effects, as is done in a working paper by Einav (2003). This prevents us from making inferences regarding the origins of seasonality.

### Acknowledgments
The authors thank The UCLA Center for Communication Policy for financial support and Olav Sorenson for his help with the data.

## Appendix. Model Specification

### 1. Demand Model
All models were estimated using Markov chain Monte Carlo simulation (MCMC). A complete model specification is available online at the *Marketing Science* website at http://mktsci.pubs.informs.org. We limit our discussion to the specification of the various distributions used. In §1.1, we describe the conditional distributions of $\eta_{i,\mathrm{d}}$, $\gamma_{i,\mathrm{d}}$, and $\beta_{i,\mathrm{d}}$. These distributions are nonconjugate and thus are handled through a Metropolis step. We describe the distribution of the parameters of the hierarchical regression ($\theta_{\mathrm{d}}^{\mathrm{Studio}}$ *and* $\theta_{\mathrm{d}}^{\mathrm{Genre}}$) in §1.2 and §1.3. As these distributions are conjugate, we handle them with a Gibbs step. For notational convenience, we define $m$ as the number of movies in the data set ($m = 404$) and we use $i$ to index movies, $t$ to index time (162 weeks), and d to refer to the demand model.

#### 1.1. Conditional Distribution of $\eta_{i,\mathrm{d}}$, $\gamma_{i,\mathrm{d}}$, $\beta_{i,\mathrm{d}}$
In §3 of the paper, we defined the demand model (in Equation (7)) as:

$$S_{it} = \eta_{i,\mathrm{d}} w_{it}^{\gamma_{i,\mathrm{d}}/\beta_{i,\mathrm{d}}} e^{(1-w_{it})/\beta_{i,\mathrm{d}}} + \varepsilon_{it\mathrm{d}}.$$

Given this specification, the likelihood function for movie $i$ is given by:

$$L(\eta_{i,\mathrm{d}}, \gamma_{i,\mathrm{d}}, \beta_{i,\mathrm{d}}) \sim \left(\prod_t \left(\exp\left(-(S_{it}-\hat{S}_{it})^2/2\sigma_{\mathrm{d}}^2\right)\right)^{I_{it}}\right) \cdot \exp\left(-\frac{1}{2}\left(\begin{bmatrix}\eta_{i,\mathrm{d}}^*\\ \gamma_{i,\mathrm{d}}^*\\ \beta_{i,\mathrm{d}}^*\end{bmatrix} - \begin{bmatrix}\hat{\eta}_{i,\mathrm{d}}^*\\ \hat{\gamma}_{i,\mathrm{d}}^*\\ \hat{\beta}_{i,\mathrm{d}}^*\end{bmatrix}\right)' V_{\mathrm{d}}^{-1}\left(\begin{bmatrix}\eta_{i,\mathrm{d}}^*\\ \gamma_{i,\mathrm{d}}^*\\ \beta_{i,\mathrm{d}}^*\end{bmatrix} - \begin{bmatrix}\hat{\eta}_{i,\mathrm{d}}^*\\ \hat{\gamma}_{i,\mathrm{d}}^*\\ \hat{\beta}_{i,\mathrm{d}}^*\end{bmatrix}\right)\right),$$

where

$$\begin{bmatrix}\hat{\eta}_{i,\mathrm{d}}^*\\ \hat{\gamma}_{i,\mathrm{d}}^*\\ \hat{\beta}_{i,\mathrm{d}}^*\end{bmatrix} = z_i\Delta_{\mathrm{d}} + \theta_{i,\mathrm{d}}^{\mathrm{Studio}} + \theta_{i,\mathrm{d}}^{\mathrm{Genre}}.$$

#### 1.2. Conditional Distribution of $\theta_{\mathrm{d}}^{\mathrm{Studio}}$
We use a multidimensional version of the massively categorical approach suggested in Steenburgh et al. (2003) to allow for an intercept shift ($\theta_{\mathrm{d}}^{\mathrm{Studio}}$) on each parameter at the studio level.

Let $\beta_{i,\mathrm{d}}^{\mathrm{Studio}} = [\eta_{i,\mathrm{d}}^* \ \ \gamma_{i,\mathrm{d}}^* \ \ \beta_{i,\mathrm{d}}^*]^{\mathrm{T}} - \theta_{i,\mathrm{d}}^{\mathrm{Genre}} - z_i\Delta_{\mathrm{d}}$ be that component of the coefficients not attributable to the continuous variables or genre. Let $\beta_{\mathrm{d}}^{\mathrm{Studio}}$ be a stacked matrix of $\beta_{i,\mathrm{d}}^{\mathrm{Studio}\prime}$ across all movies. Further, let $\tilde{\beta}_{s,\mathrm{d}}^{\mathrm{Studio}} = \sum_{i=1}^{N} \beta_{i,\mathrm{d}}^{\mathrm{Studio}} I_{\mathrm{Studio}_i=s}$ be the sum of the vectors of these partial coefficients over all movies produced by a particular studio, indexed here by $s$. Finally, let the number of movies represented by studio $s$ be $m_s$. Then

$$\theta_{s,\mathrm{d}}^{\mathrm{Studio}} \sim N\left((I(3)\times m_s + V_{\mathrm{d,Studio}}^{-1})^{-1}\tilde{\beta}_{s,\mathrm{d}}^{\mathrm{Studio}},\ (I(3)\times m_s + V_{\mathrm{d,Studio}}^{-1})^{-1}\right).$$

#### 1.3. Conditional Distribution of $\theta_{\mathrm{d}}^{\mathrm{Genre}}$
This proceeds exactly as for the studio coefficients in 1.2, except that the indexing is performed across movies that belong to each genre instead of across movies represented by each studio.

### 2. Estimation for the Market Share Model

The estimation of the market share model is similar in nature to the estimation of the demand model. The hierarchical structure on the parameters $\eta_{i,\text{ms}}$, $\gamma_{i,\text{ms}}$, and $\beta_{i,\text{ms}}$ is identical to the structure put on $\eta_{i,\text{d}}$, $\gamma_{i,\text{d}}$, $\beta_{i,\text{d}}$ except that we restrict $\eta_{i,\text{ms}}$ to be between 0 and 1 by using a logit transform, rather than log, since it can be interpreted in terms of expected market share. What is different here is that we model *market share* rather than demand; we introduce an *outside good*; and we use a *Poisson* approximation for the logit specification (see Baker 1995 and the Technical Appendix at http://mktsci.pubs.informs.org).

## References

Baker, S. G. 1995. The multinomial-Poisson transformation. *Statistician* **43** 495–504.

Dempster, A. P., N. M. Laird, D. B. Rubin. 1977. Maximum likelihood estimation from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B* **39** 1–38.

Desiraju, Ramarao, Harikesh Nair, Pradeep Chintagunta. 2004. Diffusion of new pharmaceutical drugs in developing and developed nations. *Internat. J. Res. Marketing* **21**(4) 341–357.

Einav, L. 2003. Gross seasonality and underlying seasonality: Evidence from the U.S. motion picture industry. SIEPR Discussion Paper 02-36, Stanford Institute for Economic Policy Research, Stanford, CA.

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

Eonline. 2002. MPAA: Never say "goldmember" again. http://www.eonline.com/news/items/0,1,9459,00.html.

Gelman, A., J. Carlin, H. Stern, D. Rubin. 1995. *Bayesian Data Analysis*. Chapman and Hall, London, UK.

Krider, R. E., C. B. Weinberg. 1998. Competitive dynamics and the introduction of new products: The motion picture timing game. *J. Marketing Res.* **35**(1) 1–15.

Lehmann, D. R., C. B. Weinberg. 2000. Sales through sequential distribution channels: An application to movies and videos. *J. Marketing* **64**(July) 18–33.

McGill, W. J., J. Gibbon. 1965. The general-gamma distribution of reaction times. *J. Math. Psych.* **2** 1–18.

Motion Picture Association of America (MPAA). 2004. U.S. entertainment industry: 2003 MPA market statistics. www.mpaa.org.

Neelameghan, R., P. Chintagunta. 1999. A Bayesian model to forecast new product performance in domestic and international markets. *Marketing Sci.* **18**(2) 115–136.

Nelson, P. J. 1975. The economic consequences of advertising. *J. Bus.* **48** 213–241.

Radas, S., S. M. Shugan. 1998. Seasonal marketing and timing new product introduction. *J. Marketing Res.* **35**(3) 296–315.

Red Herring. 1998. Make movies, not war. http://www.redherring.com/mag/issue50/war.html.

Sawhney, M., J. Eliashberg. 1996. A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Sci.* **15**(2) 113–131.

Shugan, S. 1998. Forecasting failure and success of new films. Working paper, University of Florida, Gainesville, FL.

Steenburgh, T. J., A. Ainslie, P. H. Engebretson. 2003. Massively categorical variables: Revealing the information in zip codes. *Marketing Sci.* **22**(1) 40–57.

Swami, S., J. Eliashberg, C. B. Weinberg. 1999. SilverScreener: A modeling approach to movie screens management. *Marketing Sci.* **18**(3) 352–372.

Zufryden, F. 1996. Linking advertising to box office performance of new film releases—A marketing planning model. *J. Advertising Res.* **36**(4).