Distribution Planning to Optimize Profits in the Motion Picture Industry

Barbara Somlo[1]
Business Analyis and Information, Amgen
One Amgen Center Drive M/S 27-3-C,
Thousand Oaks, CA 91320

Kumar Rajaram[2]
Reza Ahmadi[3]
UCLA Anderson School of Management
110 Westwood Plaza, Box 951481
Los Angeles, CA 90095-1481

Submitted: April 2009
First Revision: January 2010
Accepted: March 2010

Abstract

We consider the distribution planning problem in the motion picture industry. This problem involves forecasting theater-level box office revenues for a given movie and using these forecasts to choose the best locations to screen a movie. We first develop a method that predicts theater-level box office revenues over time for a given movie as a function of movie attributes and theater characteristics. These estimates are then used by the distributor to choose where to screen the movie. The distributor's location selection problem is modeled as an integer programming based optimization model that chooses the location of theaters in order to optimize profits. We tested our methods on realistic box office data and show that it has the potential to significantly improve the distributor's profits. We also develop some insights into why our methods outperform existing practice, which are crucial to their successful practical implementation.

Key words: Motion Picture Industry, Forecasting, Distribution Planning, Theater Selection, Optimization

[1] bsomlo@gmail.com; phone: (805) 313.6409 fax: (805) 480.9682
[2] Kumar.rajaram@anderson.ucla.edu; phone: (310) 825 4154 fax: (310) 206 3337
[3] Reza.ahmadi@anderson.ucla.edu; phone: (310) 825 2502 fax: (310) 206 3337

1. INTRODUCTION

The motion picture industry is an important sector of the U.S. economy. Movie releases at theaters generated approximately $9 billion in revenue in 2005, representing a nearly 20% increase since the beginning of the decade. In spite of this increase, the Motion Picture Association of America (MPAA) reports that only 9 of the 549 movies released in 2005 generated profit higher than $50 million (http://www.mpaa.org). The average cost of a studio-released movie was nearly $96 million in 2005, up from $54.1 million in 1995. The major component of this cost was the development and production budget, which averaged around $60 million in 2005, followed by marketing ($36 million), and distribution costs. Depending on the distribution strategy chosen by the movie distributor (e.g., platforming, wide, or saturation release), distribution costs can be up to $9 million and averaged around $3.83 million per movie in 2005. While distribution represents a smaller portion of a movie's total investment in comparison to development, production, and marketing costs, effective distribution is critical to box office success and ultimate financial return from a movie (Reardon, 1992; Thomas, 1998).

Distribution of a motion picture is handled by its distributor, who forms an important link in the motion picture industry supply chain (Figure 1). Examples of major distributors in the motion picture industry include Buena Vista (Walt Disney's distribution arm), Columbia (Sony's distribution arm), Universal, Paramount, Fox and Warner Brothers. The motion picture distribution industry is highly concentrated with these six distributors accounting for 70% of box office sales in the United States. The distributor secures rights from the producer, undertakes marketing (including advertising in television, local and national media) and has to choose at which theaters to screen the movie. The theaters are owned by exhibitors such as AMC, Regal, AVCO, General Cinema, and Mann. Historically, distributors have been more dominant and profitable than exhibitors in the motion picture industry. To achieve balance of power between these entities, there has been federal regulation and individual statutes passed by states. The aims of these statutes were to ensure that distributors picked exhibitors on a movie by movie basis using a competitive bidding process. A detailed overview of the history of anti-trust legislation in the motion picture industry can be found in Vany and Eckert (1991). However, with increasing overcapacity of screens and changes in technology that enables different channels of distribution such as DVD's, satellite and the internet, the balance of power is again being shifted in favor of the distributors. Scott (2005) provides a comprehensive discussion of these issues.

Distribution planning for a new movie is done far in advance of the actual release date. The studio (i.e., producer) announces the distributor several months before actual theatrical release of the movie. Then, the distributor solicits bids to exhibit the movie. In response, exhibitors contact the distributor with bids to screen the movie. These bids include a proposed set of theaters and their coverage territory, initial duration of play, and revenue sharing agreements. The distributor now needs to choose which exhibitors and theaters to use for the initial screening of the movie. This decision has to be made at least three months prior to the scheduled theatrical release of the movie. These decisions are then conveyed to the exhibitor. If the distributor picks a particular theater, this movie is scheduled to be shown there for the agreed period. If not, the exhibitor has the option to try other movies, as this planning takes place far in advance of when the movie is actually released in the theater. As the initial duration of play for the selected theaters expires (which is agreed by the contract), the distributor is free to add additional theaters as offers to show new movies are made by the exhibitors (Rothenberg 2003; McGrath, 2004).

INSERT FIGURE 1 ABOUT HERE

The distribution planning problem involves forecasting theater-level box office revenues for a given movie and using these forecasts to choose the best locations to screen a movie. Distribution planning in the motion picture industry is made difficult by a complex environment. Forecasting box offices revenues for a given duration is challenging as movies are experiential products (Hirschmann and Holbrook, 1982) and, consequently, it is difficult to forecast their audience appeal until they are already available in the theaters. This is because audience appeal varies widely depending on movie attributes such as genre, star presence, special effects, MPAA ratings, critical reviews, etc. Even experienced analysts in the industry have not been very successful in forecasting box office revenues (Table 1). While there has been some academic research on aggregate box office forecasting, disaggregate theater-level forecasting is even harder due to dissimilarity in location specific characteristics of theaters, such as amenities and demographics (i.e., median income, age, population density, etc.). This causes significant differences in revenue for the same movie in different markets (Table 2). All these factors conspire to make the theater-level box office revenue forecasting problem in the motion picture industry an extremely challenging problem. Even after the forecasts are generated, picking the optimal location of theaters to show a particular movie is a challenging problem due to the large number of movies released each year by multiple distributors and the abundance in potential theater locations. For example there were 500 movies released in the year 2005 there were more than 7,000 possible theater locations in the United States and Canada to show these movies.

INSERT TABLE 1 AND TABLE 2 ABOUT HERE

Despite the practical relevance and complexity of this problem, we have found nothing in the academic or managerial literature that describes how to conduct effective distribution planning in the motion picture industry. This paper presents a method for addressing these issues and addresses an important problem in entertainment operations management, an underrepresented area in the growing stream of research in service operations management (Apte et. al. 2008 and Spohrer and Maglio, 2008). Specifically, we have developed an empirical technique that provides a box office revenue estimate over time for a new motion picture at a selected theater. We use these estimates to model the distributor's location selection problem as an integer programming model. This model chooses the location of theaters to screen the movie in order to optimize the distributor's profits. We have tested our methods on actual industry data and show that our approach offers the potential to significantly improve box office profits for new movies.

This paper is organized as follows. In Section 2, we review the relevant literature. In Section 3, we develop an empirical method to estimate the theater-level box office revenue for a given movie. In Section 4, we formulate the distributor's location selection problem. We develop heuristics to solve this problem and also construct upper bounds to evaluate the quality of these heuristics. We present computational results in Section 5. We use these results to assess the performance of our forecasting method and the heuristics to solve the distributor's location selection problem. These results also provide insight into what affects theater-level box office revenue and the impact of these revenue estimates on the location choice decision. In Section 6, we test our methods on realistic box office data and show that it has the potential to significantly improve current industry practice. In the concluding section, we summarize our work and present future research directions.

2. LITERATURE REVIEW

There is extensive literature on motion pictures in the popular press and in the film and television areas (Bart, 2000; Vogel, 2001; Hayes and Bing, 2004); however much of this is descriptive in nature and relies heavily on anecdotal industry knowledge. There is a limited, but emerging, stream of academic research focused on the motion picture industry. Eliashberg et al. (2006) provide a comprehensive overview on the critical issues in practice, current research and future research directions in the motion picture industry. Areas of research include product diffusion (Neelamegham and Chintagunta, 1999; Elberse and Eliashberg, 2003), seasonal release patterns (Krider and Weinberg, 1998; Einav, 2002), ancillary markets (Lehmann and Weinberg, 2000), and contract design and competition (De Vany and Walls, 1996). In the broader context of the entertainment industry, there has been work on scheduling commercial videotapes (Bollapragada, Bussieck and Mallik, 2004), managing on-air advertisement inventory (Bollapragada and Mallik, 2008), media revenue management (Araman and Popescu, 2009) and theme park flow management (Rajaram and Ahmadi, 2003).

There has been a significant stream of research on *aggregate* box office forecasting of new motion pictures. Many previous studies have attempted to explain aggregate box office success as a function of movie attributes such as budget, star power, MPAA rating, release timing, and Academy Award nominations and winners (Litman and Ahn, 1998; Wyatt, 1994). Recent work focuses on the influence of a major distributor (Sochay, 1994), advertising and critical reviews on box office success (Zufryden, 1996; Eliashberg and Shugan, 1997; Zuckerman and Kim, 2003). Sawhney and Eliashberg (1996) develop a parsimonious aggregate forecasting model and test it on realistic data. However, none of this work considers *disaggregated* theater-level box office revenue forecasts for a new movie. There have been several papers on scheduling of screens with multiple movies (Swami et. al. 1999, Elisaberg et. al. 2009 and Dawande et. al. 2010). This is an important tactical problem tactical question once the location of the theater to screen a movie is chosen, but these papers do not consider the broader strategic question on which theaters to show the movie, as addressed in distributor's location selection problem.

This paper makes the following contributions. First, we develop a method to calculate detailed disaggregated theater-level box office sales forecasts, based on both movie attributes and theater characteristics. Second, unlike the work discussed above, we directly consider the distributor's location selection problem. Correct selection of theater location is essential to the ultimate box office success of a movie and we develop an optimization model to make this choice. Third, we test this model extensively on realistic data and show that it has the potential to significantly improve existing industry practice.

3. FORECASTING THEATER-LEVEL BOX OFFICE REVENUES

Consider a distributor who has to make box office revenue forecasts at each of the possible theaters where they could release a new movie. To provide a precise statement of this problem, we consider n possible theaters and let $j \in N = (1, \ldots, n)$ index the set of theaters. Let π_{ijt} define the expected box office revenue forecast when movie i is shown at theater j in week t, where $i \in M = (1, \ldots, m)$ indexes the set of movies and $t \in Q = (1, \ldots, q)$ indexes the set of time periods. Estimating π_{ijt} is critical for several reasons. First, when we tabulate total box office revenue across theaters over time (i.e., $\sum_{j=1}^{n} \pi_{ijt}, \forall t)$, we get the adoption pattern for movie i. This pattern provides crucial guidance for various strategic decisions made by the distributor such as determining the marketing budget and eventually the distribution strategy (i.e., platforming, wide, or saturation release) for the movie. This, in turn, is used in determining the maximum number of theaters across regions, the minimum play length at any theater, and, finally, in negotiating the revenue sharing contract during and after this minimum play length. Second, π_{ijt} is a key parameter in determining at which theaters the movie will be screened and, ultimately, its box office success.

However, estimating π_{ijt} is challenging for several reasons. First, it is difficult to understand which movie attributes and theater characteristics will affect theater-level box office revenue and how they do so. Second, it is challenging to estimate how this complex relationship between theater characteristics, movie attributes, and box office revenues changes over time. Finally, this estimation is difficult, as forecasting box office revenues requires an understanding of the individual moviegoer's decision process to see a given movie and incorporating this process into the estimation method.

Typical industry practice when forecasting box office revenues is to compare a new movie to recently released movies that are similar in *one* movie attribute and employ multiple, separate comparisons to study the effect of different movie attributes on box office revenues (Tannenbaum, 2001; Rothenberg 2003). In addition, some distributors also employ multiple regression analyses on a set of comparable attributes. While these procedures are simple and provide flexibility to incorporate subjective expertise, they are not very accurate. This is because this approach does not explicitly consider any of the above discussed aspects that make estimating π_{ijt} an immensely challenging problem. There are box office estimation models in the academic literature that incorporate some of these aspects. Most of these models run multiple regressions directly on box office revenue as a function of certain sets of movie attributes (Litman and Ahn, 1998). A major limitation of this method is that it only provides point estimates for box office revenues by assuming an unrestricted horizon for exhibiting the movie. In addition, this approach does not consider significant variations in box office revenues across time periods and differences in adoption patterns across movies. Sawhney and Eliashberg (1996) developed a parsimonious model to forecast a movie's box office success as a function of time. They employ an innovative method that incorporates an individual moviegoer's decision process to adopt (or see) a given movie and also consider the impact of different movie adoption patterns. However, the objective of this model is to estimate box office revenue at the *national* or aggregate level, and this approach cannot be used to provide *local* or disaggregate location specific, theater-level estimates.

To overcome the described challenges inherent in estimating π_{ijt}, we develop a four-step method. These steps are outlined in Figure 2. In Step 1, we extend the Sawhney and Eliashberg (1996) model to include location specific, theater-level characteristics. We estimate the parameters of this model using nonlinear regression on a historical database of realized box office revenues for selected movies. In Step 2, we use multiple regression models to link the estimated model parameters to movie attributes and theater characteristics corresponding to this historical database; then, we also consider any potential interactions between these variables. This step provides us with a function that estimates model parameters given a set of movie attributes and theater specific characteristics. In Step 3, we use this function to estimate the model parameters for a new movie given its attributes and the location specific characteristics of the theater under consideration. In Step 4, we use these estimates of the model parameters of Step 1 to estimate the box office revenue for the new movie at a given theater across time. Below, we describe each of these steps in detail.

INSERT FIGURE 2 ABOUT HERE

Step 1: Incorporation of Theater-Level Characteristics

In this step, we extend the Sawhney-Eliashberg (1996) model to include location specific, theater-level characteristics. Here, we assume that that an individual movie patron's decision process to choose whether to watch a movie depends on two independent sub-processes: (1) the decision to see a movie *p* in theater *j,* followed by (2), the decision to visit theater *j*. These processes are modeled as stochastic processes with stationary parameters λ_{pj} representing the time-to-decide parameter and γ_{pj} representing the time-to-act parameter. Then, the expected time to decide becomes ($1/\lambda_{pj}$) and the expected time to act becomes ($1/\gamma_{pj}$). Although it is plausible that the time-to-decide process is mainly influenced by movie attributes, we believe that the availability of the chosen movie in a theater that is acceptable to the patron affects the time-to-decide parameter. Once the individual has decided to watch a movie, the next decision is where to watch it, which is again influenced by theater characteristics.

Following the approach of Sawhney and Eliashberg (1996), the expected cumulative number of adopters of movie *p* at theater *j* by time τ can then be expressed as:

$$E[N_{pj}(\tau)] = N_{pj}P_{pj}(\tau) = \frac{N_{pj}}{\lambda_{pj} - \gamma_{pj}}\left[\left(\lambda_{pj} - \gamma_{pj}\right) + \gamma_{pj}e^{-\lambda_{pj}\tau} - \lambda_{pj}e^{-\gamma_{pj}\tau}\right] \quad (1)$$

where $N_{pj}(\tau)$ is the distribution of the cumulative number of adopters of movie *p* at theater *j* by time τ approximated using binomial distribution, N_{pj} is the maximum potential market size in the vicinity of theater *j*, and $P_{pj}(\tau)$ is the cumulative density function of the event when an individual decides to see movie *p* in theater *j* and acts on the decision by time τ. We estimate parameters N_{pj}, λ_{pj}, γ_{pj} by using nonlinear regression to fit (1) to historical data of theater-level box office revenues over time across a range of *w* movies at all the required theaters and let $p \in H = (1, \ldots, w)$ index these movies with historical data.

Step 2: Calibration of Regression Model Parameters from Historical Box Office Information

The second step in the estimation process is to connect the estimated value of parameters $\hat{N}_{pj}, \hat{\lambda}_{pj}, \hat{\gamma}_{pj} \ , \forall j \in N, \forall p \in H$ from Step 1 to movie attributes, theater characteristics, and their possible interactions. To do this, we rely on multiple regressions based on the historical box office information $\forall p \in H$ used in Step 1.

Let $\boldsymbol{Z_i} = (Z_{1i}, \ldots, Z_{Ai})$ define the movie attribute vector for movie *i*.. Here, movie attributes include aspects such as such as genre, star presence, special effects, MPAA ratings, critical reviews, etc. To incorporate theater characteristics, let $\boldsymbol{S_j} = (S_{1j}, \ldots, S_{Cj})$ define the vector of theater characteristics for theater *j.* These include location specific characteristics of theaters such as amenities and theater size, and demographics factors such as median income, age, population density, etc. Then, the multiple regression equations to estimate the regression coefficients are:

$$\hat{N}_{ij} = \alpha_N + \sum_{a=1}^{A} \beta_{N_a} Z_{ai} + \sum_{c=1}^{C} \delta_{N_c} S_{cj} + \sum_{a' \in \Omega'} \sum_{c' \in \Lambda'} \omega_{N_{a'c'}} Z_{a'i} S_{c'j} + \varepsilon_N , \quad (2)$$

$$\hat{\lambda}_{ij} = \alpha_\lambda + \sum_{a=1}^{A} \beta_{\lambda_a} Z_{ai} + \sum_{c=1}^{C} \delta_{\lambda_c} S_{cj} + \sum_{a' \in \Omega'} \sum_{c' \in \Lambda'} \omega_{\lambda_{a'c'}} Z_{a'i} S_{c'j} + \varepsilon_\lambda , \quad (3)$$

$$\hat{\gamma}_{ij} = \alpha_\gamma + \sum_{a=1}^{A} \beta_{\gamma_a} Z_{ai} + \sum_{c=1}^{C} \delta_{\gamma_c} S_{cj} + \sum_{a' \in \Omega'} \sum_{c' \in \Lambda'} \omega_{\gamma_{a'c'}} Z_{a'i} S_{c'j} + \varepsilon_\gamma . \quad (4)$$

Here, α_N, α_λ and α_γ denote the population intercepts and β_{N_a}, β_{λ_a} , and β_{γ_a} ($a \in \Omega = (1,\ldots,A)$), and δ_{N_c}, δ_{λ_c}, and δ_{γ_c} ($c \in \Lambda = (1,\ldots,C)$) represent the regression coefficients associated with vectors **Z** and ***S***, respectively. In addition, the interaction terms associated with regression coefficients are $\omega_{N_{a'c'}}$, $\omega_{\lambda_{a'c'}}$, and $\omega_{\gamma_{a'c'}}$, where $a' \in \Omega' \subseteq \Omega$ and $c' \in \Lambda' \subseteq \Lambda$. Finally, ε_N, ε_λ, and ε_γ are independent and identically distributed random error terms.

Step 3: Estimation of Model Parameters for a New Movie

Let $\tilde{N}_{sj}, \tilde{\lambda}_{sj}, \tilde{\gamma}_{sj}$ represent the model parameters for a new movie *s* in theater *j*. Once the regression coefficients of equations (2) through (4) are calculated, we use the known attributes $Z_s = (Z_{1s}, \ldots Z_{As})$ of new movie *s* and individual theater characteristics at theater *j* to estimate parameters $\tilde{N}_{sj}, \tilde{\lambda}_{sj}$ and $\tilde{\gamma}_{sj}$ as:

$$\tilde{N}_{sj} = \alpha_N + \sum_{a=1}^{A} \beta_{N_a} Z_{as} + \sum_{c=1}^{C} \delta_{N_c} S_{cj} + \sum_{a' \in \Omega'} \sum_{c' \in \Lambda'} \omega_{N_{a'c'}} Z_{a's} S_{c'j} \quad (5)$$

$$\tilde{\lambda}_{sj} = \alpha_\lambda + \sum_{a=1}^{A} \beta_{\lambda_a} Z_{as} + \sum_{c=1}^{C} \delta_{\lambda_c} S_{cj} + \sum_{a' \in \Omega'} \sum_{c' \in \Lambda'} \omega_{\lambda_{a'c'}} Z_{a's} S_{c'j}, \quad (6)$$

$$\tilde{\gamma}_{sj} = \alpha_\gamma + \sum_{a=1}^{A} \beta_{\gamma_a} Z_{as} + \sum_{c=1}^{C} \delta_{\gamma_c} S_{cj} + \sum_{a' \in \Omega'} \sum_{c' \in \Lambda'} \omega_{\gamma_{a'c'}} Z_{a's} S_{c'j}. \quad (7)$$

Step 4: Estimation of Theater-level Box Office Revenues for a New Movie

In this step, we use the estimates of $\tilde{N}_{sj}, \tilde{\lambda}_{sj}$ and $\tilde{\gamma}_{sj}$ from (5) through (7) in (1) to estimate $E[\tilde{N}_{sj}(\tau)]$, the expected cumulative number of adopters for new movie *i* at theater *j* until time τ, as:

$$E[\tilde{N}_{sj}(\tau)] = \frac{\tilde{N}_{sj}}{\tilde{\lambda}_{sj} - \tilde{\gamma}_{sj}} \left[\left(\tilde{\lambda}_{sj} - \tilde{\gamma}_{sj} \right) + \tilde{\gamma}_{sj} e^{-\tilde{\lambda}_{sj}\tau} - \tilde{\lambda}_{sj} e^{-\tilde{\gamma}_{sj}\tau} \right]. \quad (8)$$

Let ϕ_j be the ticket price at theater *j* and $t = \tau_2 - \tau_1$ be the time interval under consideration. Then, we calculate π_{sjt} as:

$$\pi_{sjt} = (E[\widetilde{N}_{sj}(\tau_2)] - E[\widetilde{N}_{sj}(\tau_1)])\, \phi_j \,. \qquad (9)$$

In Section 5, we conduct computational experiments to test the accuracy of this method.

4. THE DISTRIBUTOR'S LOCATION SELECTION PROBLEM

Once we forecast theater-level box office revenues, the distributor needs to use this forecast to choose at which theaters to show a new movie, in order to maximize profits. To address this problem, we consider n possible theaters and let $j, j' \in N = (1,\ldots, n)$ index the set of theaters. These movie theaters are located in r regions indexed by $r \in P = (1,\ldots, u)$.

Define the variables:

$$W_j = \begin{cases} 1 & \text{if theater } j \text{ is chosen to screen a given movie} \\ 0 & \text{otherwise} \end{cases}$$

We are given:

$K^{(MAX)}$: maximum number of theaters required across all regions

$K_r^{(MIN)}$: minimum number of theaters required in region r

$$L_{jr} = \begin{cases} 1 & \text{if theater } j \text{ is located in region } r \\ 0 & \text{otherwise} \end{cases}$$

$$b_{jj'} = \begin{cases} 1 & \text{if theater } j \text{ competes with theater } j' \\ 0 & \text{otherwise} \end{cases}$$

Here, parameter $K^{(MAX)}$ is defined by the type of distribution strategy (i.e., limited, platforming, wide, or saturation release) chosen for the particular movie. The distribution strategy is chosen to be consistent with the marketing budget for a given movie. Next, the regions chosen are often major metropolitan areas. The minimum number of theaters per region is often specified by past experience. Finally, parameters $b_{jj'}$ are derived from the coverage territory specified by the exhibitor in each territory. The coverage territory is defined by a set of competing theaters that need to be excluded in each territory, which in turn specifies $b_{jj'}, \forall j, j'$.

Recollect that π_{ijt} defines the expected box office revenue forecast when movie i is shown at theater j in week t, where $i \in M = (1,\ldots,m)$ indexes the set of movies and $t \in Q = (1,\ldots,q)$ indexes the set of time periods. Let $t_{ij} \in Q$ denote the duration chosen by the exhibitor to show movie i at theater j. Note that t_{ij} can optimally chosen using the models described in Swami et al. (1999), (2001) or Somlo (2005) and is affected by the minimum play length $P^{(MIN)}$ fixed by a contractual agreement between the distributor and exhibitor. Then, the distributor's expected box office profit during this period is given by $\theta_{ij}^{D} = \sum_{t=1}^{t_{ij}} s_{ijt}^{D} \pi_{ijt} - c_{ij}^{D}$ where c_{ij}^{D} is the distribution cost of movie i to theater j, and constant s_{ijt}^{D} represents the portion of revenues allocated to the distributor for movie i at theater j during week t. This factor depends on the duration chosen by the exhibitor and the nature of the contractual agreement between the exhibitor and distributor.

The Distributor's Location Selection Problem (DLSP) can be represented by the following binary integer program:

(DLSP) $$V_i^{D} = \text{Max} \sum_{j=1}^{n} \theta_{ij}^{D} W_j \qquad (10)$$

Such that: $$\sum_{r=1}^{u} \sum_{j=1}^{n} L_{jr} W_j \le K^{(MAX)} \qquad (11)$$

$$\sum_{j=1}^{n} L_{jr} W_j \ge K_r^{(MIN)}, \quad \forall r \qquad (12)$$

$$W_j + b_{jj'} W_{j'} \leq 1, \qquad \forall j, j' \neq j \tag{13}$$

$$W_j \in \{0,1\}, \qquad \forall j \tag{14}$$

Objective function (10) is chosen to maximize the distributor's total expected box office profits for movie *i* by the appropriate choice of theaters. Constraint (11) ensures that the total number of theaters selected to screen a movie does not exceed $K^{(MAX)}$. Constraints (12) guarantee that a set minimum number of theaters are picked for each region. Constraints (13) ensure that the distributor does not pick competing theaters. This is important because a distributor often chooses multiple exhibitors to show a movie. Consequently, to prevent dilution of sales at the selected theater, exhibitors require that competing theaters within the vicinity are not picked. Finally, 0-1 integrality of the variables is imposed by constraints (14).

The DLSP can be used by the distributor to determine the optimal set of theaters for the initial screening of a movie. As the minimum play length commitment for the theaters expires, the distributor is free to add additional theaters as new offers to show the movie are made by exhibitors. In this case, the distributor would need to consider the set of theaters in which the movie still has to be shown, reduce $K_r^{(\min)}$ at the appropriate regions, reduce $K^{(MAX)}$ by the total number of theaters where the movie is still being shown, remove the additional theaters for which constraint (13) will not be feasible, and resolve the DLSP. Note that this approach also can be used to include preferred theaters in the beginning or in later iterations of the DLSP. Here the preferred theaters are akin to the theaters where the movie still has to be shown. Such preferred theaters may be necessary to maintain a long-term relationship with the exhibitor. Also note that this model can be extended to multiple movies. Here, we would first need to index the parameters and variables of the DLSP by movie index *i*. Then, the objective function for the

DLSP for multiple movies would now be $\sum_{i=1}^{m} V_i^D$ and we would have constraints (11) through (14) for each movie *i*. Since this problem is decomposable by movie or index *i*, we could use the same solution procedure to solve the DLSP with a single movie.

Proposition 1: The DLSP is NP-Complete.

Proof: The maximum weighted independent set problem can be derived as a special instance of the DLSP by setting the coefficients $L_{jr}, \forall j, r$ to zero. Since it is known that the weighted independent problem is NP-Complete (Garey and Johnson, 2000), this reduction establishes that the DLSP is NP-Complete. ■

In light of Proposition 1, it is unlikely that we could solve large, realistic problems to optimality. In particular, we found in our computational analysis that we could not find solutions using leading commercial software tools such as the XPRESS and CPLEX solvers in GAMS (Brooke et al. 1992) when the number of theaters is large (over 1800 theaters) and when each theater has many competing theaters (averaging over seven per theater). Consequently, we elected to develop heuristics to solve such instances of this problem. We also present upper bounds to evaluate the quality of these heuristics.

4.1 Upper Bounds

To develop upper bounds on the DLSP, one could relax one or more of constraints (11) through (13) by introducing Lagrange multipliers and solving the resulting sub-problem optimally. Then, this sub-problem can be optimized over the multipliers to provide a tight upper bound. However, the upper bound from any such relaxation would be no smaller than a simple linear programming relaxation of the DLSP, in which we relax constraint (14) by allowing $W_j \in [0,1], \forall j$. This is because relaxations of the DLSP involving constraints (11) through (13) have the integrality property (Geoffrion, 1974), as established by the following proposition.

Proposition 2: Relaxations of the DLSP involving constraints (11) through (13) possess the integrality property.

Proof: We can represent the DLSP in a generalized matrix form as $\underset{x}{\text{Max}}\{fx \mid Ax \leq b, Cx \leq d, x \in X\}$, where A, b, C, d, and f are the appropriate matrices, $Ax \leq b$ represents the set of constraints we keep, $Cx \leq d$ represents the set of constraints we relax, and $x \in X$ represents the integrality constraints. Let Co$\{x \varepsilon X \mid Cx \leq d\}$ represent the convex hull formed by the constraints we relax. Since in the DLSP $K_r^{(MIN)}, K^{(MAX)} \in \boldsymbol{N}^+$, $\forall r$ and $L_{jr}, b_{jj'} \in \{0,1\}, \forall r, j, j'$ note that Co$\{x \in X \mid Cx \leq d\} = \{x \mid Cx \leq d\}$ for any relaxation involving constraints (11), (12), and (13). It follows from Geoffrion (1974) that any relaxation of the DLSP involving these constraints has the integrality property. ■

In light of Proposition 2, we generate an upper bound for the DLSP by solving its linear programming relaxation with $W_j \in [0,1], \forall j$.

4.2 Heuristics

In general, the solution provided by the upper bounds may not be feasible for the DLSP due to the violation of the integrality constraints (14). To achieve feasibility, we develop the following heuristics.

1. The Myopic Heuristic

In the myopic heuristic, we select theaters by first ignoring constraints (13) and optimally solving the resulting problem. Then, we develop an interchange procedure that is aimed at satisfying constraints (13). Recollect that $K_r^{(MIN)}$ represents the minimum number of regions required in region *r*, $K^{(MAX)}$ is the maximum number of theaters across all regions, and θ_{ij}^D is the distributors expected box office profit for movie *i* at theater *j*. This heuristic is formalized in the following steps.

Step 1: In each region, select $K_r^{(MIN)}$ theaters in descending order of θ_{ij}^D . This satisfies constraints (12). Remove the selected theaters from consideration. Sort all of the remaining theaters in descending order of θ_{ij}^D and pick an additional $\left(K^{(MAX)} - \sum_{r=1}^{u} K_r^{(MIN)} \right)$ theaters. Thus, constraint (11) is binding.

Step 2: We consider the theaters selected by Step 1 and look for those theaters that violate constraints (13). We first remove the violating theater with the lowest θ_{ij}^D . We then find a replacement theater which has the highest possible θ_{ij}^D without violating constraints (13). This procedure is repeated until all violating theaters across all regions are eliminated.

2. The Profit/Competition Heuristic

This heuristic attempts to select the most profitable theaters with the least number of competitors and consists of the following steps:

Step 1: Calculate the ratio $R_i = \theta_{ij}^D / \sum_{j' \neq j} b_{jj'}$ for each theater. R_i can be regarded as the scaled expected box-office profit for the distributor from movie *i* at theater *j*. The scale factor $1/(\sum_{j' \neq j} b_{jj'})$ decreases as $\sum_{j' \neq j} b_{jj'}$, the number of competing theaters for theater *j*, increases.

Step 2: For each region, select the theater with highest ratio and in the case of a tie select the theater with the lowest number of competing theaters. Note that by picking the theaters in this ratio, we implicitly reduce the number of competing theaters. This in turn ensures that it is more feasible to pick subsequent theaters without drastic reduction in revenues. Once this theater is selected, remove all the competing theaters from consideration. If, at this point, constraint (12) is satisfied, go to next region. If constraint (12) is not satisfied, pick the ratio with the next highest value and repeat this procedure. Continue until constraint (12) is satisfied. If this still does not lead to a feasible solution, restart this procedure with the theater with the next highest ratio and continue until this constraint is satisfied. Repeat this step for every region.

Step 3: Remove all theaters selected in Step 2 from consideration.

Step 4: Consider the theaters that have not been removed in Steps 2 or 3 and choose the remaining $K^{(MAX)} - \sum_{r=1}^{u} K_r^{(MIN)}$ theaters in decreasing order of $\theta_{ij}^D / \sum_{j' \neq j} b_{jj'}$. After each selection, remove the competing theaters associated with the chosen theater.

Note that this heuristic attempts to satisfy constraints (12) and (13) in Step 2, and constraint (11) and (14) in Step 4. In the next section, we test the performance of our method to estimate π_{ijt} and also the effectiveness of both these heuristics and the upper bound across a variety of data sets.

5. COMPUTATIONAL STUDY

The financial box office data required for the computational study was purchased from Nielsen Entertainment Data Incorporated (EDI) located in Beverly Hills, California. We selected theaters located within the continental United States and purchased data of weekly box office revenues for all movies played at a given theater between May 22, 2000 and May 25, 2001.[4] The time period was chosen to completely cover an entire major release period during summer. The sample consisted of 149 movies and 1,218 theaters. This sample of movies accounted for 98.42% of box office revenues during this period. In addition to purchasing financial data, we built two separate databases to collect information regarding movie attributes and theater characteristics corresponding to the movies and theaters in this period. These were created using Microsoft Access 2000. Data for the movie attributes were gathered from the Internet Movie Database (http://imdb.com/), Baseline Filmtracker (http://www.baseline.hollywood.com/), and Box Office Guru (http://www.boxofficeguru.com/), while data for the theater characteristics were gathered from the National Association of Theater Owners (http://www.natoonline.org/) and the U.S. Census Bureau (http://www.census.gov/).

5.1. Results of Estimation Method

We summarize our results corresponding to the sequence of steps in the estimation method outlined in Section 5.

---

[4] To check whether the sample represents a typical year of movie releases, we tested the difference between the proportions of movies released over time and across genres in the previous and following years. We could accept the null hypothesis at the 95% confidence level that the two sample population proportions are equal in each class.

Step 1: Incorporation of Theater-level Characteristics

We use nonlinear regression to approximate the model parameters: the maximum potential market size (N_j), time-to-decide (λ_j), and time-to-act (γ_j) for a given movie. To execute this regression, we used the Levenberg-Marquardt algorithm (Bates and Watts, 1988) of the NLIN procedure of SAS, a commercially available statistical software (SAS, 2003). To ensure convergence during these runs, it was critical to specify good starting values for the nonlinear regression. Therefore, we employed a grid search to obtain good starting values for the parameters. Results of the parameter approximation for selected pairs of movies and theaters are presented in Table 3.

INSERT TABLE 3 ABOUT HERE

From Table 3, we make three important observations. First, note that the magnitude of estimated box office revenue for the same movie can change significantly across theaters. For example, the estimated box office revenue for the movie WOMEN varied from $31,400 to $171,535. Thus, including location specific theater characteristics in the box office estimation procedure is an important aspect in optimizing box office profits. Second, different movies lead to vastly dissimilar revenues at the same theater. For example, consider the theater FENW in Table 3. The revenues for the three movies shown for the same duration vary from around $22,640 to $255,337. This confirms the intuition that movie attributes have a significant influence on theater-level box office revenues. Third, even when estimated box office revenues were similar, these could have been derived from very dissimilar parameter estimates and adoption patterns. For instance, the estimated box office revenue for movie LIESBTH at theater WYNN was $22,607 and the estimate for the movie VELN at theater FENW was $22,691; however, the estimate for LIESBTH at theater WYNN was based on $N = 22.994$, $\lambda = 26.168$, and $\gamma = 0.341$, while the box office estimate for VELN at theater FENW was based on

$N = 23.049$, $\lambda = 2.43$, and $\gamma = 2.043$. This meant that LIESBTH followed an exponential shaped adoption pattern, whereas VELN's adoption pattern was consistent with the shape of the Erlang-2 distribution. These differences in adoption patterns have very different implications for the distribution strategy of these movies and, ultimately, for parameter $K^{(MAX)}$.

Since the market size, time-to-act and time-to-decide parameters vary significantly across movies and theaters and lead to different adoption patterns, it is critical that they are estimated by incorporating the impact of both movie attributes and theater characteristics; however, we found that we could not develop generalizations or simple rules to determine how these aspects affected these parameters. To overcome this, we resorted to multiple regressions, described in the next step.

Step 2: Calibration of Regression Model Parameters from Historical Box Office Information

To run the multiple regression connecting the parameters of movie characteristics and theater attributes, we collected 1,218 parameter triplets $(\hat{N}_{pj}, \hat{\lambda}_{pj}, \hat{\gamma}_{pj})\ \forall p \in H$ from the nonlinear regression of Step 1. Note that each triplet corresponds to a theater across a range of movies. We divided these theater-triplets into two sets, so that each set had approximately the same number of movies. The first set is the calibration sample with 609 theater triplets and 75 movies used to calibrate the multiple regression coefficients. The second was the holdout sample with the remaining 609 theater-triplets and 74 movies, which was used to test the validity of the regression results in Steps 3 and 4. We employed the IML procedure, a multiple regression routine in SAS to run the regressions. The regression results showed that homoscedasticity (equal variance) was violated, so, to correct this problem, we transformed the dependent variables to their natural logarithm. The regression results are summarized in Table 4.

INSERT TABLE 4 ABOUT HERE

The regression results show that movie attributes and theater characteristics are good predictors of maximum theater-level box office revenues ($R_N^2 = 0.53$). As a comparison, we also ran a multiple regression model using only movie attributes. We found that the predictive power of the regression came down significantly ($R_N^2 = 0.25$). This provides strong support to include theater characteristics in our forecasting model. In addition, this regression in general possesses more predictive power for the time-to-act parameter (γ) than for the time-to-decide parameter (λ) ($R_\lambda^2 = 0.15, R_\gamma^2 = 0.27$). These results also provide interesting insight into which movie attributes and theater characteristics affect N, λ, and γ.

Movie attributes that significantly affect box office revenues include production budget, critics' reviews, genre, and release date. As expected, higher budget and positive critics' reviews add to box office success. Certain genres, specifically animation and fantasy, influenced box office revenues since fantasy and sci-fi movies usually cater to specialized crowds. These results were similar to those of Litman and Ahn (1998). We also found that spring release dates adversely impacted box office revenues, possibly because of springtime travel and the restart of outdoor activities. Significant theater characteristics affecting box office revenues included increased presence of competing theaters, amount of discount on the ticket price, median age, population density, and geographical location specified in one of seven, broadly classified regions in the United States. We found that the presence of competing theaters positively impacted box office revenues at a particular theater. While this result seems counter-intuitive, this could be due to clustering effect (Pinkse and Slade, 1998; Tannenbaum, 2001; Chisholm and Norman, 2002) in which the collective presence of stores offering the same or similar services allows better tapping into higher demand in urban areas. As expected, the size of the ticket discount was negatively correlated with box office revenues and explained why distributors often request that the number of discounted tickets be limited. We also found that median age was negatively correlated to box office revenue, while increased population density positively influenced box office revenue. Finally, the geographical location of a theater had a significant impact on box office revenue.

The time-to-decide parameter, λ, was influenced by movie attributes and theater characteristics. Several movie related attributes, such as fantasy and animation genres, star presence, and movies heavy in special effects, were positively correlated with λ. This is because these attributes pull audiences into theaters earlier, reducing the expected time to decide. On the other hand, restrictive MPAA ratings and winter release timing were negatively correlated with λ, since these attributes dampen interest and, thus, increase the expected time to decide. Significant theater characteristics included adult prices and number of screens. As anticipated, the magnitude of the discount on ticket prices was positively correlated with λ, as this reduced the time to decide. In addition, increasing the number of screens was negatively correlated with λ. This is because increasing the number of screens typically increased the time to decide presumably due to the perception that when there are more screens, the duration of movies would be longer and this could also reduce the chances of movies being sold out in subsequent weeks.

Movie attributes such as wider MPAA ratings and winter or spring opening dates were negatively correlated with γ. This is because these attributes increase the expected time to act. Conversely, genre (animation and fantasy) and sequels were positively correlated with γ, since they typically catered to special audiences whose expected time to act is smaller. A highly significant theater characteristic was the number of competing theaters within a five-mile radius.

As expected, this was positively correlated with γ since the previously described clustering effect could reduce the expected time to act.

Finally, we list the results on the interaction of theater related variables with movie related ones. None of the interaction terms proved to be significant predictors of box office revenue in our sample, but could be significant in other samples.

Steps 3 and 4: Estimation of Model Parameters and Theater-level Box Office Revenues for a New Movie

We define a new movie as a movie shown in theaters in the holdout part of our sample. To validate the multiple regression results, we estimated model parameters $\tilde{N}_{pj}, \tilde{\lambda}_{pj}, \tilde{\gamma}_{pj}$, calculated box office revenue estimates for the holdout sample, and compared those estimates to the actual, achieved box office revenue. To better assess the performance of the box office estimation procedure, we also developed a benchmark model. This was based on a multiple regression model directly running box office revenue against the complete set of movie attributes that was used for our model *without* including any theater characteristics. This benchmark model itself was an enhancement on industry practice that was based upon choosing *one* movie attribute per simple regression run and employing multiple separate regressions to study the effects of different movie attributes on box office result (Tannenbaum, 2001; Rothenberg, 2003). We observed heteroscedasticity (unequal variance) in the error terms; therefore, we transformed the actual box office measure to its logarithm. The predictive power of the benchmark model was significantly weaker than that of our model. In addition, we observed that critics' ratings were positively correlated with box office revenue estimates, while this estimate was negatively correlated to spring season release dates.

Across the entire holdout sample consisting of 609 theaters and 74 movies, the average forecast error of our method was 15%, while the average forecast error for the benchmark model was 60%. Thus, our method reduces average forecast errors from the benchmark model by 75%. Table 5 summarizes actual box office sales across all theaters, aggregate forecasts and forecast error (expressed as a percentage of actual box office sales) for a select sample of movies for our method and the benchmark model. We next use the estimates of box office revenues from our method to test the DLSP.

INSERT TABLE 5 ABOUT HERE

5.2 Results of the Distributor's Location Selection Problem

The parameters required for the DLSP, such as the portion of revenues allocated to the distributor, distribution costs, maximum number of theaters, and minimum number of theaters sought per region, were set based on specific movie level information. In addition, the key inputs to the DLSP were expected theater-level box office sales (i.e., π_{ijt}) and the duration of play (i.e. t_{ij}). We estimated π_{ijt} employing the procedure outlined in Section 3, while t_{ij} was calculated by the model described in Somlo (2005).

We used XPRESS, a mixed-integer programming solver in GAMS, to solve the DLSP. This generated optimal solutions for instances of the DLSP, when, on average, each theater had less than seven competing theaters. The average time for each run was around 53 seconds. However, we found that when each theater had more than seven competing theaters on average, GAMS could not solve the DLSP. This provided the motivation for developing the heuristics (i.e., lower bounds) to address this problem and upper bounds to evaluate the quality of the heuristics. A specialized Microsoft VisualStudio.Net program was written to calculate the lower bounds using the heuristics in Section 4.1. To derive the upper bound, we solved the DLSP as a linear program using XPRESS in which W_j is relaxed to be continuous between 0 and 1.

To examine how our models perform on larger problems, we used the reference data to construct larger problems with n theaters and m movies, where n = 3,000 and 6,000 and m = 150 and 300. We first analyzed the historical box office data and defined the probability distributions for the significant variables in the multiple regressions defined by (2) through (4). We then ran Monte Carlo simulations on these variables to generate expected theater-level revenues for these larger problems for the required choice of n and m; however, we observed that some of these simulations generated several gigabytes of data without providing additional insight on the performance of the forecasting technique and the DLSP. Therefore, after careful consideration, we elected to analyze the 3000-theater problem for the DLSP across 150 movies.

Table 6 summarizes some of our salient results from our computational tests. In this table, a row represents the solution technique used. These included the optimal solutions generated by GAMS, the upper bound generated by the LP relaxation in GAMS, and the lower bounds based on the myopic and greedy heuristics. Columns in the table represent the problem size of the DLSP represented by the number of theaters, movies, and the competition density, which is the average number of competing theaters per theater. Based on our discussion with several industry experts (Tannenbaum, 2001, Rothenberg 2003, McGrath, 2004) and to cover a broad range of scenarios, we picked the competition density to be 6, 11 and 15. The numbers in the body of the table describe the percentage gap of the given technique from a reference point, if that technique was successful in generating a solution for the given problem. Since the DLSP problems were solved to optimality at a competition density level of six competitors per theater, this was used as a reference point. However, GAMS was unable to generate optimal solutions for problems with higher competition densities. Consequently, the upper bound was used as this reference for the remaining problems.

INSERT TABLE 6 ABOUT HERE

Our computational results have been quite encouraging. From Table 6, observe that when we use the myopic heuristic for a competition density level of 15 competitors per theater, the average gap from the upper bound was 10.3%, while the corresponding gap with the greedy heuristic was 11.2%. As the competition density level decreased, the performance of both heuristics improved. For instance, the average gaps for the 6-competitors-per-theater problem reduced to 1.8% and 1.3% for the myopic and greedy heuristics, respectively. We also wanted to better understand the circumstances under which the percentage gaps change. This could provide us with insights into how to improve the upper bound and the heuristics. We observe from our analysis that these gaps were uniformly higher when the number of available theaters for selection was higher. Conversely, the gaps were significantly lower when the number of available theaters was lower. It is important to note that these gaps were reduced because the upper bound became tighter. These results show that the heuristics perform well across a range of data and there is scope to improve the upper bounds.

To test how sensitive the *value* of the heuristics were to estimates of theater-level box office revenues, we scaled π_{ijt} by $(1-x)$ and $(1+x)$ where $x = 0.1$, 0.2, and 0.3. Note that our scaling procedure resulted in 6 additional data sets for the 3,000-theater, 150-movie problem set at a density level of 15 competitors per theater. Table 7 summarizes the average gaps for the myopic and greedy heuristics. These results show that average gaps for the myopic heuristic ranged from 10.3% to 10.95%, while the corresponding gaps for the greedy heuristic ranged from 11.2% to 12.7%. These results show that the heuristics were not significantly sensitive to estimation errors in π_{ijt} and, thus, provide a reliable basis to address this problem. To see how sensitive the *solution*s of the heuristics were to the estimates of π_{ijt}, we compared the optimal theater locations selected across the heuristics for the scaled values of π_{ijt}. We found that, while the total number of locations and the composition of those locations in terms of theater types were stable, the actual locations proposed by the different solutions varied with differences of scale in π_{ijt}.

INSERT TABLE 7 ABOUT HERE

The stability in the value of the heuristics and the total number of locations can be reassuring, but can also be misleading to the movie distributor. It is reassuring because the distributor can select the number of theaters prescribed by its general distribution plan using reasonably accurate theater-level revenue forecasts over time. On the other hand, the stability might mislead the distributor into thinking that one has to only consider the same set of theaters across *different* movies. However, due to changes in movie attributes, the set of theaters may vary widely across movies. For instance, across the 150 movies, we found that on average only 43% of the theaters were common. This result reinforces the importance of including theater characteristics and movie attributes when determining the distribution plan.

We wanted to better understand the effects of minimum play length $P^{(MIN)}$ and consequently t_{ij} on the solution of the DLSP. Distributors consider this parameter vital toward achieving the desired exposure, which in turn will affect the financial potential of a movie. Consequently, they go to great lengths to ensure that the agreed-upon screen is allocated to the particular movie for the requested period of time. We tested our procedures with $P^{(MIN)}$ set to 3 weeks as the base case, and 2 and 4 weeks as alternative settings for the 3,000-theater, 150-movie, 15-competitors-per-theater problem. The sensitivity analysis of this parameter provided several interesting insights.

We compared the solutions with the alternative values of $P^{(MIN)}$ to the base case and found that a major shift occurs in the optimal theater locations when the minimum play length requirement was changed. At first glance, the change in the *number* of theaters and the actual selection were minimal, but a more detailed analysis on the new set of theaters revealed a significant difference in the *type* of theaters to target. The extent of this change in the types of theaters is surprising and goes unrecognized by distributors. For example, the total number of theaters selected by the DLSP showed a very modest increase from 740 to 750 theaters when we change $P^{(MIN)}$ from 3 to 2 weeks. However, more than 55% of the theaters recommended for the base case were replaced for the shortened commitment period. The new solution selected more mini-type theaters in neighborhoods with moderate income in contrast with the original solution's heavy dominance by multi- and mega-theaters in high income areas. When we change $P^{(MIN)}$ from 3 to 4 weeks, the total number of theaters decreased from 740 to 670, as the movie had to be shown for a longer duration, but, here again, more than 41% of the chosen theaters differed from the base case and where different theater types. Here mega and multi-theaters were preferred over mini-theaters. This shows that the minimum play length requirement strongly influences the type of theaters that needs to be chosen, and distributors should carefully examine the effect of changing $P^{(MIN)}$ on theater choice before agreeing to change it on a movie-by-movie basis. The DLSP provides a structured and robust basis for conducting this assessment.

6. APPLICATION

We have compared the methods detailed in this paper to the theater selection decisions made by motion picture distributors on realistic data for the 3,000 theater, 150-movie problem with the number of average competing theaters set to the highest level of 15 competitors per theater. We then ran the myopic and greedy heuristic to solve the DLSP for each movie and picked the best solution.

Next, we constructed a distribution plan for a given movie replicating the procedure that distributors would have used in practice, based on extensive discussions with several leading distributors (Rothenberg, 2003; McGrath, 2004; Molter, 2004). In this procedure, distributors first ranked theaters in decreasing order of historical revenues across all movies. If necessary, the distributors modified the initial ranking by weighting sales along with a key attribute rank. For instance, consider the case of when a movie is targeted towards a certain ethnicity (for example, African-American viewers). In this case, theaters in a region would be ranked in decreasing order of the proportion of that ethnicity in the vicinity of the theater's location. A final ranking will be developed by assigning weights to the sales and ethnicity rankings. Next, they picked the highest ranked theaters in each region while ensuring that the minimum number of theaters and competition constraints where met in each region. Finally, they also made sure that the total number of theaters across regions did not exceed the maximum number of theaters required, which was set based upon the distribution strategy derived from the marketing budget for each movie.

We wanted to compare our method with the distributor's procedure. To ensure that the quality of the theater-level box office forecast and the duration of play did not affect this comparison, theater rankings were developed using theater-level revenue forecasts using our estimation procedure, while the optimal duration of play for a given movie at each theater was determined by the model in Somlo (2005). Comparing our method with the distributor's procedure, we found that theaters chosen by our method were 51% different on average than those selected by the distributor. In addition, had our method been implemented, this would have increased the average box office profit by $2,950 per theater, or by an average of $1.8 million per movie. This translates to a 12% increase in expected distributor's box office profit. In addition, absolute revenue and individual percentage improvements for some movies were as high as $5 million or 33%, respectively

It is important to note that these numbers underestimate true gains. In practice, the distributor's method would have performed worse than these results indicate without the advantage of the optimal duration of play for each movie determined using the model in Somlo (2005). We believe that our method outperforms the distributor's procedure because its allocation of theaters is based not just on a ranking of historical sales volumes across all movies or a ranking based on weighting of sales along with certain attributes, but matching all the key attributes of a given new movie with the characteristics of the theater under consideration. For instance, consider the movie "What Women Want". The distributor's approach chose theaters with historically high sales in areas where the ethnicity was predominantly white at each of the regions. In contrast, our approach chose smaller theaters in urban areas with a higher percentage of singles and with higher population densities. This lead to an average increase of box office profit by $1000 per theater or around $1.3 million in total possibly due to the appeal of this movie with singles who typically lived in more densely populated urban areas. As an additional example, consider the movie "Meet the Parents". The distributors again choose theaters with historically high sales in predominantly white neighborhoods. In addition to considering ethnicity, our approach allocated more theaters in the Midwest and the South Central region and in higher income neighborhoods. This led to an average increase in box office profits per theater of $1100 or $1.5 million across all theaters perhaps due to the movies appeal with conservative and affluent audiences. These examples provide further evidence that including theater level characteristics is crucial to effective distribution planning in the motion picture industry.

7. CONCLUSIONS

Our goal in this paper is to expose the reader to an intellectually engaging problem context laden with opportunities for research that can have a high impact on profits in the motion picture industry. The following conclusions can be drawn from this research.

- There are significant differences in revenue across the same movie in different theaters due to dissimilarity in location specific characteristics at the theater such as amenities and demographics (e.g., median income, age and population density). Therefore, it is important to develop detailed, disaggregate theater-level box office forecasts and to use these forecasts to determine the distribution plan that decides on which theaters to show a given movie.
- Forecasting theater-level, box office revenues is challenging as it requires an understanding of which movie attributes and theater characteristics will affect revenues, how they do so, and how this changes over time. In addition, one needs to also understand how the decision process of individual moviegoers to see a given movie affects theater-level forecasts and how this can be integrated into the forecasting process. The estimation procedure developed in this paper incorporates these aspects into determining detailed, theater-level revenue forecasts. This procedure reduces average forecast error by over 75% compared to benchmark models based on industry practice.
- Given theater-level revenue forecasts over time, the distributor faces the problem of determining at which theaters to show a given movie in order to optimize profits. This

problem is complicated because a minimum number of theaters has to be selected in each region and because the distributor needs to ensure that competing theaters are not selected. The DLSP provides an effective basis to approach this problem. We also found that this model was robust with variations in the theater-level revenue forecast and provides a basis to understanding the impact of changes in minimum play length on theater choice. In addition, the DLSP outperformed the method used by the distributors to select theaters and has the potential to increase average distributor profits by 12%, or around $1.8 million per movie.

This paper provides several avenues for future research. First, refinements could be developed to further improve the accuracy of the theater-level box office revenue forecasting procedure. Second, improvements could be made on the heuristics to increase the profits from the DLSP. Finally, the approach developed in this paper in which we determine the best locations to show a movie by estimating profit as a function of movie (or product) attributes and theater (or location) characteristics can be applied in a variety of service industry settings. For instance, one could use this idea to choose the best locations for concerts in the music industry, to determine the optimal location of specialty boutiques in the retail industry, and to pick the locations of resorts and restaurants in the hospitality industry. The modifications required to apply our model in these contexts could be a promising area for new research.

In conclusion, we believe that the methods presented in this paper provide a useful method to forecast theater-level box office revenues and use these forecasts to choose the best locations to screen the movie to optimize the distributor's profit in the motion picture industry.

ACKNOWLEDGEMENTS

We would like to thank Thomas McGrath, Executive Vice President, Viacom Entertainment Group; Don Tannenbaum, Executive Vice President, Distribution, Warner Brothers; Thomas Molter, Vice President, International Distribution, Warner Brothers; and Steve Rothenberg, President of Distribution, Artisan Entertainment who all provided valuable industry-specific information for this research. We would also like to thank Professors Anand Bodapati, Charles Corbett, Donald Morrison, Rakesh Sarin, and Charles Weinberg for several helpful comments. Financial assistance for this research was provided by the Center for International Business Education and Research and the Entertainment and Media Management Institute at UCLA. We would also like to thank Mr. Theodore Treantafelles at the UCLA Anderson School who helped us with various aspects of this work.

REFERENCES

Apte, U., Maglaras, C., and Pinedo, M., (2006), "Operations in the Service Industries: Introduction to the Special Issue" *Production and Operations Management*, 17(3): 235-237.

Araman, V.and Popescu, I. "Media Revenue Management with Audience Uncertainty: Balancing Upfront and Spot Market Sales", Forthcoming, M&SOM, 2009.

Bart, P. (2000). *The Gross: The Hits, The Flops – The Summer that Ate Hollywood*. St. Martin's Griffin. New York, NY.

Bates, Douglas M. and Watts, Donald G. (1988). *Nonlinear Regression Analysis and Its Applications*. John Wiley and Sons. New York, NY.

Bollapragada, S., M. Bussieck, S. Mallik. (2004). "Scheduling Commercials Videotapes in Broadcast Television". *Operations Research*, 52(5), 679-689.

Bollapragada, S., S. Mallik, "Managing On-Air Ad Inventory in Broadcast Television," IIE Transactions, 40(12), 1107-1123, 2008.

Brooke, A., Kendrick, D. and Meeraus, (1992). *GAMS: A User's Guide*. The Scientific Press. San Francisco, California.

Chisholm, D.C. and Norman, G. (2002). “Spatial Competition and Demand: An Application to Motion Pictures.” Working Paper. Suffolk University.

Dawande, M, Drobouchevitch, I, Rajapakshe, T. and Sriskandarajah, C. (2010). “Analysis of Revenue Maximization Under Two Movie-Screening Policies” *Production and Operations Management*, 19(1): 111-124.

De Vany, A. and Walls, W.D. (1996). “Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry.” *The Economic Journal*. 106: 1493–1514.

Einav, L. (2002). “Seasonality and Competition in Time: An Empirical Analysis of Release Date Decisions in the U.S. Motion Picture Industry.” Working Paper. Stanford University.

Elberse, A. and Eliashberg, J. (2003). “Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures.” *Marketing Science*. 22(3): 329–354.

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

Eliashberg, J., Elberse, A., and Leenders, M (2006), “The Motion Picture Industry: Critical Issues in Practice, Current Research and New Research Directions”, *Marketing Science,* 25(6) 638-661.

Eliasberg, J. et al. (2009). “Demand-driven scheduling of movies in a multiplex. Int. J. of Research in Marketing,” 26, 75-88.

Garey, M.R. and Johnson, D.S. (2000). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company. New York, NY.

Geoffrion, A.M. (1974). “Lagrangean Relaxation for Integer Programming.” *Mathematical Programming Study*. 2: 82–114.

Hayes, D. and Bing, J. (2004). *Open Wide: How Hollywood Box Office Became a National Obsession*. Miramax Books Hyperion. New York, NY.

Hirschman, E.C. and Holbrook, M.B. (1982). “Hedonic Consumption: Emerging Concepts, Methods and Propositions.” *Journal of Marketing*. 46: 92–101.

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

Lehmann, D.R. and Weinberg, C.B. (2000). “Sales Through Sequential Distribution Channels: An Application to Movies and Videos.” *Journal of Marketing*. 64(3): 18–33.

Litman, B.R. and Ahn, H. (1998). “Predicting Financial Success of Motion Pictures.” *The Motion Picture Mega-Industry*. B.R. Litman (editor). Allyn and Bacon. Needham Heights, MA.

McGrath, T. (2004). Private communications. Executive Vice President. Viacom Entertainment Group, New York, NY.

Molter, T. (2004). Private communications. Vice President, International Distribution, Warner Brothers, Burbank, CA.

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

Pinkse, J. and Slade, M. (1998). "Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models." *Journal of Econometrics*. 85: 125–154.

Reardon, B.D. (1992). "The Studio Distributor." *The Movie Business Book*. Second Edition. J.E. Squire (editor). Fireside. New York, NY.

Rothenberg, S. (2003). Private communications. President of Distribution, Artisan Entertainment, Santa Monica, CA.

SAS version 8.2 (2001). SAS Institute Inc. Cary, NC.

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

Scott, A.J. (2005). "On Hollywood, The Place, The Industry" Princeton University Press, Princeton, New Jersey.

Sochay, S. (1994). "Predicting the Performance of Motion Pictures." *Journal of Media Economics*. 7(4): 1–20.

Somlo, B. (2005) "Distribution Planning for the Motion Picture Industry: Models, Applications and Analysis." Unpublished Doctoral Dissertation. UCLA Anderson School.

Spohrer, J. and Maglio, P.P., (2006), "The Emergence of Service Science: Toward Systematic Service Innovations to Accelerate Co-Creation of Value" *Production and Operations Management*, 17(3): 238-246.

Swami, S., Eliashberg, J. and Weinberg, C.B. (1999). "SILVERSCREENER: A Modeling Approach to Movie Screen Management." *Marketing Science*. 3: 352–372.

Swami, S., Puterman, M.L. and Weinberg, C.B. (2001). "Play It Again, Sam? Optimal Replacement Policies for Motion Picture Retailing." *Manufacturing and Services Operations Management*. 3: 369–386.

Tannenbaum, D. (2001). Private communications. Executive Vice President, Distribution, Warner Brothers, Burbank, CA.

Thomas, B. (1998). "High Tech and Old Way Get Movies to Theater on Time – Usually." Associated Press. July 28.

Vany, A.D. and Eckert, R. (1991). “Motion Picture Antitrust: The Paramount Cases Revisited.” Research in Law and Economics, Vol. 14, 51-112.

Vogel, H.L. (2001). *Entertainment Industry Economics: A Guide for Financial Analysis*. Fifth Edition. Cambridge University. Cambridge, MA.

Wyatt, J. (1994). *High Concept: Movies and Marketing in Hollywood*. University of Texas Press. Austin, TX.

Zuckerman, E.W. and Kim, T.Y. (2003). “The Critical Trade-off: Identity Assignment and Box Office Success in the Feature Film Industry.” *Industrial and Corporate Change*. 12(1): 27–67.

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

Table 1. Examples Of Missed Box Office Forecasts

Title	NRG Estimate[(1)] ($M)	Actual Box Office ($M)	Relative Percentage Error [(2)]
X-Men	29.5	54.5	-45.9%
The Mummy Returns	50	70.1	-28.7%
Charlie's Angels	28	40.1	-30.2%
The Perfect Storm	21.5	41.3	-47.9%
Chicken Run	9	17.5	-48.6%
The Patriot	25	22.4	+11.6%
The Story of Us	18	9.7	+8.6%
Fight Club	14	11	+27.3%
Titan A. E.	12.5	9.4	+33%
Star Wars -Episode I.	150	105.7	+41.9%

(1) National Research Group's estimate of the opening weekend box office.
(2) Relative percentage error = (NRG Estimate – Actual BO) / Actual BO.

Source: IMDB and Variety.

Table 2. Box Office Revenues At Selected Markets

Markets*	The Matrix Reloaded	Star Wars: Episode II
Boston	$38,371	$35,829
Cleveland	$21,105	$21,559
Dallas	$31,207	$35,385
Denver	$31,423	$28,663
New York	$36,968	$53,499
Pittsburgh	$23,193	$18,683
San Francisco	$60,336	$60,169
St. Louis	$22,813	$23,410
Tampa	$25,043	$25,763

* Movies are shown in the same number of theaters in each market and revenues reported are Friday to Sunday averages over duration of play.

Data Source: Nielsen EDI's Box Office Sample.

Table 3. Results of the Nonlinear Regression for Selected Movie-Theater Pairs

Theater Type	Theater	Film	T (wks)	Actual BO ($1,000)	N	λ	γ	Type of Pattern	Predicted BO ($1,000)	Absolute	MSE
Multi	WYNN	LIESBTH	12	23.009	22.994	26.168	0.341	Exponential	22.607	1.75%	1.172
Multi	WYNN	PROOF	5	4.798	5.712	0.659	0.659	Erlang-2	4.802	0.01%	0.387
Multi	WYNN	WOMEN	11	32.248	32.318	2.946	0.335	Gen. Gamma	31.400	2.63%	1.773
Multi	FENW	VALEN	3	22.640	23.049	2.043	2.043	Erlang-2	22.691	0.23%	0.097
Multi	FENW	CHARLIE	9	255.337	254.914	7.011	0.573	Gen. Gamma	253.314	0.79%	6.193
Multi	FENW	CHICKEN	10	151.711	154.752	59.795	0.391	Exponential	151.615	0.06%	2.723
Multi	KTLA	CHICKEN	7	48.535	53.029	98.852	0.379	Exponential	49.283	1.54%	1.744
Multi	KTLA	WOMEN	7	59.411	64.666	3.559	0.369	Gen. Gamma	59.221	0.32%	0.988
Multi	KTLA	FOCKER	11	76.058	90.369	16.231	0.183	Exponential	78.126	2.72%	3.762
Mega	ALIS	CHICKEN	7	131.560	134.969	5.662	0.550	Gen. Gamma	131.783	0.17%	1.846
Mega	ALIS	WOMEN	8	171.214	175.080	5.031	0.501	Gen. Gamma	171.535	0.19%	3.506
Mega	ALIS	LIESBTH	9	146.500	146.282	45.437	0.487	Exponential	144.430	1.41%	4.836
Mega	HADL	CHARLIE	8	50.418	51.558	11.083	0.565	Exponential	50.965	1.09%	1.903
Mega	HADL	CHICKEN	6	37.547	41.732	10.523	0.385	Exponential	37.429	0.07%	0.522
Mega	HADL	WOMEN	11	56.044	56.213	6.311	0.364	Gen. Gamma	55.126	1.64%	1.980
Mini	FAUL	HEADOV	3	1.626	1.911	1.195	1.195	Erlang-2	1.668	2.58%	0.122
Mini	FAUL	SHAFT	5	6.581	7.352	6.509	0.476	Gen. Gamma	6.616	0.53%	0.086
Mini	FAUL	DISKID	4	10.375	13.230	4.536	0.412	Gen. Gamma	10.434	0.56%	0.189
Mini	HANF	CHARLIE	10	26.963	27.751	89.985	0.399	Exponential	27.235	1.01%	0.779
Mini	HANF	SPYKIDS	9	24.654	25.281	3.691	0.399	Gen. Gamma	24.497	0.64%	0.889
Mini	HANF	FOCKER	11	30.190	37.810	17.694	0.151	Exponential	30.537	1.15%	1.166
Single	GRAH	NUTTY	4	23.089	24.612	10.631	0.715	Exponential	23.100	0.05%	0.027
Single	GRAH	DRSEUSS	5	41.800	44.285	14.218	0.518	Exponential	40.841	2.30%	2.170
Single	CMEO	MISSCON	7	36.838	38.496	36.113	0.406	Exponential	36.231	1.65%	1.154
									Average Errors	**1.05%**	**1.67%**

Table 4. Multiple Regression Results Relating Model Parameters with Theater Characteristics and Movie Attributes

	Dependent Variables (Standard Errors In Parentheses)			
	Multiple Regression Model			Benchmark
	Log N	Log λ	Log γ	Log Actual
Intercept	3.584* (0.281)	2.374* (0.397)	-0.831* (0.193)	3.000* (0.200)
Theater Related Variables				
Rockies Region	-0.282** (0.144)	0.010 (0.203)	0.161 (0.099)	
North Central Region	-0.584* (0.119)	-0.091 (0.168)	0.064 (0.082)	
South Central Region	-0.231** (0.113)	0.156 (0.160)	0.108 (0.078)	
Midwest Region	-0.173 (0.147)	0.127 (0.207)	0.137 (0.100)	
North Eastern Region	-0.440* (0.088)	-0.025 (0.124)	0.092 (0.060)	
South Eastern Region	-0.291* (0.105)	0.059 (0.147)	0.062 (0.72)	
Median Age	-0.015*** (0.008)	-0.007 (0.011)	0.008 (0.005)	
Percentage Of Singles	0.009 (0.005)	0.005 (0.008)	0.004 (0.004)	
Population Density (1,000s)	0.012* (0.004)	-0.011** (0.005)	-0.003 (0.003)	
Median Household Income ($1,000s)	0.001 (0.002)	-0.001 (0.003)	0.001 (0.001)	
Adult Ticket Price ($)	0.269* (0.057)	0.169** (0.080)	0.043 (0.039)	
Ticket Discount ($)	-0.143*** (0.085)	0.019 (0.120)	0.022 (0.058)	
Stadium Seating	0.364* (0.065)	0.032 (0.092)	0.010 (0.045)	
Mini Type (2-7)	-0.172 (0.265)	-0.780** (0.375)	-0.067 (0.182)	
Multi Type (8-15)	-0.148 (0.264)	-0.799** (0.373)	0.086 (0.181)	
Mega Type (16+)	-0.071 (0.275)	-0.737 (0.389)	0.192 (0.189)	
Numbers Of Neighboring Theaters (5 Miles)	0.015** (0.007)	-0.003 (0.009)	0.012** (0.005)	
Numbers Of Neighboring Theaters (10 Miles)	-0.002 (0.009)	-0.018 (0.013)	-0.008 (0.006)	
Numbers Of Neighboring Theaters (15 Miles)	-0.004 (0.004)	0.013** (0.005)	0.001 (0.003)	
Movie Related Variables				
Runtime (Min)	0.000 (0.002)	0.004 (0.003)	-0.002 (0.002)	-0.002 (0.003)
Production Budget ($M)	0.008* (0.001)	-0.001 (0.002)	-0.001 (0.001)	0.004** (0.002)
MPAA Rating	0.053 (0.050)	-0.136*** (0.07)	-0.087* (0.034)	0.049 (0.065)

	Dependent Variables (Standard Errors In Parentheses)			
	Multiple Regression Model			Benchmark
	Log N	**Log λ**	**Log γ**	**Log Actual**
Critics' Rating	0.305* (0.033)	0.031 (0.046)	0.148* (0.022)	0.330* (0.044)
Special Effects	0.063 (0.086)	0.276** (0.122)	0.046 (0.060)	0.119 (0.113)
Star Presence	0.042 (0.077)	0.211*** (0.109)	0.003 (0.053)	0.030 (0.101)
Sequel	0.192 (0.171)	-0.120 (0.242)	0.154 (0.118)	0.165 (0.224)
Animation Genre	-0.325*** (0.173)	0.555** (0.244)	0.397* (0.119)	-0.328 (0.223)
Comedy Genre	(0.097 (0.082)	0.215 (0.115)	-0.063 (0.056)	0.157 (0.106)
Drama Genre	-0.113 (0.108)	0.176 (0.156)	0.054 (0.075)	-0.051 (0.142)
Horror Genre	0.024 (0.127)	0.280 (0.180)	0.063 (0.088)	0.072 (0.181)
Fantasy Genre	-0.360* (0.128)	0.374** (0.180)	0.263* (0.088)	-0.227 (0.168)
Holiday Period Opening	0.050 (0.079)	-0.499* (0.111)	-0.067 (0.054)	0.167 (0.103)
Winter Period Opening	0.051 (0.095)	-0.220 (0.135)	-0.198* (0.065)	-0.046 (0.125)
Spring Period Opening	-0.229** (0.098)	-0.231 (0.139)	-0.050 (0.066)	-0.253*** (0.133)
Fall Period Opening	-0.094 (0.094)	-0.145 (0.133)	-0.190* (0.065)	0.015 (0.123)
Interaction Terms				
Median Age * MPPA Rating	-0.003 (0.006)	-0.002 (0.008)	-0.006 (0.004)	
% Of Singles * MPAA Rating	-0.006 0.004)	-0.003 (0.005)	-0.003 (0.002)	
Median Household * Rating	0.001 (0.001)	0.001 (0.002)	-0.001 (0.001)	
Population * Rating	-0.002 0.002)	0.005 (0.004)	0.003 (0.002)	
Star Presence * Median Age	0.013 (0.010)	-0.002 (0.015)	-0.001 (0.007)	
Star Presence * % Of Singles	-0.007 0.005)	0.003 (0.008)	-0.003 (0.004)	
Model R^2	0.5337	0.1510	0.2719	0.2004
Adjusted R^2	0.4998	0.0893	0.2190	0.1774
F Value	15.74	2.45	5.14	8.70
Pr > F	<0.0001	<0.0001	<0.0001	<0.0001
N	605	605	605	607

* Statistically significant at 1%. ** Statistically significant at 5%. *** Statistically significant at 10%.

Table 5: Actual Vs Forecast Revenues Across All Theaters for Forecasting Method and Benchmark Model for Selected Movies

Title	Actual Box Office Revenues ($M)	Revenues Estimate of Method (($M)	Relative Percentage Forecast Error(3)	Revenues Estimate of Benchmark Model (($M)	Relative Percentage Forecast Error (4)
Shaft	70.3	61.9	-12	34.4	-51
Disney' The Kid	69.7	77.4	11	36.2	-48
What Lies Beneath	155.4	183.4	18	265.7	71
The Nutty Professor II	123.3	106	-14	61.7	-50
Meet the Parents	166.2	191.1	15	275.9	66
What Women Want	182.8	199.3	9	283.3	55
Miss Congeniality	106.8	97.2	-9	71.6	-33
The Wedding Planner	60.4	69.5	15	105.7	75
Enemy at the gates	51.4	59.6	16	92.5	80
Spy Kids	112.7	98	-13	67.6	-40

(3) Relative percentage forecast error = (Method Estimate – Actual BO) / Actual BO.
(4) Relative percentage forecast error = (Benchmark Model Estimate – Actual BO) / Actual BO.

Table 6. Average Percentage Gaps from Tightest Bound across 150 Problems for the DLSP

Average (Minimum/ Maximum)	Problem Size (Theaters/ Movies/ Competition Density)		
	3000/ 150/ 6	3000/ 150/ 11	3000/ 150/ 15
Optimal Solution: Gams	*	N/A	N/A
Upper Bound: Linear Programming Relaxation	5.24%	*	*
Lower Bounds:			
Myopic Heuristic	1.8%	8.2%	10.3%
Greedy Heuristic	1.3%	8.7%	11.2%

Table 7 Average Percentage Gaps From Upper Bound with Scaled π_{Ijt}.

Scale Factor	-30%	-20%	-10%	0%	10%	20%	30%
Myopic Heuristic	10.95	10.6	10.5	10.3	10.7	10.8	10.9
Greedy Heuristic	12.6	12.2	11.7	11.2	11.7	12.3	12.7

Figure 1. The Motion Picture Industry Supply Chain

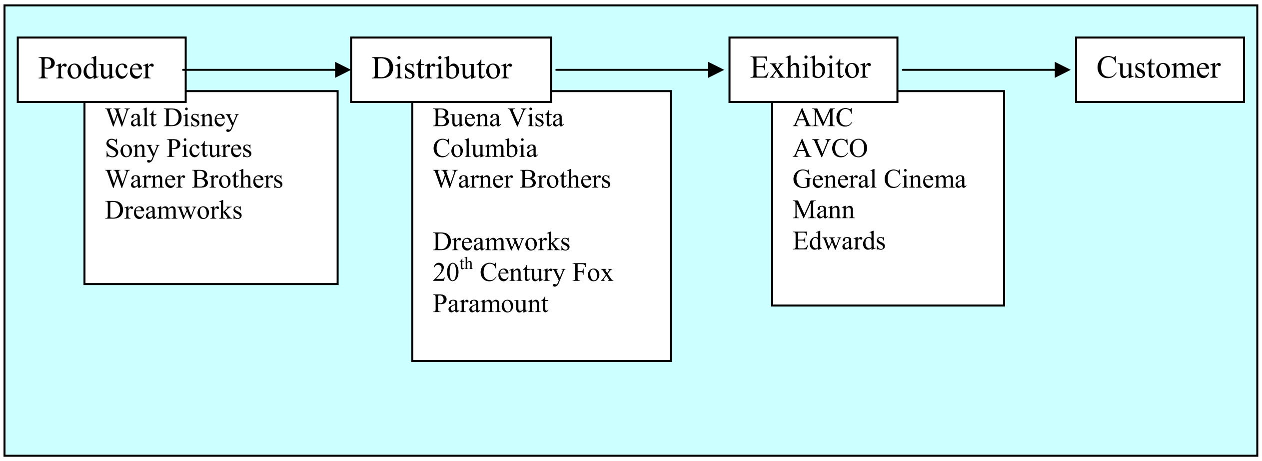

Figure 2. Methodology For Estimation Of π_{ijt}: Theater-level Box Office Revenues For A New Movie