Modeling Movie Release Strategies

Abstract

This research examines the impact of release strategies on the diffusion of motion-picture movies at the US domestic box-office. A model is developed that captures consumer choice as a behavioral process accounting for the movie's intrinsic attributes, seasonality, word-of-mouth, network effects, consumer heterogeneity, marketplace competition, and managerial inputs. The model estimates weekly box-office receipts for 137 movies and achieves a median r-squared of 0.98 and fits 91 percent of the movies with an r-squared greater than 0.75. The study demonstrates that accounting for this full range of factors not only improves the model's fit, but also leads to a parameter set that depicts a richer description of the movie industry. Managerial decisions regarding the selection of a movie's release date and its distribution strategy are found to significantly impact box-office performance.

Key Words

Motion pictures distribution strategy, box-office revenue forecasting, marketing support models, decision support models, prerelease market evaluation, new product diffusion

Christopher Ryan Hughes
San Francisco, United States of America
PricewaterhouseCoopers
June 27, 2012

1. Introduction

The US domestic box office is the motion picture industry's Constantinople: a gateway through which all new movies must first pass before they can be released in new markets. The box office's position at the forefront of the product lifecycle means that not only is it the first opportunity for distributors to recoup their massive upfront investments, but also that its performance will set the movie's baseline value for subsequently products released later in the movie's lifecycle (Ainslie 2005). Its strategic importance stimulates fierce competition, and the stakes are high. In 2009, the average wide-releasing movie cost $90.1 million to produce ($57.7 million) and market ($32.4 million). A staggering figure considering that only 30-40 percent of movies successfully recoup expenses (Vogel 1994), and distributors only release a small portfolio of 5 to 20 films per year. Even among successful films, returns are strikingly unequal, and the top 20 percent of movies earn 65 percent of total box office receipts (Delre 2007). Given its economic, cultural, and strategic importance, it is surprising that only a handful of papers have been published on issues relevant to the design and implementation of decision models in the motion picture industry.

Expanding upon work by Silk and Urban (1978) in the consumer goods industry, Eliashberg et al. (2000) designed a "macro-flow" model that forecasts weekly box-office revenue by progressing consumers through a series of behavioral states as they are exposed to word-of-mouth and marketing information. Their model is calibrated from "consumer clinics" and has been successfully deployed by movie distributors in the Netherlands. Liu (2006) developed a second pre-release model that predicts box office revenues as a function of early-audience word-of-mouth. He finds that the inclusion of word-of-mouth significantly reduces forecasting errors for both weekly and cumulative box office revenue predictions.

Ainslie et al. (2005) examines weekly box-office revenue as a market-share problem, and develops a "sliding-window" logit model with a gamma diffusion pattern to capture the buildup/decay of a movie's potential audience over time. They find that incorporating competition improves the model predictive ability, and that an increase in competition in any week causes a movie's box-office revenue to decrease, although some of the lost sales may be recaptured in subsequent weeks.

Einav (2007) explores the effect of seasonality on box-office revenue by using a market share model that separate the effects of seasonality, competition, and movie quality on consumer demand. He finds that when quality is accounted for, the underlying seasonality in the movie market is amplified by suboptimal distribution strategies, and concludes that total industry revenues would increase if the highest quality movies were released more uniformly over the course of the year as opposed to being clustered around a few major release dates.

This research adds to the extant set of models and describes a pre-release model that distributors can use to select robust movie release strategies for their movies. Specifically, the study will seek to answer the following research questions:

1. *What is the optimal release date for a movie given marketplace competition?*

2. What distribution strategy will maximize a movie's box-office revenue given its intrinsic attributes and the behavioral characteristics of consumers?

The rest of this article is organized as follows: section 2 describes the model, section 3 describes the dataset and the empirical results of the study, section 4 illustrates how the model can be used to guide distribution strategy, and section 5 summarizes the research insights and suggests directions for future research.

2. Model Description

The model is described in three parts: first, the stock-flow consistent *Consumer Adoption Structure* is outlined; second, the mathematical underpinnings of the *Choice Probability Equation* are discussed; and third, the impact and underlying assumptions of the model's *Managerial Inputs* are explained.

2.1. Consumer Adoption Structure

Consumer Behavioral State Representation

Following Mahajan et al. (1984) and Eliashberg et al. (2000) the model partitions moviegoers into six mutually exclusive behavioral states (Figure 1).

1. **Undecided:** consumers who are unaware the movie is playing in theaters, and/or are undecided about viewing it in theaters
2. **Intender:** consumers who are aware the movie is playing and intend to watch it in theaters, but have not yet acted upon their intention
3. **Rejecter:** consumers who are aware the movie is playing but have decided not to watch the movie in theaters
4. **Positive Spreader:** consumers who have seen the movie, enjoyed the movie, and are actively spreading positive word-of-mouth
5. **Negative Spreader:** consumers who have seen the movie, did not enjoy the movie, and are actively spreading negative word-of-mouth
6. **Inactive:** consumers who have seen the movie but are no longer actively spreading word-of-mouth

Alternative behavioral state representations were tested with more/less behavioral states, but none were found to significantly improve the model fit. The six-state behavioral representation was therefore adopted stay consistent with the previous literature, to simplify the interpretation of model outputs, and to increase computational efficiency during the calibration process. For simplicity, all consumers are assumed to start in the Undecided state.
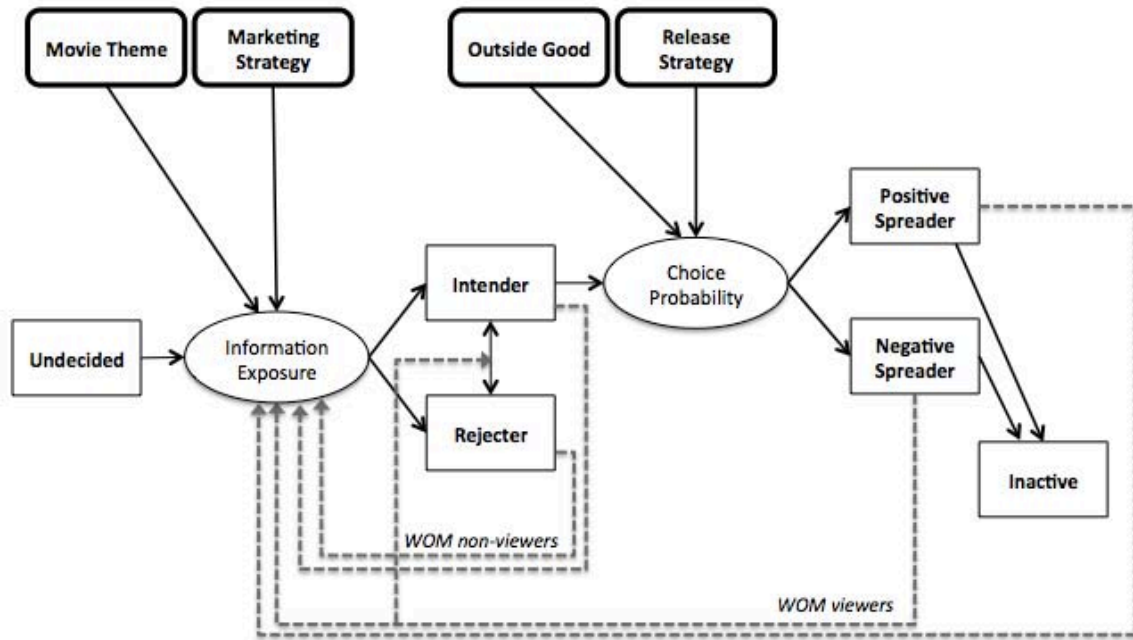
Description of State Transitions

State transitions are stock-flow consistent and ensure the population of consumers is constant throughout the simulation (Forrester 1961). These equations are common in the diffusion literature, and provide constant denominators for calculating the impact of word-of-mouth, network effects, and market saturation as a function of the number of consumers in each behavioral state (Urban 1970, 1990). To illustrate the mathematical formulations of the state transition equations, the transition consumers make between the Undecided and Intender states is described in detail below.

$$f_t = \frac{U_t}{N} * (MKTG_t + WOM_t) * MSA_t$$

Where:
f_t = The flow rate at time t
U_t/N = The fraction of consumers in the undecided state at time t
$MKTG_t$ = The number of consumers in response to marketing activities at time t
WOM_t = The number of consumers flowing in response to word-of-mouth at time t
MSA_t = The effect of consumer heterogeneity on market saturation at time t, $MSA_t \geq 0$

The population fraction term, U_t/N, decreases the probability consumers in the Undecided state will receive marketing and word-of-mouth messages in proportion to the fraction of consumers currently in the Undecided state. Thus, as the market saturates and the number of consumers in the Undecided state approaches zero, the flow rate will also approach zero. The consumer heterogeneity term, MSA_t, gives the location of the

4

inflexion point of the adoption curve flexibility, allowing the market to saturate before every consumer has adopted (Mahajan et al., 1990). Distinguishing between the effect of marketing and the effect of word-of-mouth is standard practice in marketing diffusion models (Bass 1969), and both terms have both been empirically proven to be important to the adoption of motion pictures. The model assumes independence between the exposures to different types of information sources (Urban et al. 1990), and ignores second-order effects such as the simultaneous exposure to advertising and word-of-mouth (Elaishberg et al. 2000).

The probability that marketing messages will trigger consumers to transition between the Undecided and Intender states is a function of the marketing reach at time t, and the probability the movie's theme will be acceptable to the viewer.

$$MKTG_t = SPEND_t * kTHEME$$

Where:
$SPEND_t$ = The marketing reach associated with the advertising expenditures at time t
$kTHEME$ = The movie's theme acceptability, and $0 \leq kTHEME \leq 1$

The probability of exposure to positive/negative word-of-mouth depends on the number of Positive and Negative Spreaders in the population (Eliashberg et al., 2000), the number of the Intender and Rejecter in the population, the frequency of interaction between consumers in each state, the magnitude of the interaction frequency (Mahajan et al., 1990), and the probability that consumers in the sending state will influence consumers in the receiving state. The impact of word-of-mouth messages sent from Positive Spreaders on Undecided consumers is used illustrates the general framework of the word-of-mouth transition below. The impact of word-of-mouth from consumers in other states follows the same principles.

$$WOM_t = PW_t * cIMP * kMSG * MAG_t$$

Where:
PW_t = The number of positive word-of-mouth spreaders at time t
$cIMP$ = Coefficient for the probability that each message will influence its receiver, where $0 \leq cIMP \leq 1$
$kMSG$ = Number of messages sent by spreaders in each time period, where $0 \leq kMAG$
MAG_t = Multiplier that varies the number of messages sent proportional to the movies perceived availability at the time period t, and $0 \leq MAG_t$

The transition between the Undecided and Rejecter states is the inverse of the transition between the Undecided and Intender states; consumers will transition to the Rejecter state if they are exposed to marketing messages but do not find the movie's theme to be acceptable, or if they receive negative word-of-mouth.

The transitions between the Intender and Rejecter states are formulated similarly except the transition is only influenced by word-of-mouth and not by marketing. The impact of marketing is ignored because the model assumes the theme of marketing is constant over

the weeks around the movie's release and once a consumer has accepted/rejected a movie's theme, new marketing messages will not change their opinion.

Consumers within the population of Intenders may view the movie in any time period, including the one in which they first become aware. The decision to view is depicted by the choice probability. Upon seeing the movie, intenders transition to the Positive or Negative Spreader state based on their satisfaction with their viewing experience. Viewers who have satisfactory experiences are assumed to transition into the Positive Spreader state. Viewers with unsatisfactory experiences are assumed to transition into the Negative Spreader state. Positive and Negative Spreaders are assumed to actively spread word-of-mouth messages for a finite period before they transition into the Inactive state.

Consumers in the Intenders or Rejecters states are allowed to flow back to the Undecided state if their purchase is delayed: namely, if the movie is not playing in a nearby theater during the period when they intend to view the movie, or if they found it less attractive than the outside good for multiple periods.

Model Parameters
The state transitions are influenced by the following sets of parameters.
- Movie attributes
- Marketing strategy
- Distribution strategy
- Adoption structure

Movie attribute are unique for each movie and are outside the control of distributors. The attributes include *theme acceptability*, the likelihood Undecided consumers will transition to the Intender or Rejecter state; *viewer satisfaction*, the likelihood viewers will transition to the Positive or Negative Spreader state; and *perceived movie quality*, a variable that enters into the choice probability equation. Note that price is not included because box office admission tickets tend to be relatively constant across movies. Marketing strategy parameters influence the rate that Undecided consumers become aware of the film. Distribution strategy parameters include the movie's *release date*, and the *initial number of theaters* the movie opens on. Adoption structure parameters impact the average time for consumers to forget that a movie is playing, and the average time for Positive and Negative Spreaders to actively spread word-of-mouth messages.

2.2. Choice Probability Equation
The Intender's decision to view is a function of three elements: expected utility, market saturation, and product availability. The equation takes a multiplicative form to ensure that if any element of the equation is zero, the choice probability will also fall to zero. Thus, if the movie offers the marginal consumer zero utility relative to the outside good, or if the market is completely saturated and there is no marginal consumer to adopt, or if the movie is no longer available in the marketplace, then the choice probability will fall to zero and no new Intenders will view the movie.

$$CP_t = UP_t * MS_t * PA_t$$

Where:

CP_t = Choice probability at time t, and $CP_t \leq 1$

UP_t = Expected Utility probability at time t

MS_t = Market saturation at time t

PA_t = Product availability at time t

Expected Utility Probability

The probability of viewing increases as the consumer's expected utility increases relative to the outside good. The model captures the influence of expected utility on choice probability with a logit-choice equation (Ainsle et al. 2005, Einav 2007). Specifically, the model uses a binary logit equation that examines the consumer's choice to *view* or *not view* each movie without attempting to specify an exhaustive choice set.[1] The binary approach is possible because competition enters the model an exogenous input, and preferable because it is more computationally efficient than a multinomial formulation. The elements of the binary logit equation are displayed below.

$$UP_t = \frac{U_t}{OG_t + U_t}$$

Where:

U_t = The perceived utility at time t

OG_t = Outside good at time t

The perceived utility parameter is modeled as an exponent of the movie's *perceived movie quality* (Krider & Weinberg 1998), and a series of network effects used to estimate the level of buzz surrounding the movie (Delre 2007). Utility is adjusted by a scaling constant, and errors related to the unexplained portion of the utility function, ε_t, are assumed to be distributed iid extreme value.

$$U_t = e^{PMQ_t + NW_t + kU + \varepsilon_t}$$

Where:

PMQ_t = Perceived movie quality at time t

NW_t = Network effects at time t

kU_t = Utility constant at time t

ε_t = Unobserved error at time t

Krider and Weinberg (1998) describe a movie's drawing power on two dimensions: *marketability*, the level of interest the movie generates before it is released (based on factors such as directors, story line, and special effects); and *playability*, the level of interest the movie generates after it has been release and more signals about the movie's quality become available to the public. The model captures changes in the movie's attractiveness over time, or its *perceived movie quality*, with the following equation.

---

[1] Train (2009) discusses the many forms of the logit equation and under what assumptions they can be deployed

$$PMQ_t = PMQ_0 + \int_0^t \frac{(PMQ_F - PMQ_{t-1})}{tPMQ}$$

Where:

$PMQ_0 =$ The movie's *marketability*, or its perceived movie quality before release
$PMQ_F =$ The movie's *playability*, or its perceived movie quality after release
$tPMQ =$ Time for perceived movie quality to adjust from initial value to its final value

In a study of diffusion within social networks that spanned the movie industry, Delre (2007) reached the conclusion that "the success of movies depends more on the buzz surrounding the movie than on the quality of the movie itself." To capture the effect of buzz on a movie's attractiveness, the model deploys a set of network effects to account for the impact of the population of Intenders, Spreaders, and Inactives on potential viewers.

$$NW_t = NWI_t + NWW_t + NWV_t$$

$$NWI_t = I_t * cNWI$$

$$NWW_t = W_t * cNWW$$

$$NWV_t = V_t * cNWV$$

Where:

$I_t =$ Fraction of the population in the Intender state at time *t*
$W_t =$ Fraction of the population in the Positive/Negative WOM Spreader states at time *t*
$V_t =$ Fraction of the population in the Positive/Negative Inactive states at time *t*
$cNWI =$ Coefficient for the impact of the number of intenders on the utility to the marginal viewer
$cNWW =$ Coefficient for the impact of the number of current WOM spreaders on the utility to the marginal viewer
$cNWV =$ Coefficient for the impact of the number of previous viewers now inactive on the utility to the marginal viewer

Variations in the outside good enter the denominator of the expected utility equation via the term, OG_t, and fluctuate with changes in calendar events (Einav 2007) and competition (Ainslie et al. 2005). The outside good is equivalent to the opportunity cost of viewership, and as the outside good increases the probability a consumer will view decreases. Thus, during the holidays when consumers have more free time, the outside good decreases, and the probability of viewership increases. Similarly, when movies face little competition in the marketplace and consumers have fewer attractive alternatives, the outside good decreases, and the probability consumers will view increases. The contrapositives are also true.

$$OG_t = CAL_t * COMP_t$$

$$CAL_t = HOL_t * cHOL * SEA_t * cSEA$$

$$COMP_t = 1 + MPAA_t * cMPAA + GENRE_t * cGENRE$$

Where:
OG_t = The outside good at time t
CAL_t = Impact of calendar events at time t
$COMP_t$ = Impact of competition at time t
HOL_t = Impact of holidays on calendar events at time t
SEA_t = Impact of seasonality on calendar events at time t
$cHOL$ = Coefficient for the impact of holidays on calendar events
$cSEA$ = Coefficient for the impact of seasonality on calendar events
$MPAA_t$ = Relative number of competitive films with the same MPAA ratings at time t
$cMPAA$ = Coefficient for the impact of the number of films with the same MPAA rating
$GENRE_t$ = Relative number of competitive films with the same genre ratings at time t
$cGENRE$ = Coefficient for the impact of the number of films with a the same genre

While movie are individually unique, they all compete for moviegoers in a common marketplace. To capture the impact of competition on choice, the model follows Ainsle et al. (2005) and assumes movies with the same MPAA rating and/or the same genre are direct substitutes and therefore in competition for viewers. The variables $MPAA_t$ and $GENRE_t$ track the relative number of theaters that competitive films are playing on at any given point in time. Theater count is selected over alternative metrics as a competitive proxy for: consistency (the same units are used for the "product availability" variable); convenience (datasets are complete, publically available, and relatively accurate); and because they correspond with a movie's drawing power at each stage of its lifecycle (e.g., Avatar was less of a competitive threat when it played on 47 theaters 30 weeks after its initial release, then when it first opened on 3,452 theaters). To standardize the two variables, the number of theaters is interpreted relative to its historical mean and standard deviation.

$$MPAA_t = \frac{TMPAA_t - kTMPAA}{sdMPAA}$$

$$GENRE_t = \frac{TGENRE_t - kTGENRE}{sdGENRE}$$

Where:
$TMPAA_t$ = The number of theaters that competitive movies with the same MPAA rating were playing on at time t
$kTMPAA$ = Constant representing the average number of theaters that competitive movies with the same MPAA rating were playing on in any given week in the dataset
$sdTMPAA$ = The standard deviation for the number of theaters that movies with the same MPAA rating were playing on in any given week in the dataset
$TGENRE_t$ = The number of theaters that competitive movies with the same genre were playing on at time t
$kTGENRE$ = Constant representing the average number of theaters that competitive movies with the same genre were playing on in any given week in the dataset
$sdGENRE$ = The standard deviation for the number of theaters that movies with the same genre were playing on in any given week in the dataset

Market Saturation & Consumer Heterogeneity

Hanssens & Parsons (1993) note: "many commonly used sales-response functions work well despite not formally modeling saturation, even thought it must exist, because most firms do not operate near saturation levels." However, the observation that weekly revenue typically declines exponentially over time despite increases in surveyed *top-of-mind awareness* and *intention-to-view* (see Figure 2) suggests that the marginal consumer is increasingly difficult to convert. Thus, the model assumes that market saturation, defined here as the variation in the consumer's willingness to pay across the market, influence adoption.

The consumer heterogeneity hypothesis segments of the market and allows each segment of consumers to saturate at different rates (Mahajan et al., 1990). Heterogeneity can enter the choice function via a von Neumann-Morgenstern risk-aversion framework (Chatterjee & Eliashberg 1990), as a variable on the individual consumer's utility function (Horsky et al. 2006), as the distribution among the vector representing the individual consumer's preferences (Liechty et al. 2005), or by explicitly representing different groups of consumers (Rahmandad & Sterman 2008). However, because no publically available data exists to estimate consumer preferences and risk-tolerances, the model aggregates the effect of heterogeneity by directly decreasing the choice probability of the marginal adopter via the term MS_t (Hanssens & Parsons 1993).

$$MS_t = \frac{kMS}{1 + e^{V_t * cSAT}}$$

Where:

$kMS =$ Constant setting the y-axis intercept for the affect of market saturation when the number of viewer is equal to 0. Where $kMS \geq 2$

$cSAT =$ Coefficient for the impact of the number of films with a similar genre

Note that when the constant kMS is set to 2 and the number of viewers, V_t, is equals to zero, MS_t will be equal to one, implying that the market saturation does not influence the first consumer to adopt. Under this assumption the equation can be simplified to a single parameter, however, the two-parameter formulation gives the calibration process the freedom to scale kMS greater than 2 to explore scenarios where movies may have natural audience bases.

Theater Capacity and Product Availability

Bruno and Vilcassim (2008) observed that "not accounting for availability penalizes products with low distribution coverage" and empirically demonstrate that accounting for product availability yields more realistic coefficient estimates when logit models are used to recover parameter values. The model uses theater count as a proxy for product availability, and captures the effect of availability on choice with a decreasing returns-to-scale equation. First the current number of theaters the movie is playing on is divided by a reference value (the median number of theaters wide-releasing movies open on), and then raising the resulting number by an exponent less than one. The resulting dimensionless variable, PA_t, enters the choice probability equation (Sterman et al. 2007).

$$PA_t = \left(\frac{T_t}{kT}\right)^{cT}$$

Where:

T_t = Theaters the movies is playing on at time t

kT = Constant for the number of theaters a widely available movie plays on

cT = Coefficient transforming the impact of the relative number of theaters into a decreasing returns to scale function, where cT ≤ 1

2.3. Managerial Influences

Movie release strategies have a significant impact on ultimate box office revenues, and the process of selecting strategies mimics a high-stakes multi-player game (Krider & Weinberg, 1998). Distributors have three potential intervention points: the selection of the movie's release date, the selection of the movie's distribution strategy, and the selection of the movie's marketing strategy.

Krider and Weinberg (1998) developed a game-theoretic model to investigating marketplace competition between any two movies in direct competition, and empirically demonstrated that not only do studios recognize the impact of competition, but they also shift release dates in an optimal manner: simultaneously releasing strong movies to compete head-to-head during peak weeks, and delaying the release of weak movies when they face strong competition. In a pair of related papers, Einav (2007, 2010) explores the same issue from a multiple-movie macro-industry perspective, and empirically determines movie distributors overcompensate for the effects seasonality and cluster too many good movies around too few major release dates. He concludes that total industry revenue could be increased if distributors spaced out their best movies over the course of the year to avoid competition.

To capture the impact of managerial decisions on box-office performance, the model adopts Einav's (2007) taxonomy and recognizes three distinct theatric release strategies.

1. Wide-release strategy: screening begins in a large number of theaters and is supported by an extensive national advertising campaign. Roughly 84 percent of all movies that reach national distribution follow wide-release strategies.
2. Platform-release strategy: screening begins in a small number of theaters, often in big cities, with advertising concentrated in local newspapers. In most cases the movie expands to additional screens in more rural areas within two to four weeks of the initial screening. Distributors use platform releases for movies do not have obvious appeal to mainstream audiences (e.g., the movie's actors are unknown, the subject matter is controversial) to stimulate word-of-mouth with the hope of mitigating audience uncertainty regarding the movie's theme or quality. Roughly 14 percent of all movies that reach national distribution follow platform-release strategies.
3. Limited-release strategy: screening begins in a small number of theaters without expectations of a subsequently wider release. In a few rare cases, the movie will perform exceptionally well and distributors will expand distribution. Less than 2 percent of movies that reach national distribution originate from a limited-release strategy.

Defining the theatric release strategies is important because distributors vary the magnitude and timing of their advertising for each type of release, and advertising has been proven to impact box-office revenue. Zufryden (1996) studied the impact of marketing on movie revenues using an awareness/intention behavioral model, and demonstrated that both the total marketing spend and the timing of the spending was critical to a movie's success. Ishii et al. (2012) used a bass-diffusion model that incorporated daily advertising spending to reproduce daily box-office revenue and blog-postings (a proxy for audience awareness) in the Japanese market.

To determine the probable release strategies deployed for past movies, the model uses rule-based logic to assign the most appropriate marketing strategy for each movie in the dataset based on its historic theater count. The three strategies correspond to the three release strategies identified by Einav (2007).

1. Wide-release marketing strategy: marketing is concentrated on the weeks surrounding movies release. For simplicity, the movie's national marketing campaign is assumed to begin five weeks before the movie opens.[2]
2. Platform-release marketing strategy: marketing spending is distributed in the same fashion as the wide-release marketing strategy, except that it is shifted from the movie's official release date (in a small number of theaters) to its first week of wide-release. A small fraction of the movie's marketing budget is allocated to each displaced marketing week before the movie's official release.
3. Limited-release marketing strategy: the movie's marketing spending is focuses on the initial opening and each significant expansion in theater count.

3. Empirical Analysis

This section describes the empirical analysis in three parts: first, the dataset used to estimate the model is described; second, the estimation methodology is outlined; and third, the estimation results are interpreted.

3.1. The Dataset

A dataset was constructed of the 155 wide-releasing movies of 2009. For each movie, weekly box-office receipts and theater counts were collected along with a set of descriptive variables.[3] However, the initial calibration run revealed an inconsistency in the dataset for 18 movies with non-Friday release dates. Because most movies are released on Friday many data aggregators restart their weekly box-office counts each new Friday, however this habit causes the opening-week revenues of the non-Friday releasing movies to be wrongly spread over two weeks in the dataset, which subsequently led to misspecified estimates. To clarify the interpretation of the models estimates, these movies were removed from the dataset, leaving 137 movies for analysis. Summary statistics for remaining movies are displayed in Table 1.

---

[2] In reality, marketing can begin up to a year before the movie is released, but it is rare for a movie to make a significant national push earlier than five weeks before a movie is released

[3] The dataset was constructed with data from The-Numbers, Rotten Tomatoes, Box-Office Mojo, and IMDB.

Table 1: Summary Statistics for Dataset

|	Total Revenue (2009 $)	Population Share (%)*	First Week Revenue (2009 $)	Production Cost (2009 $)	Advertising Expenditures (2009 $)
Titles	137	137	137	100	137
Mean	72,630,441	3.15	30,249,889	60,121,150	33,115,707
Median	42,100,625	1.83	18,082,380	40,000,000	31,889,147
Standard Deviation	93,785,853	4.07	38,353,422	57,387,566	14,241,656

*Population share (box-office revenue divided by ticket price and by U.S. Population)
Average ticket price 2009: $7.50
US population 2009: 307,000,000

To supplement the box-office figures, a second dataset was constructed with time-series data on audience awareness, intention-to-view, and marketing exposures for 20 movies. While the dataset was too small to be used directly in the calibration process, it did provide three important insights on audience behavior that could be used to interpret the validity of the calibration results. First, both awareness and intention-to-view peaked after the movie's initial release, often almost doubling between the first and second week. Second, both awareness and intention-to-view decreased at an increasing rate every week after the initial peak. Third, the probability the marginal consumer would execute on their intention-to-view decreased over time as more consumers adopted. Collectively, these three insights indicate that the dynamics underlying the exponential decay pattern observed at the box-office are more complex than a simple decay-function or Bass-diffusion process can explain. Figure 2 illustrates the insights of the second dataset for *Oceans Thirteen*, an action film released in 2007.

Figure 2: Audience Tracking Data: *Oceans Thirteen (2009)*

13
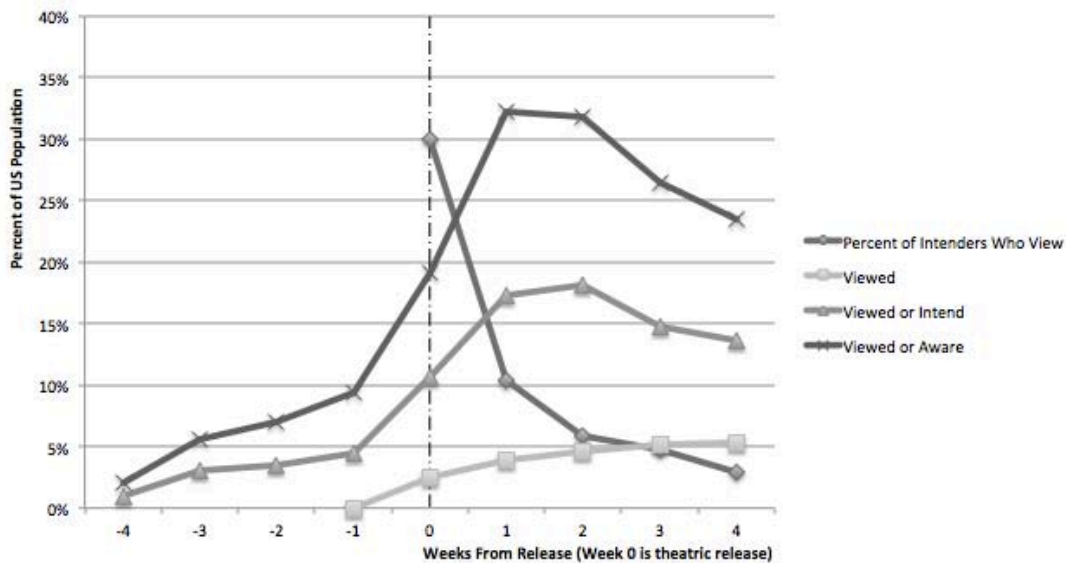
3.2. Estimation Method

The model was estimated using the Gauss-Marquardt-Levenberg algorithm within the PEST software package (Doherty, 2004). The algorithm's goal was to minimize the squared errors between the weekly simulation predictions and the historical data, while ensuring that awareness peaked after the movie's initial release and that the percent of intenders who viewed each week were within feasible ranges.

For computational efficiency the calibration process was estimated via a two-step procedure. First, an initial calibration run converged a generic set of parameter values towards a feasible set of movie-specific values in a partial model that ignored the effect of seasonality and competition. The resulting parameters were then used to kick off a second calibration run within the full model. The generic set of parameter values were pulled from previous academic research, industry reports, and through conversations with industry analysts.

3.3. Estimation Results and Interpretation

At this moment only the first half of study's estimation process is complete.[4] Nonetheless, several insights can already be observed. This section first describes the preliminary results for the 137 movies estimated by the partial model, and then previews the estimation results for a small set of movies within the full model to illustrate the importance of using a full-set of parameters.

Interpretation of the Estimation Results for the Partial Model

Although the purpose of the initial run is only to estimate a feasible set of parameter values in order to reduce the computational burden of estimating movies in the full model, the results indicate that the model accurately reproduces a wide range of weekly box-office patterns. Table 2 displays the summary statistics for the first estimation run.

The partial model alone achieved a median r-squared of 0.98 and fit 91 percent of the movies with an r-squared greater than 0.75. Additionally, every movie passed the peak-awareness test (indicating that in the simulated audience awareness increased after the movie's initial release). However, the probability that consumers in the Intender state would view a movie in its opening weeks was higher than expected, suggesting that the model may be systematically underestimated the true number of consumers who intend-to-view, and overestimated the movie's expected utility to the consumer. However, until a complete dataset can be constructed that tracks audience awareness and intention-to-view, this critique cannot be confirmed.

---

[4] An updated model that corrects for the non-Friday released movies has been designed but the estimation process will not be complete until after the study's deadline because the entire estimation process requires roughly 8 weeks for the estimation process to estimate 155 movies in the dataset on the researcher's cpu

Table 2: Summary of Calibration Outputs

	Movie R-Sq	Awareness Test*	Peak Awareness (%)	Choice Prob Opening Week (%)	Choice Prob Second Week (%)	Theme Acceptability (%)	Viewer Satisfaction (%)	Movie Marketability**	Movie Playability**
	-	-	(%)	(%)	(%)	(%)	(%)	-	-
Titles	137	137	137	137	137	137	137	137	137
Mean	0.91	1	8.47	63.0	41.3	29.5	36.9	0.53	0.25
Standard Deviation	0.17	0	6.67	27.0	16.0	9.7	2.1	0.10	0.11
Median	0.98	1	6.67	65.5	44.6	30.3	36.7	0.50	0.26
Max	1.00	1	58.34	100	62.8	49.7	49.5	0.95	0.79
Min	0.05	1	1.18	0	0	6.7	32.1	0.26	0.02

*Value of 1 indicates audience awareness peaked after the movie's initial release date
**Dimensionless value impacting *Perceived Movie Quality* bound between 0 and 1

Drilling down into the estimation results reveals that the strength of the model's fit is sensitive by the movie's release strategy and its cumulative box-office revenue (Figure 3). Results were strongest for movies that deployed the wide-release marketing strategies. Performance for platform-release movies varied significantly: six were estimated with r-squared values greater than 0.90, seven with an r-squared greater than 0.50, but four were under 0.50. The lone limited-release movie in the dataset failed to be estimated better than the generic parameter values in the first run. Regarding cumulative revenues, the model achieved a median r-squared of 0.91 for movies that generated less than $10 million, 0.97 for movies that generated between $10 and $100 million, and 0.99 for movies that generated more than $100 million.

Figure 3: Estimated Movie R-Squared in the Partial Model: *All Movies (2009)*
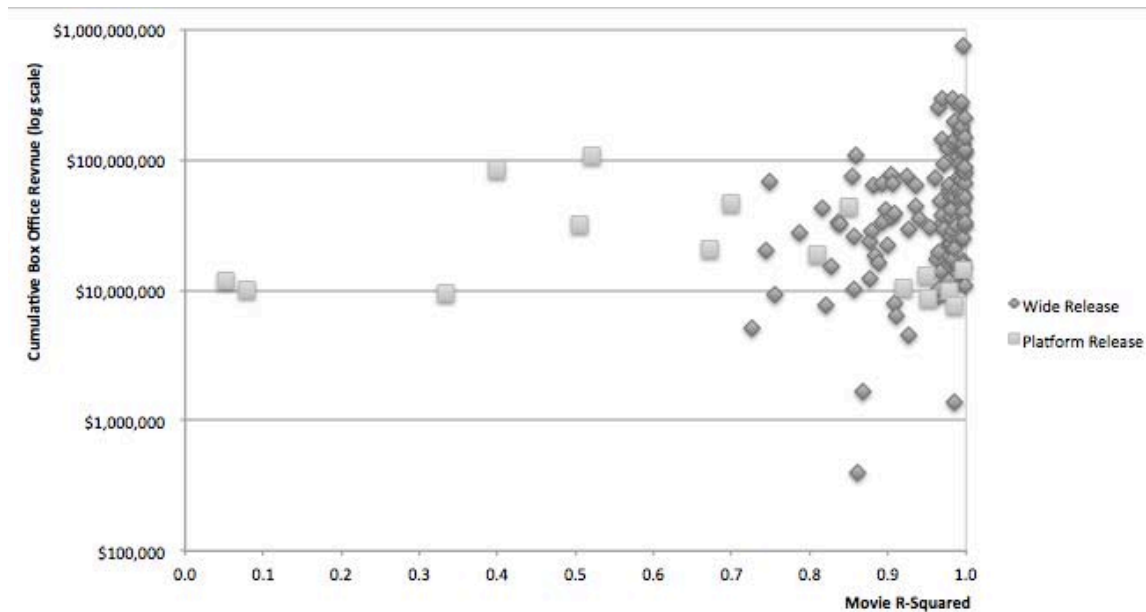

To better understand the weaknesses of the estimation methodology, a quick manual calibration of the lone limited-release movie and the two poorest performing platform-release movies yielded fits better than 0.90 for each film, suggesting the model may be

capable of explaining the entire dataset but that the automated estimation methodology needs to be revisited and improved upon. Additionally, it is expected that the second estimation process will enhance the fit for all movies, and it is very possible that accounting for seasonality and competition may dramatically improve the fit for movies with limited initial drawing power, such as low-revenue and platform release movies.

Interpretation of the Estimation Results for the Full Model

A small number of movies have been estimated in the full model to illustrate the importance of incorporating seasonality and competition in the full model. Figure 4 uses the movie *Paul Blart: Mall Cop*, a wide-releasing comedy that grossed $146 million, to compare the box-office estimates for the partial and the full models. The movie benefited from two long-weekend holidays: Martin Luther King weekend (week 0) and President's weekend (week 4). The partial model achieves an r-squared of 0.97, but entirely misses uptick in week 4. In contrast, the full model precisely catches the uptick, and fits the historic data with an r-squared greater than 0.99. Beyond providing a better numerical fit, the parameter values shift in a subtle but important way: by accounting for the movie's holiday opening, the full model estimates a significantly lower value for the movie's *marketability* parameter (which falls from 0.58 to 0.45) that is much closer to the *playability* parameter's value (unchanged at 0.40), suggesting a more feasible scenario in which the movie's quality was roughly in line with the audience's initial expectations.

Figure 4: Estimated Weekly Box-Office Revenue for *Paul Blart: Mall Cop (2009)*
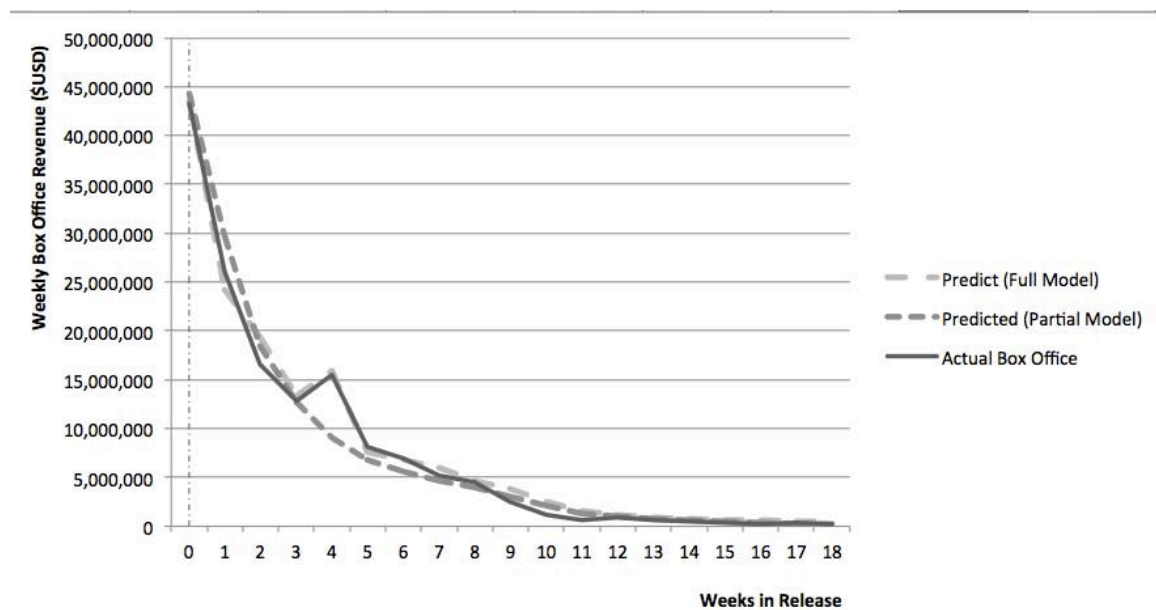
4. Managerial Implications

The model can be used by movie distributors to select movie release strategies by estimating model parameters as a function of the movie's intrinsic attributes and/or by

pulling the parameter values from *comparable* films in the database,[5] and then testing alternative strategies over a range of scenarios via Monte-Carlo methods. This section will describe how the model can be used to select revenue-maximizing strategies through two examples: optimizing the release dates for the movie *Coraline*, and optimizing the distribution strategy for *Taken*.

4.1. Selecting Optimal Movie Release Dates: *Coraline (2009)*

The model can be used to determine a movie's optimal release date following a two-step process. First constructing a competitive release calendar to estimate the potential drawing power of competitive films, and then searching for the revenue-maximizing release date given the competitive landscape. To illustrate this process, Table 3 shows the expected revenues for five potential release dates for movie *Coraline*, a wide-releasing stop-motion animation film. Holding everything else constant, the model predicts that although the movie's would have made an additional $10 million in its opening week had been released one week later, its expected cumulative revenue were highest for its original weekend.

Table 3: Comparison of Expected Box Office Revenue by Release Date: *Coraline (2009)*

Release Date Shift	23-Jan-2009 Shift - 2 Weeks	30-Jan-2009 Shift -1 Week	**6-Feb-2009** **Actual Release Date**	13-Feb-2009 Shift +1 Week	20-Feb-2009 Shift +2 Week
Opening Week Box ($ USD)	20,035,056	20,740,587	19,959,879	**30,327,269**	22,210,118
Second Week Box ($ USD)	14,645,448	13,431,580	**20,328,290**	13,914,517	16,268,174
Cumulative Box ($ USD)	71,876,130	73,414,089	**83,271,812***	77,389,193	74,978,329
Percent of Maximum	86.3%	88.2%	**100%**	92.9%	90.0%

*Actual cumulative box was $75 M. Estimated r-squared for weekly revenue was 0.98
Bold values indicates the maximum value among the 5 release strategies

Interpreting the model outputs in the context of the movie's parameters tells a deeper story. By releasing the movie one week before the holiday weekend, the distributors capitalized on a high volume of viewer word-of-mouth to help mitigate the audience's initial uncertainty regarding the movie's theme (*Coraline* had a relatively low *theme acceptability* and a relatively high *word-of-mouth effectiveness*) just in time for the upcoming holiday weekend. In fact, *Coraline's* second week actually outperformed its first, a rare feat for wide-releasing movies which typically experience 40-60 percent

---

[5] Data aggregators often described movies as "*comparable to…*" previously produced films. Two example include: Box Office Mojo (2012) who generates lists of movies with similar audience appeal, genre, tone, timeframe, and release patterns for free, and Netflix, who uses sophisticated matching algorithms to suggest new films to their members. Note that directly comparing movie-specific parameters between movies in the dataset must be made with the understanding that parameter values are estimated within a network of differential equations and their values must be interpreted within the context of the other parameters. Thus, while comparing the *theme acceptability* parameter between two movies with similar target audiences and marketing budgets is certainly useful (e.g., Disney's *Up* & DreamWorks' *Monsters versus Aliens*), comparing the parameter between two movies with entirely different target audiences and marketing budgets would be inappropriate (e.g., James Cameron's *Avatar* and Michael Moore's *Capitalism: A Love Story*).

declines in box-office sales between their first and second weeks. It is not difficult to imagine how this example could be expanded upon to optimize a studio's entire portfolio of upcoming releases to maximize individual movie revenues while avoiding the cannibalization of sales between movies.

4.2. Selecting Optimal Distribution Strategies: *Taken (2009)*

The model can be used to optimize a movie's distribution strategies given its intrinsic attributes. Specifically, the model can assist distributors to select between wide- and platform-release strategies, and to determine if boosting advertising spending and/or negotiating to release on a greater number of theaters might secure the network effects necessary to achieve blockbuster success. Figure 5 displays the estimated weekly box-office revenue for *Taken*, a wide-releasing movie that displayed characteristics common to successful platform-releases (namely, a relative low initial week at the box office, and an unusually slow attrition of weekly ticket sales), following wide- and platform-release strategies while holding everything else constant.

Figure 5: Comparison of Expected Box-Office Revenue by Distribution Strategies: *Taken (2009)*
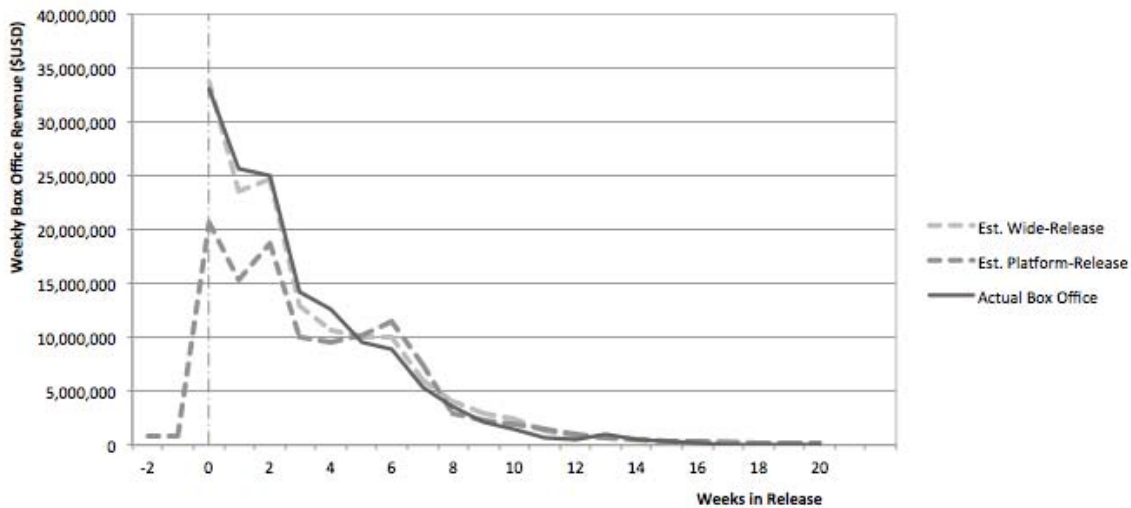
The model's estimates for the movie's wide-release corresponds almost perfectly with the historical data, achieving an r-squared greater than 0.99 and precisely hitting the true cumulative revenue of $145 million. To explore the platform-release scenario, the movie's was run on 20 theaters for two weeks before being released widely on its historic release date.[6] Interestingly, the model predicts the movie would do substantially worse following the platform-release strategy: generate only $117 million, 80 percent of the historical total. Again, the movie's estimated parameter values tell the story: the movie's *marketability* was significantly greater (0.58) than its *playability* (0.42), indicating the audience expected the movie to be more entertaining than it actually was, and as information about the movie's true quality diffused through the general pubic, it diminished the movie's potential audience in its first week of wide release. The smaller

---

[6] Shifting the platform-release date two weeks prior to the wide release was done to ensure the competitive effects on the movie are the same throughout the movie's box-office run. Running the same test without the shift produced equivalent results: estimating the expected cumulative revenue to be $114 million

initial audience reduced the buzz that surrounded the movie in its opening week, which evaporated the impact of network effects on consumer choice in each successive week after its initial release. The effect compounded from one week to the next, and ultimately reduced the movie's expected cumulative box by $38 million.

5. Conclusion

This paper has described a pre-release marketing model that can be used to optimize movie release strategies. The model was tested with data from 137 movies released in 2009 and achieved a median r-squared of 0.98 and fit 91 percent of the movies with an r-squared greater than 0.75. Additionally, the model matched known patterns for audience awareness and intention-to-view over time.

The model captures adoption as a behavioral process and accounts for word-of-mouth, consumer heterogeneity, and the product availability. A logit-choice equation is postulated that captures choice as a function of the movie's perceived quality, network effects, seasonality, and marketplace competition. The impact of managerial factors such as the movie's release date, distribution strategy, and weekly advertising exposures are also considered.

The paper describes how the model can be used to optimize a movie's release date and distribution strategy. As the database of movies calibrated by the model grows, the model will be able to forecast the box-office revenue of new releases by using the parameter values of *comparable* films that have been previously calibrated by the model.

The study yields five insights useful to future researchers: one, audience awareness and intention-to-view climb significantly after movies are released; two, the probability consumers will execute on their intention-to-view decreases each week after the movie is released; three, behavioral models are capable of accurately replicating a large number of box-office adoption patterns; four, incorporating seasonality and marketplace competition improves the ability of models to fit historical data at the box-office; and five, managerial influences have a significance influence on the speed and the scale that movies are adopted.

I would like to highlight two gaps in the academic literature that would be useful for future researchers to pursue.
1. Research into the evolution of audience awareness and intention-to-view over time, which would provide researchers with additional data points for calibrating models
2. Research exploring the attributes most important to purchase-point decisions at the box office, which would facilitate the development of richer and more robust consumer choice function

Many more research questions can be spun off from this study, however, I refer researchers to Eliashberg et al. (2006) for an excellent summary of previously conducted industry research, and a well-categorized list of potential research directions.

Works Cited

Ainslie, A., Dreze, X., & Zufryden, F. (2005). Modeling Movies Life Cycles and Market Share. *Marketing Science* , 508-517.

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

Box Office Mojo. (2012, 06 25). *Box office tracking by time*. Retrieved 06 25, 2012, from http://boxofficemojo.com/about/boxoffice.htm

Box Office Mojo. (2012, 06 25). *New Chart: Similar Movies*. Retrieved 06 25, 2012, from Box Office Mojo: http://boxofficemojo.com/siteinfo/?id=1625&p=.htm

Bruno, H. A., & Vilcassim, N. J. (2008). Structural Demand Estimation with Varying Product Availability. *Marketing Science* .

Chatterjee, R., & Eliashberg, J. (1990). The Innovation Diffusion Process in a Heterogeneous Population: A Micromodeling Approach. *Management Science , 36* (9), 1057-1079.

Delre, S. A. (2007). The Effects of Social Networks on Innovation Diffusion and Market Dynamics. *PhD Thesis* . Ridderkerk, The Netherlands: Labyrinth Publications.

Doherty, J. (2004). *PEST: Model-Independent Parameter Estimation User Manual: 5th edition.* Retrieved 06 01, 2010, from http://www.pesthomepage.org/Downloads.php

Einav, L. (2007). Seasonality in the U.S. Motion Picture Industry. *The RAND Journal of Economics* , 127-145.

Eliashberg, J., Elberse, A., & Leenders, M. A. (2006). The motion picture industry: critical issues in practice, current research, and new research directions. *Marketing Science* , 638-661.

Eliashberg, J., Jonker, J.-J., Sawhney, M. S., & Wierenga, B. (2000). MOVIEMOD: AN Implementable Decision Support System for Prerelease Market Evaluation of Motion Pictures. *Marketing Science , 19* (3), 226-243.

Forrester, J. W. (1961). *Industrial dynamics.* Cambridge: MIT Press.

Hanssens, D. M., & Parsons, L. J. (1993). Econometric and Time-Series Market Response Models. In G. L. J. Eliashberg, *Handbooks in OR & MS, Vol 5.* New York: Elsevier Science Publishers.

Horsky, D., Mirsa, S., & Nelson, P. (2006). Observed and Unobserved Preference Heterogeneity in Brand-Choice Models. *Marketing Science , 9* (4), 342-365.

Ishii, A., Matsuda, N., Arakaki, H., Umemura, S., Urushidani, T., Yamagata, N., et al. (2012). The 'hit' phenomenon: a mathematical model of human dynamics interactions as a stochastic process. *New Jornal of Physics* .

Krider, R. E., & Weinberg, C. B. (1998). Competitive dynamics and the introduction of new products: the motion picutre timing game. *Journal of Marketing Research* , 1-15.

Liechty, J. C., Fong, D. K., & DeSarbo, W. (2005). Dynamic Models Incorporating Individual Heterogeneity: Utility Evolution in Conjoing Analysis. *Marketing Science* .

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

Mahajan, V., Muller, E., & Kerin, R. A. (1984). Introducing strategy for new products with positive and negative word-of-mouth. *Management Science* , 1389-1404.

Pautz, M. (2002). The Decline in Average Weekly Cinema Attendance: 1930-2000. *Issues in Political Economy , 11*.

PwC. (2011). *Global entertainment and media outlokk: 2011-2015.* New York: PwC.

Rahmandad, H., & Sterman, J. (2008). Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models. *Management Science , 54* (5), 998-1014.

Silk, A. J., & Urban, G. L. (1978). Pre-test market evaluation of new packaged goods. *Journal of Marketing Research* , 171-179.

Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modeling in a Complex World.* Boston: McGraw-Hill Higher Education.

Sterman, J. D., Henderson, R., Beinhocker, E. D., & Newman, L. I. (2007). Getting Big Too Fast: Strategic Dynamics with Increasing Returns and Bounded Rationality. *Management Science* , 683-696.

Train, K. E. (2009). *Discrete Choice Methods with Simulation: Second Edition.* New York, NY: Cambridge University Press.

Urban, G. L. (1970). Sprinter Mod III: a model for the analysis of new frequently purchased consumer products. *Operations Research* , 805-854.

Urban, G. L., Hauser, J. R., & Roberts, J. H. (1990). Prelaunch Forecasting of New Automobiles. *Management Science* , 401-421.

Vogel, H. L. (1994). *Entertainment Industry Economics: A Guide for Financial Analysis.* Cambridge, UK: Cambridge University Press.