The Dual Impact of Movie Piracy on Box-Office Revenue: Cannibalization and Promotion

Liye Ma, Alan L. Montgomery, and Michael D. Smith

Carnegie Mellon University
Tepper School of Business and Heinz College
5000 Forbes Ave
Pittsburgh, Pennsylvania

February 2016

Liye Ma (liyema@rhsmith.umd.edu) is an Assistant Professor at the Smith School of Business at the University of Maryland. Alan L. Montgomery (alanmontgomery@cmu.edu) is an Associate Professor at the Tepper School of Business, and Michael D. Smith is a Professor at the Heinz College at Carnegie Mellon University.

The Dual Impact of Movie Piracy on Box-Office Revenue: Cannibalization and Promotion

Abstract

There are two main hypothesized effects from movie piracy: a cannibalization effect which reduces legitimate sales, and a promotional effect which increases word-of-mouth and stimulates sales. While these two effects are commonly discussed, there has been no research to measure their relative impact on motion picture sales. In this paper we use a hidden Markov model adapted from MOVIEMOD to decompose, and separately measure, the cannibalization and promotional impacts of piracy.

Using data from all wide release movies in the US from 2006 to 2008 we show that if piracy could be eliminated from the theatrical window then box-office revenues would increase by 15% or $1.3b per year. An analysis for the time period from 2011 to 2013 shows a similar increase of 14%. Our decomposition of piracy into separate cannibalization and promotional effects shows that the negative effects from piracy due to cannibalization dwarf any positive, promotional benefits: if piracy did not generate promotional effects through word-of-mouth communication then box-office revenues would drop by another 1.5%. We also find that, in rare instances (less than 3% of movies) promotional effects from pre-release piracy could increase revenue compared to piracy that occurs at release. Nonetheless, all of the movies in our counterfactual analysis would experience increased box-office revenue if piracy were eliminated altogether.

Keywords: Motion Pictures, Piracy, Box Office Revenue, Hidden Markov Models, Hierarchical Bayesian Models

1. Introduction

Quantifying the effect of piracy is an important research question for marketers (Eliashberg, Elberse and Leenders 2006). Most piracy is due to theft or negligence, but there is speculation in the press (Sandoval 2010, Moraski 2013, Barr 2015) and among industry bloggers (Crossley 2009, Mick 2009, Wallenstein 2010, de Moraes 2015) that some piracy may be deliberately orchestrated by marketers. Such deliberate piracy leaks would be consistent with theoretical analyses that propose piracy can increase revenues due to sampling, network effects, promotion and competition (Conner and Rumelt 1991, Prasad and Mahajan 2003, Chellappa and Shivendu 2005, Jain 2008, Vernik et al 2011). However, empirical verification of such theories is lacking (Liebowtiz 2006) and the predominant empirical finding (see the recent review by Danaher, Smith, and Telang 2014) has been that overall piracy is harmful to sales.

Empirical research on movie piracy has focused on measuring the overall effect of piracy (Ma et al 2014), but has not considered decomposing it into its two hypothesized effects (Danaher et al 2010, Smith and Telang 2010): cannibalization and promotion. Cannibalization is a direct reduction in sales of an authorized version due to consumption of a pirated copy. For example, moviegoers who would have paid to see the movie in the theater instead watch an illegal—but free—pirated version. An indirect promotional effect of piracy is that it can build awareness through word-of-mouth communication. The implication is that even though consumers do not watch the movie at the box-office, they still spread word-of-mouth messages about the movie and inform other consumers just as if they had seen a legal copy. Word-of-mouth messages would be positive for good movies or negative for bad ones.

Decomposing piracy into these two effects is important for marketing managers since their response to piracy depends upon the strength of each effect. If piracy has a strong promotional effect then it is possible that managers could craft strategies to profitably leverage its benefits but minimize its damage. Potentially there are specific movies or specific periods during their product life cycle when the promotional effects of piracy might outweigh the cannibalization effects. We believe this discussion points to the need for a methodology to be able to separately measure these effects, which is the focus of our research.

Measuring piracy's combined effects is straightforward since sales are observed. However, awareness and word-of-mouth messages are not observed so measuring the component effects of piracy is challenging. To address this challenge we propose a hidden Markov model of consumer behavior based upon MOVIEMOD (Eliashberg et al 2000). This model provides an explanation of how behavioral elements (awareness, word-of-mouth, and forgetting), marketing mix variables (advertising—both pre- and post-release, and availability—through the number of screens), and the intrinsic quality of the movie impact box-office revenue. It is flexible enough to model all movies within the same framework, both blockbusters which have exponentially declining sales curves as well as "sleeper" movies that have bell-shaped sales curves (Ainslie et al 2005). We also adjust for early partial week effects, for example when a movie is released on a Thursday instead of the usual Friday release, and limited release effects. Finally, we estimate the effect of both pre-release and post-release piracy through natural variation in when the pirated version of the movie is introduced, allowing us to estimate the impact of piracy compared to a counterfactual of what sales would be in the absence of piracy. This approach has many advantages over existing reduced form regression models of piracy like Ma et al 2014, such as being able to decompose piracy into promotional and cannibalization effects and make counterfactual predictions. However, this gain comes at the cost of making assumptions about consumer behavior that are not made in reduced form models.

To illustrate our method we estimate our model using data from all major movie releases in the United States during a three-year period from 2006 to 2008. Our data include piracy information collected from a unique Internet file-sharing site, allowing us to analyze the timing of piracy relative to the movie's box-office release date. We find that although movie piracy generates both cannibalization and promotional effects, the cannibalization effect vastly outweighs any promotional effect. Specifically, our estimates suggest that box-office revenue would be $1.3b (15%) higher per year if piracy could be eliminated entirely from the theatrical window and that piracy is particularly damaging to early sales. Although our model shows that while piracy is unequivocally harmful for overall sales, there is a small silver lining to piracy because of its promotional role. If piracy did not generate word-of-mouth promotion lift to movie attendance then box-office revenue would be another $158m (1.5%) lower per year. We repeated our analysis for 2011 through

2013 and find a similar result that box-office revenue would increase by 14% per year if piracy could be eliminated entirely from the theatrical window. In summary, we show that piracy is harmful to movie sales and that the word-of-mouth generated by pirated viewers helps to lessen, but does not offset, the negative cannibalization effect of piracy.

2. Literature Review

Movies have been a focus for marketing research (Eliashberg, Elberse and Leenders 2006). Some of the most important managerial questions for movie studios relate to forecasting a movie's box-office sales, timing release dates, and effectively directing promotion for movies prior to release. These questions have been widely studied in the marketing and information systems literatures (see, for example Chen et al. 2012; Duan et al. 2008; Eliashberg et al. 2007; Elberse 2007; Eliashberg and Shugan 1997; Liu 2006; Swami et al. 1999; Krider and Weinberg 1998 with respect to sales forecasting of movies and Prag and Casavant 1994, Zufryden 1996, and Elberse and Eliashberg 2003 with respect to advertising effectiveness), and have become established parts of movie studio practice.

More recently, studios have faced additional managerial questions around the impact of piracy on sales, and appropriate managerial responses to this potential threat. While the vast majority of empirical papers find that piracy harms sales (see Danaher, Smith, and Telang (2014) for a recent review of this literature), the analytic literature hypothesizes that piracy can also have a promotional impact through increased sampling (e.g., Peitz and Waelbroeck 2006; Chellappa and Shivendu 2005; Cheng, Sims and Teegen 1997, Sinha et al 2010) or through word-of-mouth promotional effects (e.g., Givon, Mahajan, Muller 1995, Prasad and Mahajan 2003, Liu, et al. 2011, Waters 2013), or can help alter the strategic interaction of firms (Jain 2008, Vernik et al 2011). However, awareness is a critical component of these hypothesized effects, and since awareness is not typically observed empirically, it is not possible to use standard reduced form approaches to disentangle the promotional and cannibalizing effects of piracy.

While there have been many papers in the literature that have studied the impact of piracy on music, there have been relatively few papers studying the impact of piracy on box-office sales of movies. The earliest of these papers, Bounie et al (2006), used a 2005 survey of piracy and purchase behavior by French university students and found that while piracy has a strong negative impact on VHS and DVD rentals and purchases, it had statistically no impact on theatrical revenue. To our knowledge, each subsequent published paper in the literature has found that piracy harms theatrical revenue. Specifically, Hennig-Thurau, Henning, and Sattler (2007) used a survey of movie purchase and piracy intentions among German consumers and found that piracy caused substantial losses in both DVD rentals and purchases and in theatrical revenue, and may have been responsible for $300 million of revenue losses in Germany. Rob and Waldfogel (2007) use a 2005 survey of movie purchase and piracy behavior for University of Pennsylvania students and find that each unit of pirated consumption reduces paid consumption by approximately one unit. DeVany and Walls (2010) model box-office revenue and the supply of pirated content for an unnamed movie and find that piracy accelerated this movie's box-office decline and may have caused $40 million of losses for this movie alone. Bai and Waldfogel (2013) use a 2008-2009 survey of Chinese university students' movie consumption behavior and estimate that each incidence of piracy among these students reduces paid consumption by about 0.14 units. Finally, Ma et al (2014) use a reduced form model and find that pre-release movie piracy reduces box office sales by 19% relative to the revenue that would have occurred if piracy were delayed until after release.

Our approach is to adapt the MOVIEMOD model developed by Eliashberg et al (2000) to the context of movie piracy. MOVIEMOD presents a validated and accepted structure from the marketing literature for understanding how awareness turns into adoption, and second, the parameters can be interpreted directly in terms of consumer behavior. However, our approach has many significant differences from Eliashberg et al (2000). Specifically our states are not observable while their states were, the likelihood for our states are defined differently (we combine exposure and impact together), we estimate our model using observed weekly box-office data as opposed to survey data, we incorporate piracy into the model and distinguish between pre-release and post-release periods, and we frame our model within a hierarchical Bayesian context. Our dataset and estimation methods reflect the difference between our objectives. Namely, the interest of Eliashberg et al (2000) was using survey data to predict box-office revenues for creating a pre-release decision support system, while we are interested in a different substantive question altogether: the impact of piracy on box-office revenues. In summary, the structure of MOVIEMOD is retained, but our implementation is quite different.

Our approach advances past empirical piracy research by allowing us to conduct counterfactual simulations to estimate box-office sales in the absence of piracy and by allowing us to disentangle the promotional and cannibalizing impact of piracy at a movie level. These results contribute to our knowledge of the promotional effects of piracy, specifically a branch of the literature that uses the Bass Model (Bass 1969) to understand the effect of word-of-mouth effects on software piracy (e.g., Givon et al 1995, Prasad and Mahajan 2003, Liu et al 2011, Waters 2013).

We use MOVIEMOD instead of the Bass model or a derivative because while the Bass model may be well suited to software sales, which typically present an asymmetric bell-shaped sales curve, this approach is inappropriate to model movie sales for several reasons. First, movie box-office revenue typically has an exponentially declining sales pattern: sales peak in the first week and then decline rapidly each week thereafter. Second, sales are strongly influenced by promotion and availability. Specifically, advertising promotion tends to be highly clustered in the weeks prior to release in order to generate awareness and interest amongst the movie-going population. Likewise, availability tends to be high in the opening weekend and then begins to taper quickly. Third, there are exceptions to this typical pattern in the case of sleeper movies (Ainslie et al 2005), movies which have slow builds and the peak occurs weeks or months after release. Fourth, the Bass model assumes that the purchase occasion and awareness are the same. Finally, the introduction of covariates, like promotion and availability, is not straightforward in the context of the Bass model, and the appropriate approach is often debated. In summary, the Bass model is not well suited for predicting movie sales.

3. Modeling Movie Box-office

3.1 A Hidden Markov Model of Movie Box-office

Our model is based on a discrete-time hidden Markov model and can be understood by the movements through a series of six decision states, listed in Table 1. These states are the same as those used in Mahajan et al. (1984) and Eliashberg et al (2000) and were retained to maintain consistency with the prior research. The potential transitions between our six states during one period are given in Figure 1 along with the parameters of our Markov transition matrix. Initially all consumers start as undecided. Consumers are then exposed to information about the movie from advertising or word of mouth effects. The exposure from word of mouth is dependent on past viewer behavior, hence our model falls in the class of interactive Markov processes (Conlisk 1976), which can yield S-shaped diffusion models as described by Bass (1969). If the movie theme is acceptable to the consumer based on this information, then consumers either immediately watch the movie (if available) or they may wait in the consideration state. Movies with unacceptable themes result in consumers moving to a rejection state. Upon watching the movie at either the theater or a pirated copy (either pre- or post-release) consumers form either positive or negative evaluations, and enter the respective spreading state in which they spread corresponding messages until they become inactive.

Index	*State*	*Description*	*A customer in this state has …*
1	U	Undecided	…not made a decision whether to watch the movie.
2	C	Consider	…decided to watch the movie.
3	S^+	Watched and Positive Spreader	…watched the movie, felt good about it, and is telling other people the movie is good.
4	S^-	Watched and Negative Spreader	…watched the movie, did not like it, and is telling other people the movie is bad.
5	I	Watched and Inactive	…watched the movie, and is not spreading information about the movie.
6	R	Rejecter	…decided to not watch the movie.

Table 1. Consumer States in Markov Model.

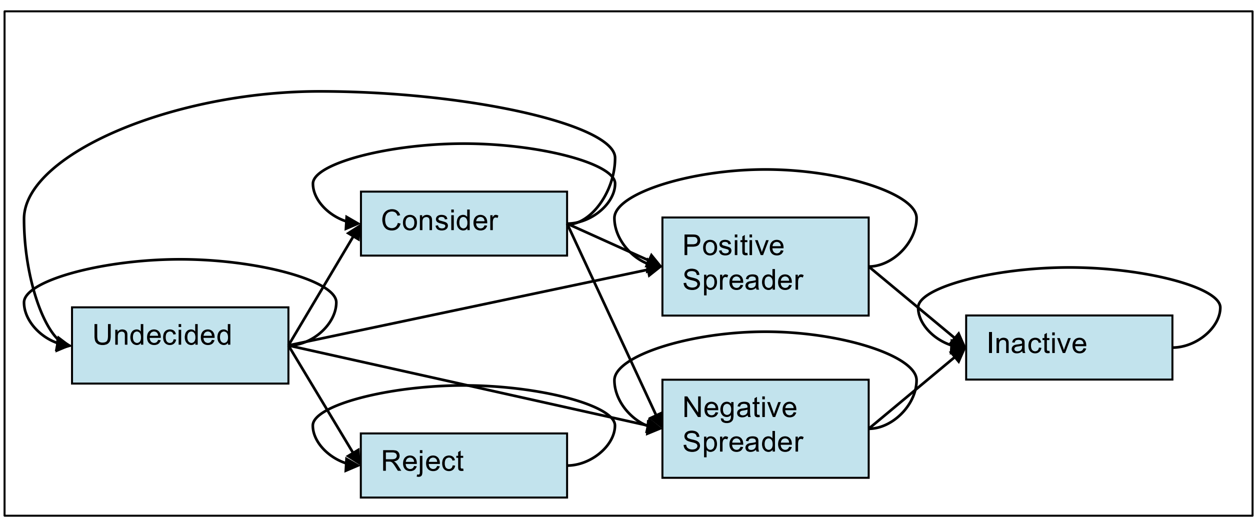

Figure 1. Consumer State Transitions for the Markov Model. The arrows denote permissible flows between states. Arrows that point back to a state represent consumers who may stay in the state.

3.2 State Transition Probabilities

Formally, we define the vector of probabilities associated with the states given in Table 1 as $\mathbf{X}_{it}$, where *i* is the movie and *t* is time.

$$\mathbf{X}_{it} = \begin{bmatrix} U_{it} & C_{it} & S_{it}^{+} & S_{it}^{-} & I_{it} & R_{it} \end{bmatrix}' \tag{1}$$

Initially we assume that all consumers begin in the undecided state for the initial time period[1]:

$$\mathbf{X}_{i1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}' \tag{2}$$

Our Markov process allows us to updates the probabilities from the previous time period as:

$$\mathbf{X}_{it} = \mathbf{A}_{it}\mathbf{X}_{i,t-1}, \text{ if } t > 1 \tag{3}$$

and our transition matrix at time *t* is:

$$\mathbf{A}_{it} = \begin{bmatrix} 1-\mathcal{E}_{it} & (1-\mathcal{V}_{it})(1-\phi_i) & 0 & 0 & 0 & 0 \\ \mathcal{E}_{it}\mathcal{L}_{it}(1-\mathcal{V}_{it}) & (1-\mathcal{V}_{it})\phi_i & 0 & 0 & 0 & 0 \\ \mathcal{E}_{it}\mathcal{L}_{it}\mathcal{V}_{it}\rho_i & \mathcal{V}_{it}\rho_i & \mu_i & 0 & 0 & 0 \\ \mathcal{E}_{it}\mathcal{L}_{it}\mathcal{V}_{it}(1-\rho_i) & \mathcal{V}_{it}(1-\rho_i) & 0 & \mu_i & 0 & 0 \\ 0 & 0 & 1-\mu_i & 1-\mu_i & 1 & 0 \\ \mathcal{E}_{it}(1-\mathcal{L}_{it}) & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where $\mathcal{E}_{it}, \mathcal{L}_{it}, \mathcal{V}_{it}, \rho_i, \phi_i, \mu_i$ are conditional probabilities whose construction are defined below. Notice the transition matrix has many null values due to the small number of allowable transitions as depicted in Figure 1. Each column of the transition matrix corresponds to the conditional probabilities of users who move from the previous state (column) to the new states (rows) where the index of the state is given in Table 1, and where, by construction, values in each column must sum to one.

[1] The initial time period typically occurs weeks or even months before the release date, since there may be significant pre-release advertising efforts or pre-release piracy.

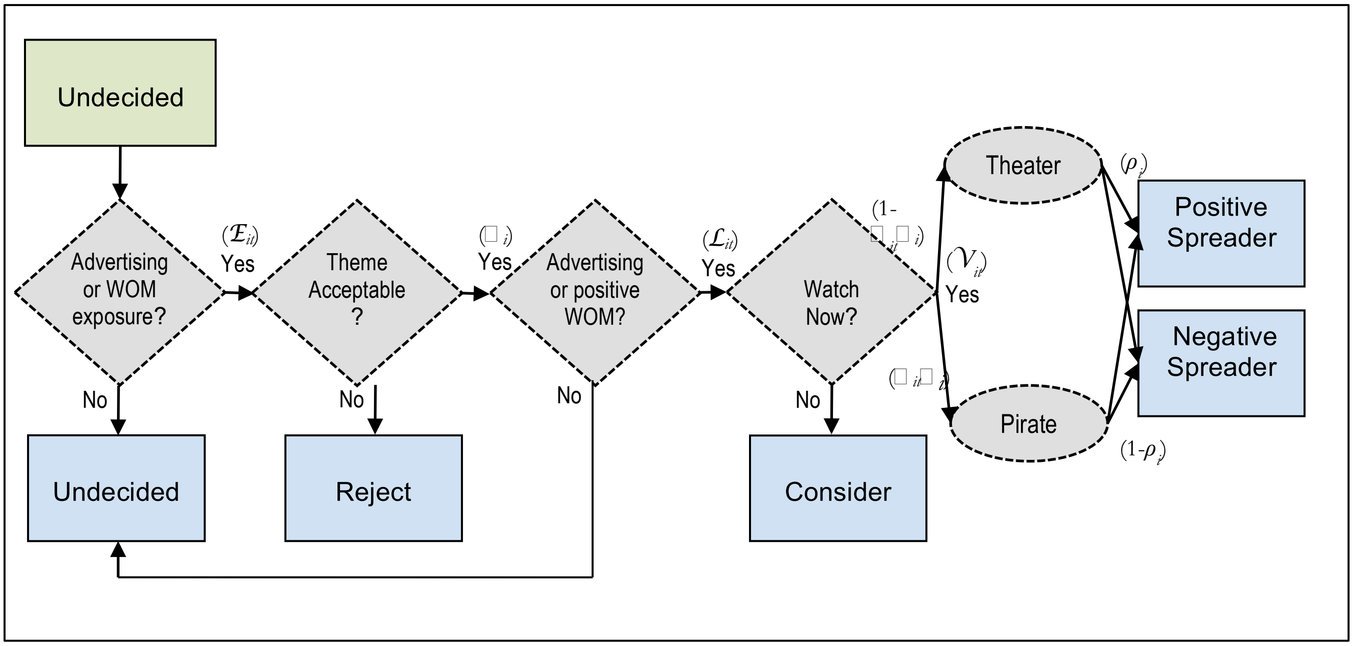

Figure 2a. Transition probabilities from undecided to alternative states. Probabilities of movements from the states are given next to the flow in parentheses. Gray boxes (outlined with dashed borders) denote intermediary states and are not an allowable terminal state for each period.

Transitions from Undecided. The potential transitions of an undecided consumer to become a positive or negative spreader, consider, reject, or remain undecided are illustrated in Figure 2a. The first step in moving away from the undecided state is to determine the probability of exposure to advertising or word-of-mouth. The probability of becoming exposed from advertising (α_{it}) depends on advertising expenditures, A_{it}, which vary over time as follows:

$$\alpha_{it} = 1 - \exp\{-\nu_i A_{it}\} \tag{5}$$

Note that advertising can occur pre- or post-release. Pre-release advertising is meant to help consideration even before the movie is available. We assume that the probability of being exposed to word-of-mouth spreaders is proportional to the number of exposed individuals in either state:

$$\omega_{it} = \lambda_i S^{+}_{i,t-1},\ \ \psi_{it} = \lambda_i S^{-}_{i,t-1} \tag{6}$$

The dependence of WOM on the probability of consumers having viewed the movie and actively spreading information allows S-shaped curves for the cumulative diffusion process and places our model in the class of interactive Markov Chains (Conlisk 1976).

We assume that the exposure to advertising, positive-word-of-mouth, or negative-word-of-mouth are independent, such that the probability that a consumer receives at least one exposure is:

$$\mathcal{E}_{it} = 1-(1-\alpha_{it})(1-\omega_{it})(1-\psi_{it}) \tag{7}$$

We assume the probability of a user being exposed to positive communication, either advertising or positive word-of-mouth, given that they have been exposed is:[2]

$$\begin{aligned}\mathcal{P}_{it} = \frac{1}{\mathcal{E}_{it}}\Big(&\tfrac{\alpha_{it}+\omega_{it}}{\alpha_{it}+\omega_{it}+\psi_{it}}\alpha_{it}\omega_{it}\psi_{it} + \alpha_{it}(1-\psi_{it}) + \tfrac{\alpha_{it}}{\alpha_{it}+\psi_{it}}\alpha_{it}(1-\omega_{it})\psi_{it} \\ &+ \tfrac{\omega_{it}}{\omega_{it}+\psi_{it}}(1-\alpha_{it})\omega_{it}\psi_{it} + (1-\alpha_{it})\omega_{it}(1-\psi_{it})\Big)\end{aligned} \tag{8}$$

The motivation for this equation is to account for potential multiple exposures. For example, consider the first term in parentheses from (8). This term is the probability that a consumer is simultaneously exposed to advertising, positive-word-of-mouth, and negative-word-of-mouth ($\alpha_{it}\omega_{it}\psi_{it}$) weighted by the fraction of these communications that are positive (e.g., either advertising or positive word-of-mouth): $\frac{\alpha_{it}+\omega_{it}}{\alpha_{it}+\omega_{it}+\psi_{it}}$. The remaining terms are computed similarly.

The product of those who have a positive exposure and accept the movie theme is:

$$\mathcal{L}_{it} = \mathcal{P}_{it}\theta_i \tag{9}$$

The probability that a user will instantly view a movie given a positive exposure and theme acceptance is:

$$\mathcal{V}_{it} = \delta^*_{it}\pi_i\tau_i + (1-\delta^*_{it}\pi_i)\tau_i\gamma_{it} \tag{10}$$

We assume that some proportion of viewers will be cannibalized by pirated consumption, π_i, but only if the pirated version is available. Pirated availability is denoted by a weekly indicator: $\delta^*_{it}=1$ if the pirated version is available and $\delta^*_{it}=0$ otherwise. The probability that a consumer of a movie with a pirated version will

[2] Eliashberg et al (2000) did not adjust for multiple exposures, and as a consequence its exposure terms cannot be strictly thought of as conditional probabilities. Additionally, we have combined exposure and impact, so all our exposures are impactful according to their model. Unfortunately, in our setting identification issues prevent us from separately identifying exposure and impact.

immediately watch is τ_i. Those who watch at the theater, $1-\delta_{it}^{*}\pi_i$, also consider the availability of the movie, which we assume is a function of the number of screens at which it is shown in week *t*, S_{it}:

$$\gamma_{it} = 1 - \exp\{-\upsilon_i S_{it}\} \tag{11}$$

Note that if the movie is not available at the box-office then S_{it}=0 and $\gamma_{it} = 0$. If theater distribution is saturated then $\gamma_{it} \to 1$ and both the pirated and theater versions have the same probability of being watched; otherwise the pirated version has an advantage in distribution over the theater version.

Given that a viewer has watched the movie, regardless of whether it is a pirated version or at the theater, we assume that ρ_i become positive spreaders, and the complement, $1-\rho_i$, become negative spreaders, where ρ_i can be thought of as a measure of the quality of the movie.

Finally, to illustrate the conditional probability of users from the undecided to positive spreader state we can take the product of these conditional probabilities: $\mathcal{E}_{it}\mathcal{L}_{it}\mathcal{V}_{it}\rho_i$, which gives us the (1,3) element of our transition matrix $\mathbf{A}_{it}$.

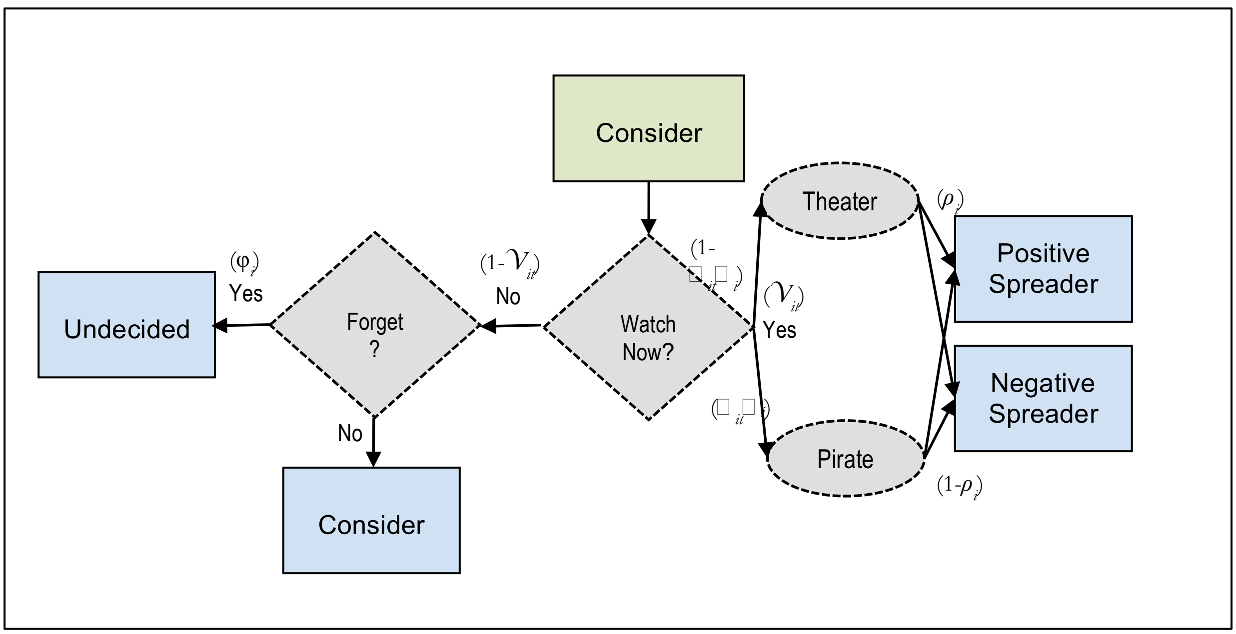

Figure 2b. Transition probabilities from consider to alternative states. Probabilities of movements from the states are given next to the flow in parentheses. Gray boxes (outlined with dashed borders) denote intermediary states and not allowable terminal states for each period.

Transitions from Consider. The potential transitions of those consumers who previously decided to consider the movie include moving to the positive or negative spreader states, remaining in the consider state, or moving to the reject state, and are illustrated in Figure 2b. The process in which consumers decide to watch now is the same as in the transitions from the undecided state. However, if the consumer decides not to watch then some proportion, $1-\phi_i$, forget and transition back to the undecided state, while the remainder, ϕ_i, remain in the consider state.

Transitions from Spreader States, Inactive and Reject States. Spreaders, either positive or negative, remain active with probability μ_i, and the remainder $1-\mu_i$ transition to the inactive state. The columns that correspond to reject and inactive are trivial since they are absorbing states and their probabilities are unity.

3.3 The Likelihood Function and Hierarchical Bayesian Specification

Our model can be used to predict the percentage of consumers who attend a movie *i* during week *t*, which we designate as $\overline{w}_{it}$. These predictions are formed by computing the number of transitions from undecided to watch due to advertising and word-of-mouth and also those who transition from the consider state to the watch state:[3]

$$\overline{w}_{it} = \left(1-\delta_{it}^{*}\pi_i\right)\tau_i\gamma_{it}\left(\mathcal{E}_{it}\mathcal{L}_{it}U_{i,t-1}+C_{i,t-1}\right) \tag{12}$$

Note that sales are only predicted to occur when the release is available, and if a pirated copy is available then π_i percent of moviegoers are lost due to cannibalization. Multiplying these share predictions by the market potential, which we designate as $\mathcal{M}$, allows us to predict sales as $\mathcal{M}\overline{w}_{it}$. We assume that the market potential[4] is fixed and known across all movies.

[3] We deviate from MOVIEMOD by allowing sales to derive from contemporaneous negative word-of-mouth.
[4] It is clear that market potential must exceed the box-office sales of the largest blockbuster (which in our data is Avatar, with box-office revenue of $750m), and conceptually one could argue its value is based on the total population size (300 million Americans times the average ticket price of $7.50). Once the parameter value of the potential exceeds some large value then it is difficult to distinguish the potential from the data alone, and therefore we fix its value at $1b.

To reflect the stochastic nature of sales we assume that weekly box-office receipts follow a Gamma distribution[5] around these predictions:

$$y_{it} \sim \Gamma\left(\sigma_i \mathcal{M}\,\overline{w}_{it}, 1/\sigma_i\right), \text{ where } \mathrm{E}\left[y_{it}\right] = \mathcal{M}\,\overline{w}_{it} \text{ and } \mathrm{Var}\left[y_{it}\right] = \mathcal{M}\,\overline{w}_{it} / \sigma_i \qquad (13)$$

Alternatively we could have chosen a normal distribution (Lehoczky 1980) or a multinomial distribution (Schmittlein and Mahajan 1982). A primary consideration in our choice of this likelihood is that the gamma and normal should approximate each other in this setting, since sales are large positive values. If the parameters of the multinomial follow a Dirichlet prior then the posterior will also be Dirichlet. In our setting the relative market shares (or ratio of a gamma distribution to the sum of gamma distributions) will be Dirichlet. Hence our proposed model can be thought of as a limiting case of these alternative specifications.

We frame our model in the context of a hierarchical Bayesian model, where the parameter vector for each movie is defined as follows:

$$\boldsymbol{\beta}_i = \left[\ln(\nu_i)\ \ln(\upsilon_i)\ \mathrm{l}\left(\lambda_i\right)\ \mathrm{l}\left(\tau_i\right)\ \mathrm{l}\left(\phi_i\right)\ \mathrm{l}\left(\mu_i\right)\ \mathrm{l}\left(\theta_i\right)\ \mathrm{l}\left(\rho_i\right)\ \mathrm{l}\left(\pi_i\right)\ \mathrm{l}\left(\tau_i\right)\ \ln\left(\sigma_i\right)\right]' \qquad (14)$$

where β_{ij}, the *j*th element of $\boldsymbol{\beta}_i$ the logistic transform, is defined as $\mathrm{l}\left(z\right) = \ln\left(z/_{1-z}\right)$ and has the corresponding inverse $z = 1/_{1-\exp\left\{-\mathrm{l}\left(z\right)\right\}}$. This ensures that the corresponding parameters fall between 0 and 1 and have the full range that corresponds to the normal distribution. We assume that these parameters are drawn from a hierarchical distribution as follows:

$$\boldsymbol{\beta}_i = \overline{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}_i, \text{ and } \boldsymbol{\varepsilon}_i \sim N\left(\mathbf{0}, \boldsymbol{\Sigma}\right) \qquad (15)$$

This random effects model allows each movie to have its own response. We could introduce covariates that capture movie genre, director appeal, star appeal, movie rating, and production budget into (15) to explain

[5] In our notation $X \sim \Gamma\left(\alpha, \beta\right)$ is defined as $p(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} \exp\left\{-x/\beta\right\}$.

this variation or encourage shrinkage in our estimates.[6] Please see the Appendix for an MCMC algorithm to estimate the model.

It is important to understand that there are several parameters that serve similar purposes in our model, which leads to estimation challenges. For example, slowly declining sales at the box-office can be attributed to retaining high levels of active spreaders, positive movie evaluations, or high exposures to word-of-mouth effects. Box-office revenues are positively related to an acceptable theme, positive evaluations, effective advertising, availability, or strong immediacy effects. In contrast, if the cannibalization effect of piracy is high then box-office revenue declines. Therefore the parameters compete with one another and each can effectively scale the entire box-office time series. However, the parameters enter distinctly into the likelihood function and can be uniquely estimated if sufficient information is available. Our challenge is that each movie tends to have a short-life at the box-office, which makes it difficult to find a precise estimate of the parameters for a particular movie without the hyperdistribution introduced in a hierarchical model. Therefore we are relying on the hyperdistribution to impart information about what range of parameter values are likely by distilling information across the cross-section of all movies.

4. Empirical Results

In this section we present the estimates from the model proposed in §3. First, §4.1 describes our dataset. In §4.2 we illustrate these parameters for a selected "blockbuster" movie and a "sleeper" movie to better motivate the model. We then present the mean and standard errors of the parameters for our set of movies in §4.3, which provides evidence for the heterogeneity in response across movies. Finally, in §5 we consider the implications of these parameters on our estimates of the effects of piracy on movie box-office revenue through a series of counterfactual predictions.

---

[6] If we regress our parameter estimates against movie characteristics we explain between 7% and 24% of the variation in these parameters across movies. This does not mean that this variation is not important, just that it is difficult to predict. In other words can we predict with high confidence whether a movie will be a blockbuster based upon these covariates, and our finding is that it is difficult to make this prediction based upon the covariates alone. We did try introducing such covariates into (15), but found similarly weak relationships and have omitted these results since it appears that the random effects associated with the individual-level component of the estimates is the more critical component and we wish to keep the presentation of the hierarchy as straightforward as possible.

4.1 Description of Dataset

We collect our data from four sources: BoxOfficeMojo, Nielsen Research, Kantor Research, and vcdquality.com.[7] Our data consist of all movies released between February 2006 and December 2008 that were listed by BoxOfficeMojo as "wide release titles". We obtained weekly advertising data from Kantor Research, which includes television, magazines, newspapers, radio, Internet and outdoor advertising. Our data from Nielsen Research reports the weekly box-office revenue and the number of screens upon which the movie is showing. Our information about movie piracy comes from vcdquality.com. This is not an Internet file-sharing site. Instead vcdquality.com monitors various Internet file sharing sites and posts a message on its website once a pirated copy of a movie becomes available at other piracy sites. Each message includes the date of availability, which allows us to infer that a pirated version of the movie was available. We know the date a pirated copy is posted from vcdquality.com and the official theatrical release date of the corresponding movie from BoxOfficeMojo. The difference between these two dates allows us to infer the week of release that piracy occurs, negative values indicate pre-release piracy, zero indicates piracy occurs in the same week as release, and positive values indicate post-release piracy.

Table 2 provides a table of the descriptive statistics for each of these variables. The average across the 533 movies in our dataset shows that the average box-office revenue was $48.1m, advertising was $19.8m, and the number of opening screens was 2,161. Almost half of movies (47%) were pirated during the week of release, 42% of movies were pirated after release, 10% of movies had pre-release piracy, and only seven movies had no piracy. The movies with no piracy tended to be low-budget films with low box office revenue, for example no pirated copies of "U2 3D" were identified in our dataset. The distribution of piracy is characterized by long tails. For example the movie "88 Minutes" was pirated via a screener over 65 weeks before its release in the US. It is likely piracy coincided with an overseas premiere in this example. The other extreme was "Expelled: No Intelligence Allowed" which was a Ben Stein documentary that was not pirated until nearly half a year after its release.

[7] All information is available on the Internet either through public or subscription services.

Variable	Mean	Std. Dev.	Minimum	Median	Maximum
Box-office Revenue ($ million)	$48.1	$58.5	$0.1	$29.4	$509.5
Advertising ($ millions)	$19.8	$11.5	$0.5	$17.3	$53.4
Opening Screens (at wide-release)	2161	1035	2	2344	4366
Week pirated version available	1.5	6.3	-65	0	25

Table 2. Descriptive Statistics for Movie Characteristics. There are 533 movies in our dataset that have wide releases between Feb. 2006 and Dec. 2008. The week for the first week a pirated version is available is expressed relative to the wide release (e.g., 0 represents at release, -1 represents pre-release piracy since the copy was available 2 weeks before release, 1 represents post-release piracy since the copy was available 2 weeks after release). In our dataset there were seven movies in which no piracy was reported during the initial theatrical release, so the week of piracy is conditional upon the 98.7% of movies for which piracy occurs during the theatrical window.

4.2 Illustration for a "Blockbuster" and a "Sleeper" Movie

To better understand the mechanics of the model in §3 we first consider its application to a prototypical blockbuster movie, "Mission: Impossible III." Figure 3 gives the box-office sales of the movie along with the predictions, the corresponding 95% confidence region, and number of screens and advertising. Notice that sales start at their peak and then decline exponentially. We also see that advertising, the main driver in building awareness and interest in this particular movie, begins three months before release, although most advertising occurs in the month before release. Typically movie studios start their pre-release advertising up to two months in advance of the release, but the most intense periods are the few weeks before launch. The number of screens begins to drop after box-office revenue declines, and as such is a lagging indicator of box-office rather than a constraint on potential sales. Finally, this movie is pirated in the week of release. This pattern of sales, advertising, screen release, and piracy is typical among most movies, particularly "blockbusters".

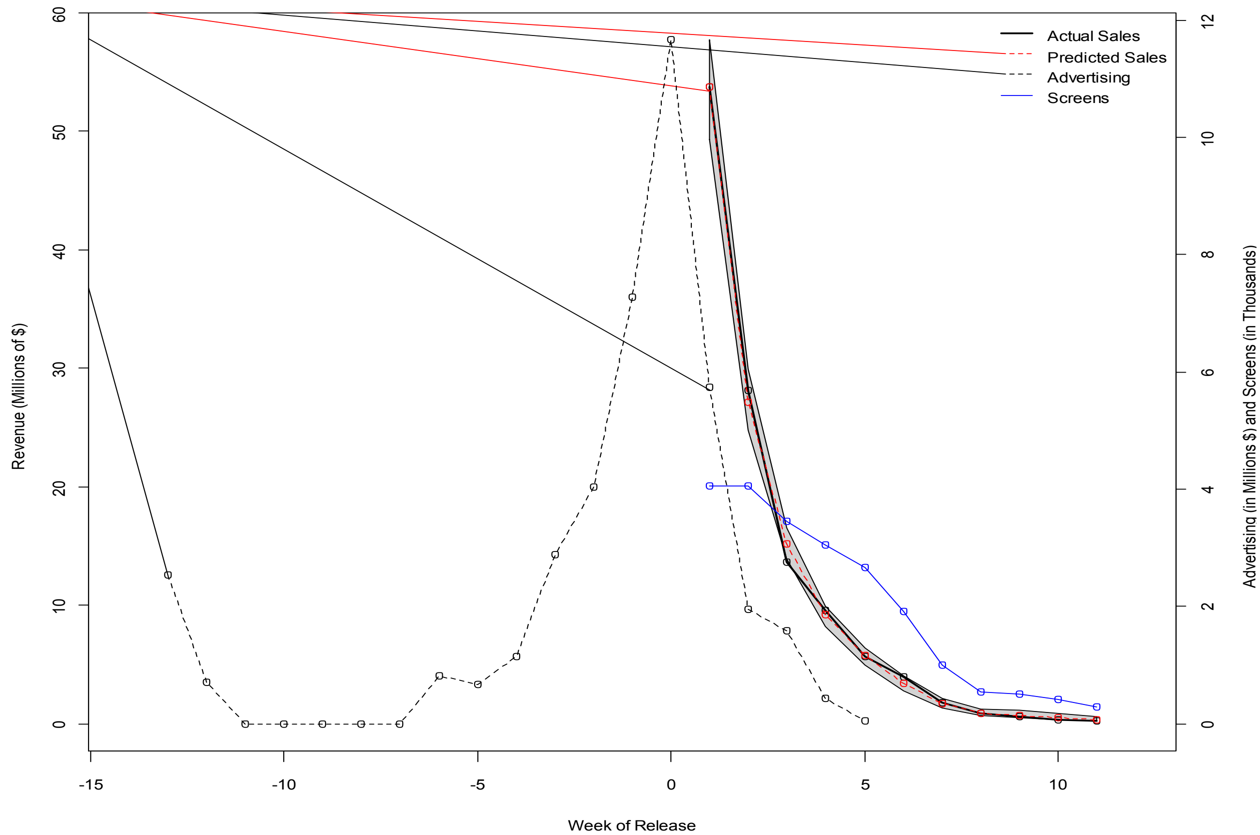

Figure 3. Actual and predicted box-office revenue, advertising, and screens for "Mission: Impossible III". The first week of release is 1. The shaded area denotes the 95% confidence interval for our sales forecasts.

Parameter	*Transform Mean*	*Transform Std. Err.*	*Mean*	*Std. Err.*
ν_i	-4.31	0.34	0.014	0.0044
υ_i	-8.03	0.58	0.00039	0.00027
λ_i	-0.56	0.66	0.38	0.14
τ_i	0.57	0.70	0.62	0.14
ϕ_i	0.55	0.50	0.63	0.10
μ_i	1.30	0.72	0.77	0.09
π_i	-2.74	1.36	0.10	0.11
θ_i	0.44	0.88	0.58	0.16
ρ_i	1.23	1.24	0.72	0.20
σ_i	-10.96	0.58	0.000020	0.000011

Table 3. Parameter Estimates of "Mission: Impossible III"

Table 3 provides the resulting parameter estimates and their corresponding transformed values. The parameters themselves can be interpreted in light of consumer behavior, since most can be thought of as conditional probabilities. For example, θ_i implies that 58% of exposed consumers find the movie's theme acceptable, and ρ_i states that 72% of consumers positively evaluate the movie. Both of these values imply that this is a well-received movie. Awareness from advertising is driven by ν_i and the amount of advertising, which for this movie was $12m at its peak. This yields a conditional exposure of 16% from advertising per week. λ_i measures awareness from word-of-mouth at 38% conditional upon exposure, and 77% of word-of-mouth spreaders remain spreaders in the next week as estimated by μ_i. This indicates that word-of-mouth is much more potent than advertising for this movie.

Once a consumer becomes a considerer there is a 63% chance of remaining a considerer in the next week as indicated by ϕ_i. Therefore if the movie is not available, awareness can decay quickly. Notice that consideration does not immediately led to consumption, based on υ_i and availability at 4,000 screens implies that 79% of considers will want to watch, and τ_i =62% will convert immediately. Hence, intentions do not translate into immediate action. The parameter π_i measures the cannibalization effect of piracy at 10% of consumers, although this effect is moderated by the movie being watched immediately. Finally, the scale parameter of the Gamma distribution is inversely related to our confidence in the forecasts, which in this case appear to be fairly accurate as illustrated by our confidence region of our sales forecasts in Figure 3.

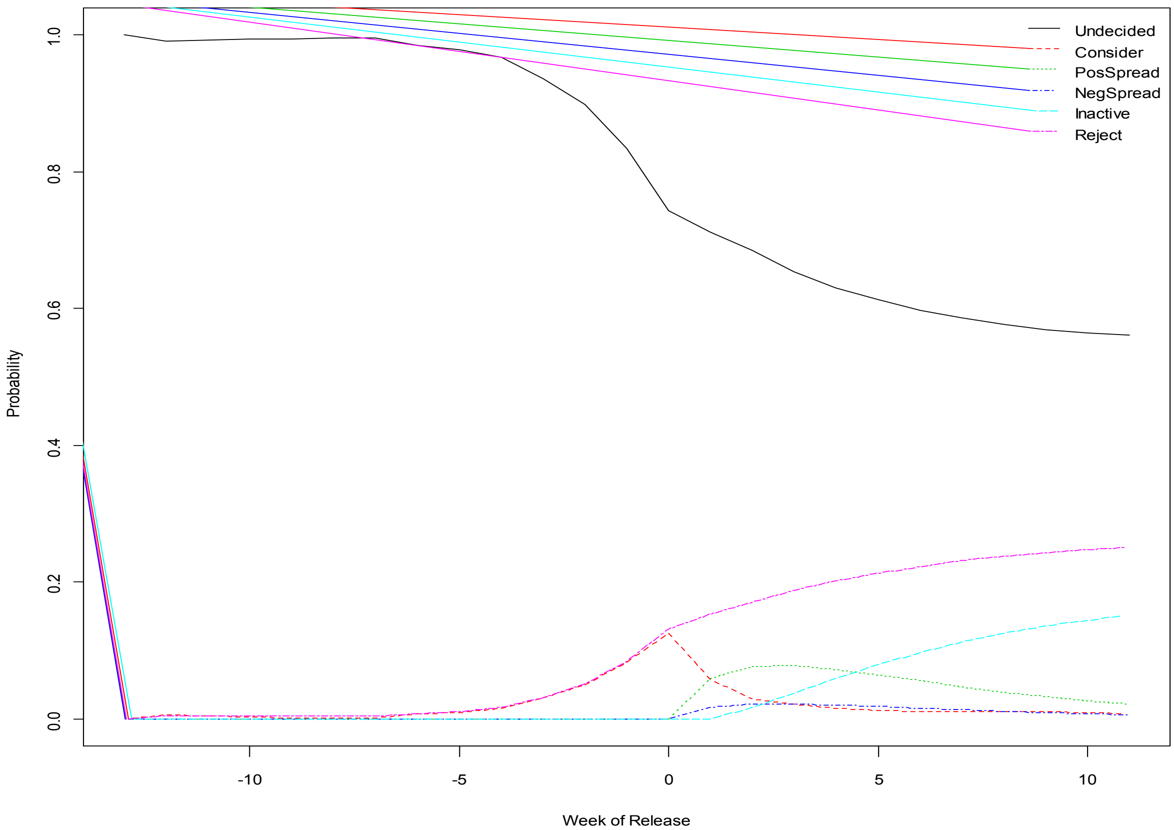

Figure 4. Probabilities for latent states by week of release for "Mission: Impossible III"

Our box-office revenue predictions can be understood by considering the latent state predictions in our hidden Markov model, which are plotted in Figure 4. Initially all consumers start in the undecided state. Note that prior to a movie's opening, the large amount of pre-release advertising has resulted in a large number of consumers who become aware of the movie and transition from the undecided state to the consider state. This movie is pirated at release so we do not attribute any word-of-mouth to piracy until release. Additionally, there is a notable increase in those consumers who reject the movie based on its theme alone. Once the movie is released, the diffusion from word-of-mouth effects begins. Note that once the release begins, advertising drops off quickly, which implies that the new considerers are largely the result of word-of-mouth effects. Positive-spreaders contribute to more consumers transitioning to the consider state, while consumers who hear from negative-spreaders transition to the rejection state. Consistent with the blockbuster status of the movie, almost half of the population (44%) has become aware of the movie.

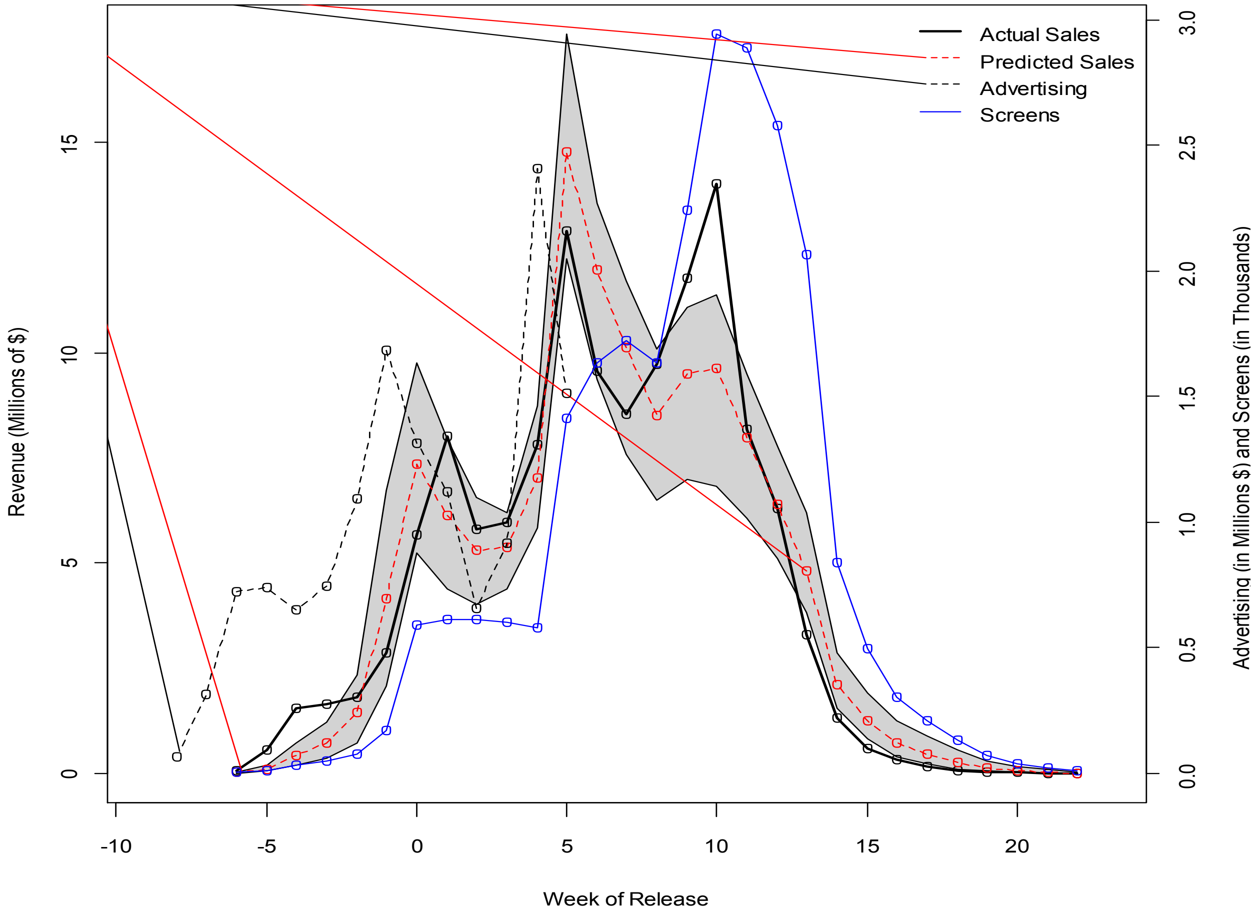

Figure 5. Actual and Predicted Sales for "Slumdog Millionaire".

In addition to explaining the behavior of "blockbuster" movies, the proposed model is flexible enough to describe sleeper movies that exhibit very different sales patterns than the typical blockbuster. These "sleeper" movies, like "Slumdog Millionaire," start with a limited release, but exhibit characteristics where increased interest builds the number of available screens over time such that the movie experiences a slower decline in sales than typical blockbuster movies do. This type of movie was one of the most problematic to model since its peak could occur over two months after release. The plot of actual and predicted sales for Slumdog Millionaire is given in Figure 5, the probabilities of the latent states are given in Figure 6, and the parameter estimates are given in Table 4. Note that although the model does a good job of tracking the general pattern of sales, it misses the very strong pre-release sales and under-predicts the third peak. Even though this pattern differs from the prototypical exponential declining sales pattern substantially, our model is still able to track this movie's pattern fairly well.

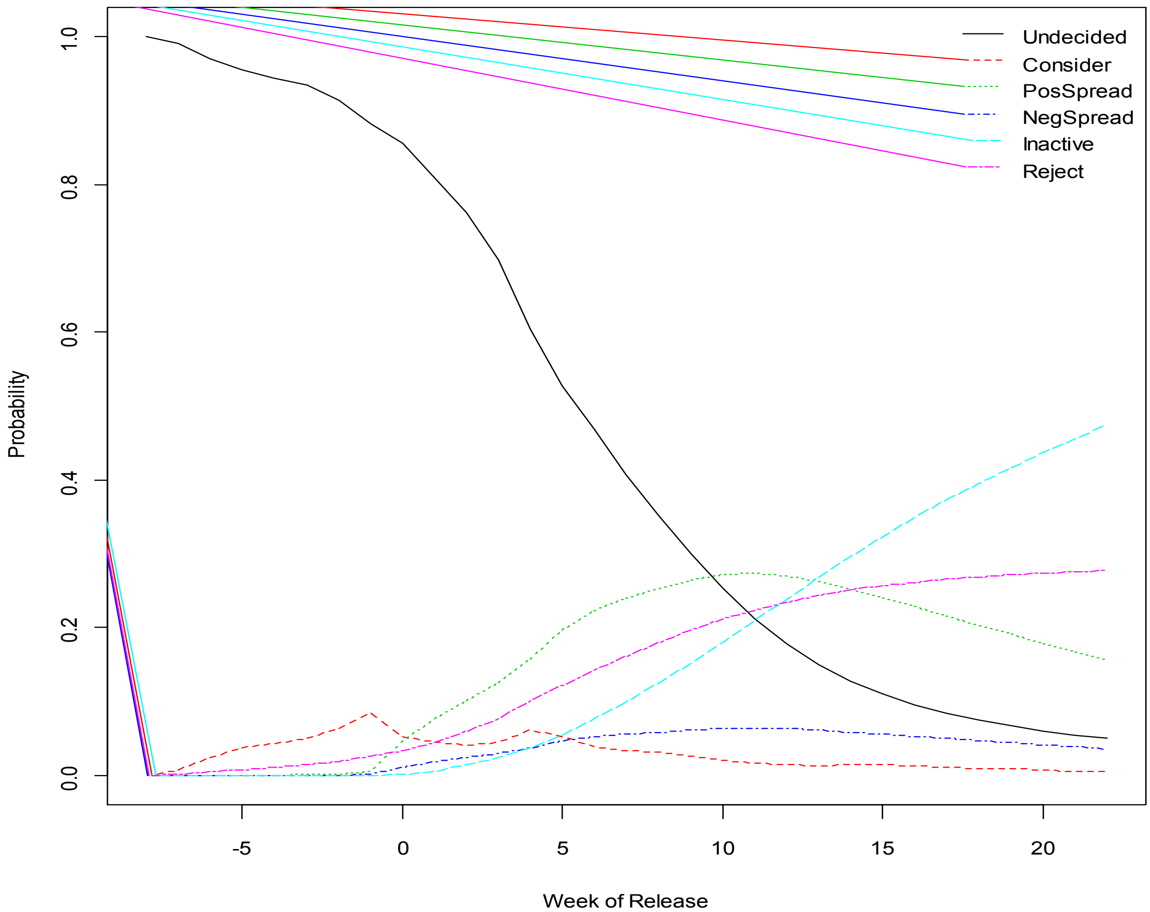

Figure 6. Probability of latent states for "Slumdog Millionaire"

Parameter	*Transform Mean*	*Transform Std. Err.*	*Mean*	*Std. Err.*
ν_i	-3.44	0.34	0.034	0.011
υ_i	-7.98	0.36	0.00037	0.00013
λ_i	0.08	0.42	0.52	0.101
τ_i	1.14	0.65	0.74	0.114
ϕ_i	0.91	0.68	0.70	0.135
μ_i	2.48	1.09	0.89	0.093
π_i	-0.06	0.54	0.49	0.114
θ_i	1.75	0.82	0.83	0.097
ρ_i	1.39	1.20	0.75	0.171
σ_i	-12.70	0.31	0.0000032	0.0000010

Table 4. Parameter Estimates of "Slumdog Millionaire"

Total box-office revenues for "Mission: Impossible III" and "Slumdog Millionaire" were $129m and $119m respectively, while advertising differ markedly at $41.5m and $14m. Again movie "quality" as measured by θ_i and ρ_i are both higher for "Slumdog Millionaire" than for "Mission: Impossible III". The largest differences between the movies are the susceptibility to piracy (lower for "Slumdog Millionaire") and theme acceptability (higher for "Slumdog Millionaire"). Also, the fact that "Slumdog Millionaire" was a victim of pre-release piracy while "Mission: Impossible III" was not is critical to our predictions. Based upon our estimated model we predict that "Slumdog Millionaire" sales would have been $15.5 million higher if it had been able to hold off piracy until its release.

Parameter	*Transform Mean*	*Transform Std. Dev.*	*Mean*	*Std. Dev.*	*Lower*	*Upper*
ν_i	-4.34	0.46	0.015	0.008	0.006	0.036
υ_i	-7.29	0.76	0.0011	0.0016	0.0002	0.0048
λ_i	-1.16	0.46	0.263	0.081	0.122	0.432
τ_i	-0.26	0.83	0.445	0.170	0.147	0.778
ϕ_i	0.61	0.56	0.633	0.118	0.358	0.837
μ_i	2.58	0.81	0.879	0.083	0.663	0.976
π_i	-3.11	0.84	0.096	0.100	0.026	0.400
θ_i	0.20	1.06	0.524	0.200	0.139	0.860
ρ_i	0.73	0.84	0.618	0.147	0.249	0.846
σ_i	-9.81	1.28	0.00016	0.00025	0.00001	0.00092

Table 5. Distribution of mean of parameter estimates across all movies along with standard deviations. Both the transformed and inverse of the parameters are given. Lower and Upper values correspond with the 2.5% and 97.5% percentiles.

4.3 Results

To illustrate the variation across our 533 movies, in Table 5 we provide the mean and standard deviation of the parameter estimates across out set of movies as well as their corresponding transformed estimates. Our conclusion is that there is substantial variation in response across movies. For example, the relative response in movies to advertising has a ten-fold increase from the lower to the upper bound. Of particular interest, we see the piracy parameter has a mean of 10% across movies, with 95% of movies ranging between 2.6% and 40.0%. This parameter is distributed in an inverse-J shape across movies, and its histogram is given in Figure 7.

The mean of the hierarchical distribution represents the central tendency across the movies. As expected, the mean of the parameter estimates is close to the mean of the hierarchical distribution. While there is a great deal of heterogeneity across movies, we can still reliably estimate the hierarchical distribution. For example, the mean of the piracy parameter is 10.0% with a standard error of 1.0%. Hence, we are confident that the cannibalization effect of piracy is detrimental to box-office sales.

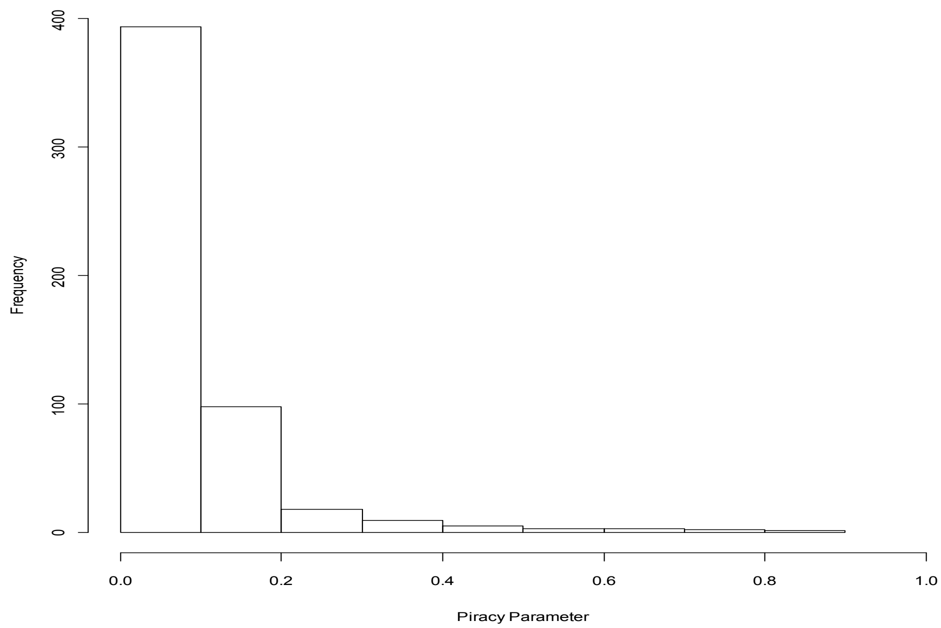

Figure 7. Histogram of the piracy parameter estimates across all movies.

Table 6 reports the mean and Table 7 gives the diagonals of the covariance matrix and correlation matrix for the hierarchical distribution. This covariance matrix controls for variation across movies, and the correlation matrix provides information about how the parameters co-vary. The correlation matrix is important in understanding that information about one parameter can improve the precision of the estimates of the other parameters. Note that the estimate of the standard deviation of the transformed piracy parameter is 1.644 (Table 7). However, the actual standard deviation observed across the movies is 0.84. The reduction in variation is due to the correlation among the parameters. For example, the conditional variance of the piracy parameter is reduced by 26% given knowledge of the other parameters.

Parameter	*Transform Mean*	*Transform Std. Err.*	*Mean*	*Std. Err.*	*Lower*	*Upper*
ν_i	-4.34	0.04	0.013	0.001	0.012	0.014
υ_i	-7.29	0.05	0.00069	0.00003	0.00063	0.00075
λ_i	-1.16	0.08	0.238	0.014	0.209	0.264
τ_i	-0.26	0.06	0.435	0.014	0.407	0.464
ϕ_i	0.61	0.06	0.649	0.015	0.622	0.677
μ_i	2.58	0.14	0.929	0.009	0.909	0.945
π_i	-3.11	0.24	0.044	0.010	0.028	0.067
θ_i	0.20	0.11	0.550	0.028	0.494	0.599
ρ_i	0.73	0.21	0.673	0.045	0.592	0.766
σ_i	-9.81	0.07	0.000055	0.000004	0.000048	0.000063

Table 6. Distribution of mean and standard error of the hierarchical distribution. Both the transformed and natural values of the parameters are given. Lower and Upper correspond to the 2.5% and 97.5% percentiles.

Parameter	*Estimate*	ν_i	υ_i	λ_i	τ_i	ϕ_i	μ_i	π_i	θ_i	ρ_i	σ_i
ν_i	0.589	1.000									
	(0.030)	(0.000)									
υ_i	0.889	-0.333	1.000								
	(0.035)	(0.066)	(0.000)								
λ_i	0.791	0.205	-0.584	1.000							
	(0.070)	(0.083)	(0.065)	(0.000)							
τ_i	1.063	0.087	-0.488	0.529	1.000						
	(0.058)	(0.077)	(0.041)	(0.058)	(0.000)						
ϕ_i	0.790	0.021	-0.268	0.061	0.421	1.000					
	(0.070)	(0.080)	(0.076)	(0.125)	(0.101)	(0.000)					
μ_i	1.409	0.070	-0.003	-0.114	-0.293	-0.147	1.000				
	(0.111)	(0.069)	(0.074)	(0.073)	(0.069)	(0.082)	(0.000)				
π_i	1.644	0.029	0.072	-0.051	0.025	0.046	-0.142	1.000			
	(0.150)	(0.067)	(0.072)	(0.074)	(0.074)	(0.063)	(0.101)	(0.000)			
θ_i	1.453	-0.167	0.151	-0.238	-0.436	-0.415	0.072	0.168	1.000		
	(0.080)	(0.065)	(0.062)	(0.065)	(0.051)	(0.052)	(0.077)	(0.085)	(0.000)		
ρ_i	1.627	-0.008	0.100	-0.085	-0.062	-0.151	-0.257	0.224	-0.096	1.000	
	(0.140)	(0.077)	(0.087)	(0.080)	(0.100)	(0.082)	(0.085)	(0.114)	(0.106)	(0.000)	
σ_i	1.449	-0.112	0.157	-0.085	-0.084	-0.185	-0.055	-0.006	-0.164	0.295	1.000
	(0.058)	(0.056)	(0.053)	(0.064)	(0.059)	(0.058)	(0.071)	(0.083)	(0.056)	(0.071)	(0.000)

Table 7. Estimate of the standard deviation of the diagonals of the covariance matrix for the hierarchical distribution of the transformed parameters, and the correlation matrix of the hierarchical distribution. The standard errors are given below the estimate in parentheses.

5. Counterfactual Studies of Piracy

Measuring the impact of piracy has been a major focus of the academic literature since the introduction of Napster. However, while many academic papers have analyzed the impact of piracy, these analyses typically are conditional upon when piracy actually occurs. The goal of our model and analysis is not to simply understand what the impact of piracy is on movie sales as they happen, but to consider hypothetical situations that change whether or when piracy would occur. We estimate our counterfactuals by setting the point at which piracy occurs—which may be substantially different than when piracy actually occurred—but holding other variables such as advertising and screens at their actual values. We construct a Monte Carlo simulation[8] so that we can marginalize over the posterior distribution of the parameter estimates, and we then compute the expectation of box-office revenue. Using this method we predict that the total box-office revenue[9] for our 533 movies over our 35 month period would be $25.66b (with a standard error of $0.04), which is the same sales amount as what we actually observe (or within the precision reported), indicating there is little bias in our predictions.

5.1 Predicting box-office revenue based upon when piracy occurs

Recall that in our dataset piracy occurs on average 1.5 weeks after release but can occur weeks or months before or after the release date (see Table 2). We wish to consider what would happen if all pirated releases were actually to occur in the same week of release. Specifically, if a pirated copy were made available at 8, 4, or 1 weeks before release, at release, and at 1, 4, or 8 weeks after release then how would these changes in piracy availability effect box-office revenue? Our estimates and standard errors of the box-office revenue are reported in Table 8. The actual revenue and predicted revenue based on the actual piracy pattern, as well as the predictions when no piracy occurs at all, are given for comparison. Additionally, we decompose box-office revenue by when the revenue occurs: either before widespread release (e.g., during limited release), the week of release (or at release), post-release, and the total.

---

[8] To speed up our simulations we thin our MCMC draws to every hundredth draw of the parameters. Since our original draws were highly autocorrelated, these new samples can be considered to be almost independent.

[9] Again we are only considering box-office revenues, and one could consider how piracy affects later markets such as DVD sales, cable releases, and television revenues.

Scenario	*Pre-Release*	*Std Err*	*At-Release*	*Std Err*	*Post-Release*	*Std Err*	*Total*	*Std Err*
Actual	$0.24		$10.08		$15.34		$25.66	
Predicted	$0.22	$0.01	$10.14	$0.02	$15.31	$0.03	$25.66	$0.04
Release-8	$0.19	$0.01	$9.34	$0.08	$14.61	$0.09	$24.14	$0.17
Release-4	$0.23	$0.01	$9.37	$0.08	$14.68	$0.08	$24.29	$0.16
Release-1	$0.31	$0.04	$9.62	$0.05	$14.91	$0.06	$24.84	$0.11
At release	$0.35	$0.05	$9.97	$0.03	$15.13	$0.05	$25.45	$0.09
Release+1	$0.35	$0.05	$11.31	$0.12	$15.30	$0.05	$26.96	$0.16
Release+4	$0.35	$0.05	$11.31	$0.12	$17.08	$0.21	$28.75	$0.33
Release+8	$0.35	$0.05	$11.31	$0.12	$17.59	$0.27	$29.26	$0.40
No Piracy	$0.35	$0.05	$11.31	$0.12	$17.90	$0.31	$29.57	$0.43

Table 8. Effects of box-office revenue (in billions of US$) summed over all movies under various scenarios. For example, "Release-4" means that piracy occurs 4 weeks before release, while "Release+4" means that piracy occurs 4 weeks after release.

Comparing our predicted box-office revenue with no-piracy ($29.57b) to our predicted revenue of $25.66b from Table 8, we see that our model predicts a $3.91b reduction in box-office revenue due to piracy. Our results occur over a 35-month period, so we can annualize these effects to estimate that piracy is a $1.34b problem for the first-run theater business in the U.S. Said another way, if piracy could be eliminated then box-office revenue would increase by 15%. Piracy currently occurs at any point during the release or perhaps never. If we use the benchmark of piracy always happening at release which is more typical of what currently happens, then box office-revenue would be $25.45b. This reference point suggests that if piracy always occurred at release, box-office revenue would decrease by $4.12b versus current revenue where piracy typically occurs after release.

Figure 8 illustrates the results in Table 8 using a scatterplot of the total predicted U.S. box-office across all movies in our database versus when the piracy occurs. The dotted line at the top of this figure illustrates what box-office revenue would be if no piracy occurs, while the dashed line near the middle illustrates actual sales. What is clear from this figure is that any form of piracy is bad, and that the longer piracy can be delayed the less the loss of revenue. A more nuanced perspective of piracy is that it is asymmetric around piracy occurring at release. In other words, the gains of delaying post-release piracy are larger than the losses attributed to pre-release piracy occurring. For example, the predicted loss of revenue from piracy occurring one week early relative to at release is a $0.61b loss, versus the gain of delaying piracy by one week which results in a $1.51b gain relative to piracy occurring at release.

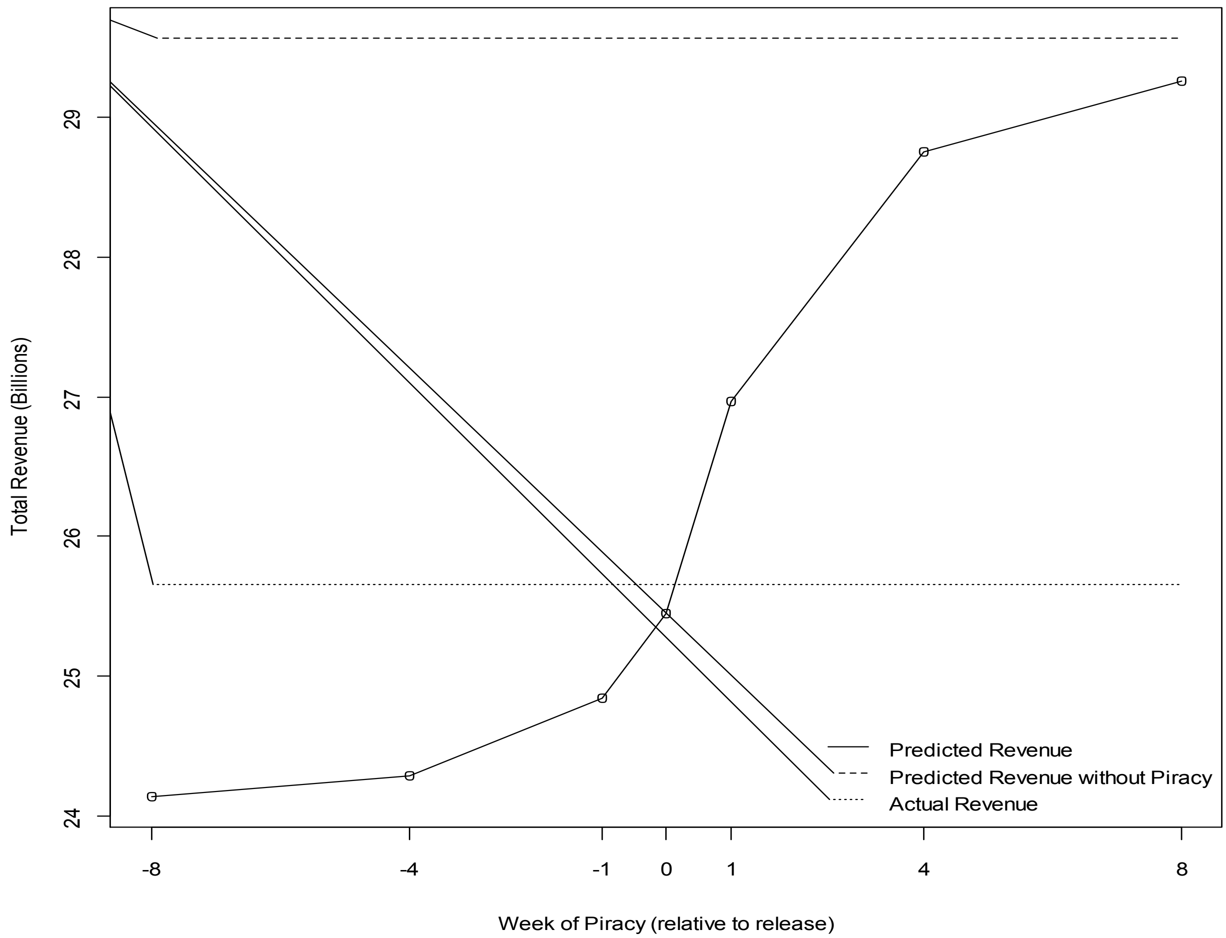

Figure 8. Total predicted box office revenue versus the week piracy occurs from a counterfactual simulation in which all movies have piracy that occurs at the specified week relative to the release week.

We can further decompose the losses associated with piracy depending upon when the revenues are earned during the release cycle. Specifically, if we compare no-piracy to piracy at release we find a $0.13b reduction in sales during pre-release, a $1.18b reduction in revenue during the opening week, and a $2.59b reduction in sales the post-release period. (Note that in our terminology the release week is the first week of wide release and thus lost pre-release revenue comes from sales during limited or staged releases of movies such as for a premiere, a specific city, or region.) Although pre-release sales are a small proportion of overall revenue, they would experience a 61% increase if piracy could be eliminated versus a 12% and 17% increase for launch week and post-launch periods, respectively. Hence, pre-release sales are especially vulnerable to piracy.

5.2 Heterogeneity in movie-level effects of Pre-Release Piracy

Pre-release piracy has become a more pressing issue given prominent pre-release piracy leaks such as *The Expendables 3* and *X-Men Wolverine*. Our results allow us to simulate the impact of pre-release piracy. These simulations show that, across all 533 movies in our data, piracy that occurs eight weeks prior to release causes a mean reduction in total box-office revenue of $1.31b, or about 5% of box-office revenue, compared with piracy that occurs at release. Almost half of the loss, or $0.63b, comes in the opening week. These results are helpful in understanding aggregate response, but we can examine these simulations at the individual movie level to gain further insight into the effect of pre-release piracy.

To do this, in Figure 9 we report the histogram of the effects of piracy across movies. In this case we calculate the percentage loss in box-office revenue from piracy occurring eight-weeks prior to release versus piracy occurring at release. The mean percentage loss in box-office revenue across movies is 3.7%, the median loss is 2.6%, and the 25th and 75th percentiles of losses are 1.3% and 4.3%, respectively. One notable characteristic of Figure 9 is that a distinct minority of movies (15 out of 533 movies in our data or about 3% of our movies) show an increase in box-office revenue if piracy occurs eight-weeks prior to release versus if piracy occurs at release. However, the average percentage gain in revenue is only 0.5% increase over piracy occurring at release, with only two movies showing a gain of more than 1% (specifically a gain of 2.1% and 1.5%).

To be clear, this does not mean that piracy is good for these movies: all of the movies in our counterfactual analysis would experience increased box-office revenue if piracy were eliminated altogether. Rather, this result suggests that for a small number of movies, pre-release piracy might have a promotional impact relative to piracy that only occurs after release. We also note that from a promotional standpoint this effect is small, but it leads to an interesting question of why pre-release piracy increases box-office revenue compared with piracy at release? We believe answering this question leads to greater insight into understanding the cannibalization versus promotional effects of piracy. We discuss this question further in the next subsection.

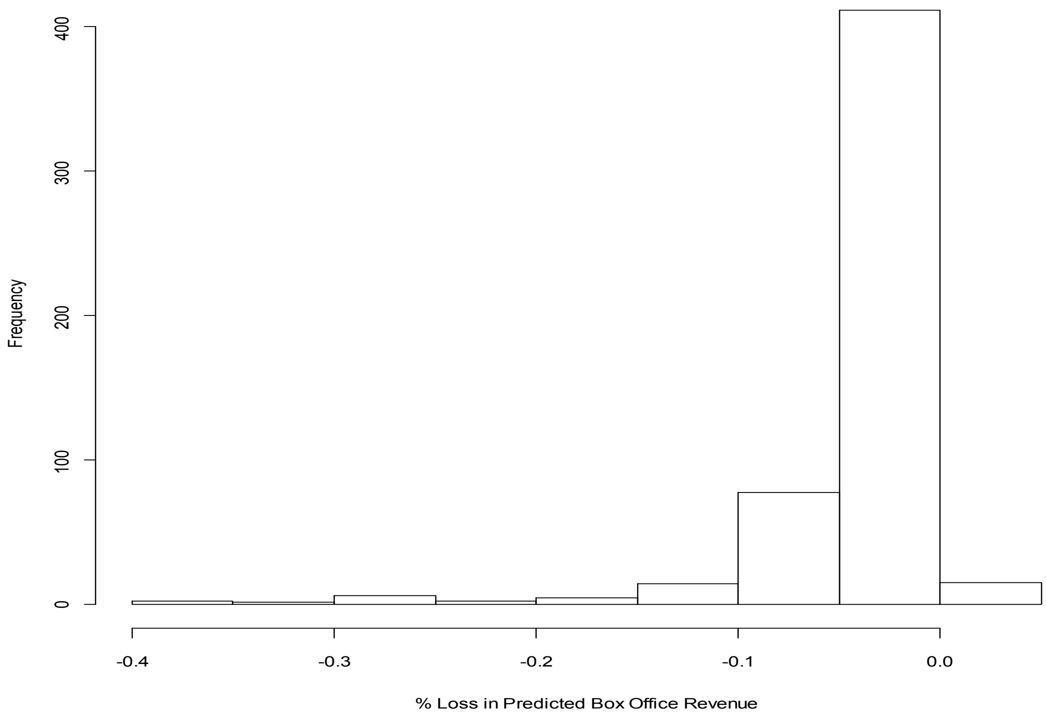

Figure 9. Percentage loss in predicted box-office from pre-release piracy occurring at eight weeks prior to release versus to at release. All losses of more than 40% are included in the left most bar due to the long-tail of the distribution.

5.3 Decomposing Piracy into Cannibalization and Promotional Effects

The finding that pre-release piracy could help some movies compared with piracy at release may seem counterintuitive, but our model provides a rationale. There are two competing effects of piracy. The first is a negative effect due to cannibalization which leads those who watch pirated copies to not go to the movie theater, thus reducing box-office revenue. The second is a positive effect[10] which comes from consumers who watch pirated copies conveying positive word-of-mouth to their friends. Our discussion earlier shows that overall the cannibalization effect outweighs the promotional effect. However, if the secondary positive effect of word-of-mouth is strong enough, then for particular movies it may offset piracy's

[10] Our results suggest that there is more positive than negative word-of-mouth, hence this effect, on balance, is positive. However, our model does not require this to be true: for low quality movies, our results allow word-of-mouth effects to reduce the likelihood of attendance. This factor suggests that piracy could be more harmful for low-quality movies than for high-quality movies.

primary negative effect of cannibalization. This explains why in certain circumstances pre-release piracy could increase box-office revenue over piracy that occurs at release.[11] But this possibility motivates a more general thought exercise to understand the relative magnitude of each of the effects of piracy. Specifically, in this subsection we are interested in what would happen to box-office revenue if piracy only had a negative, cannibalization effect but not a positive, promotional effect?

We can answer this counterfactual exercise directly from the model by removing the flows from those who watch pirated copies to positive and negative word-of-mouth states and redirecting these flows to the inactive state. In other words, we can use our parameters to analyze a counterfactual world where those individuals who watch pirated copies move directly to an inactive state and do not spread any word-of-mouth. Notice that we are not re-estimating the model, but using the parameters estimated in §4 and re-specifying the flows in §3 such that flows from those who view piracy go directly to the inactive state and not to the spreader states as in Figures 2a and 2b. Specifically, our previous transition matrix in (4) becomes:

$$\mathbf{A}_{it} = \begin{bmatrix} 1-\mathcal{E}_{it} & (1-\mathcal{V}_{it})(1-\phi_i) & 0 & 0 & 0 & 0 \\ \mathcal{E}_{it}\mathcal{L}_{it}(1-\mathcal{V}_{it}) & (1-\mathcal{V}_{it})\phi_i & 0 & 0 & 0 & 0 \\ \mathcal{E}_{it}\mathcal{L}_{it}\mathcal{V}_{it}^{M}\rho_i & \mathcal{V}_{it}^{M}\rho_i & \mu_i & 0 & 0 & 0 \\ \mathcal{E}_{it}\mathcal{L}_{it}\mathcal{V}_{it}^{M}(1-\rho_i) & \mathcal{V}_{it}^{M}(1-\rho_i) & 0 & \mu_i & 0 & 0 \\ \mathcal{E}_{it}\mathcal{L}_{it}\mathcal{V}_{it}^{P} & \mathcal{V}_{it}^{P} & 1-\mu_i & 1-\mu_i & 1 & 0 \\ \mathcal{E}_{it}(1-\mathcal{L}_{it}) & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{15}$$

Where $\mathcal{V}_{it} = \mathcal{V}_{it}^{M} + \mathcal{V}_{it}^{P}$, $\mathcal{V}_{it}^{M} = \left(1-\delta_{it}^{*}\pi_i\right)\tau_i\gamma_{it}$ and $\mathcal{V}_{it}^{P} = \delta_{it}^{*}\pi_i\tau_i$. In other words, viewing effects are separated into effects from watching at the movie theater (*M*) and pirated copies (*P*). Our goal is to consider a world that we do not observe, namely one in which those individuals who watch pirated copies never pass along information about the movies that they have watched. We are not asking the question of whether the data supports whether piracy has an impact on sales, since the data through its estimates of the

[11] The type of movie where we might expect to see the beneficial, secondary effect of pre-release piracy on box-office sales is when a movie has higher quality than expected by the market, word-of-mouth effects are strong and persistent, the cannibalization effect of piracy is relatively small, and advertising spending low.

cannibalization effect shows that piracy has an impact (e.g., $\pi_i > 0$). Rather, we are trying to assess how much worse piracy would be without its positive, promotional aspect.

Scenario	*Predicted Box-office*	*Std Err*	*Percentage Increase*	*Predicted Box-office (Cannibalization only)*	*Std Err*	*Percentage Increase*
Release-8	\$24.14	\$0.17	22%	\$23.47	\$0.24	26%
Release-4	\$24.29	\$0.16	22%	\$23.59	\$0.23	25%
Release-1	\$24.84	\$0.11	19%	\$24.29	\$0.17	22%
At release	\$25.45	\$0.09	16%	\$25.06	\$0.11	18%
Release+1	\$26.96	\$0.16	10%	\$26.73	\$0.14	11%
Release+4	\$28.75	\$0.33	3%	\$28.67	\$0.33	3%
Release+8	\$29.26	\$0.40	1%	\$29.24	\$0.39	1%
No Piracy	\$29.57	\$0.43		\$29.57	\$0.43	

Table 9. Effects of box-office revenue (in billions of US\$) summed over all movies under various scenarios. The percentage increase represents how much of a gain would occur from this scenario if piracy could be eliminated entirely.

The results of our counterfactual world in which piracy no longer has a promotional word-of-mouth effect but only a cannibalization effect are reported in Table 9 and Figure 10. In this analysis we use a benchmark of the predicted box-office from our previous counterfactual, which predicts what would happen if piracy were to occur across all movies at the specified release week. For example, if piracy occurs at release—which is a close to what currently happens—then we would expect box-office revenue to be \$25.45b. However, if piracy only has a cannibalization effect then this would drop to \$25.06b. In other words, the promotional gain of piracy helps to offset the cannibalization effect of piracy by \$0.39b, although this is a small offset compared with the potential loss of \$4.51b if piracy only had a cannibalization effect[12].

We can see in Figure 10 that the promotional gains of piracy are most critical for pre-release piracy. For example, notice that box-office revenue is offset by promotional effects by \$0.67b when piracy occurs 8 weeks before launch, while it only represents a \$0.02b offset when piracy occurs 8 weeks after launch. The intuition is that if piracy occurs early then word-of-mouth effects can diffuse over a longer period, which

[12] This \$4.51b loss is the reduction in predicted box-office revenue with no piracy of \$29.57 to \$25.06b if piracy occurs at release.

would help offset more of the cannibalization effects of piracy, versus piracy that occurs after release where most consumers have already discovered the movie through other promotional channels.

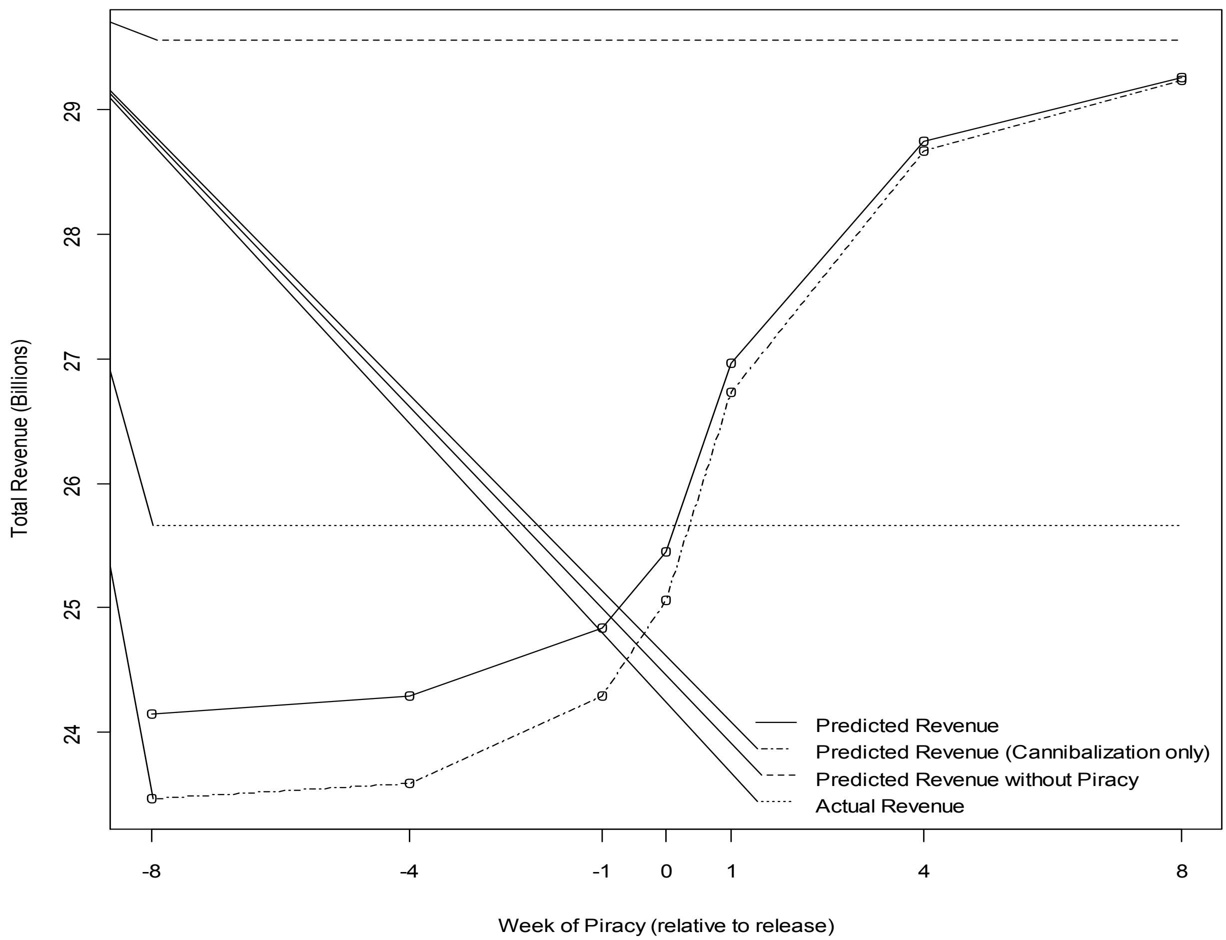

Figure 10. Total predicted box-office revenue versus week of piracy when piracy only has a cannibalization effect. This plot considers the counterfactual of predicted revenue when piracy does not have a promotional effect, but instead those who watch pirated copies do not spread any word-of-mouth.

5.4 Comparison with previous estimates of pre-release piracy

Our estimate of pre-release piracy indicates that pre-release piracy would reduce box office revenue by 5% compared with piracy that occurs at release. This estimate is consistent with—but lower than—estimates in Ma et al. (2014). Ma et al. (2014) use a reduced form approach, and found that pre-release piracy reduces box-office revenue by 19% relative to piracy that occurs at- or after-release. An important difference between these two approaches is that the estimates of Ma et al. (2014) use a different benchmark. Our benchmark for the effects of piracy is to compare predicted box-office revenue with what would have happened if all piracy occurs at release. In contrast, Ma et al (2014) compare revenue to when piracy actually occurred with those movies that did not experience pre-release piracy. In our dataset there were 474 movies that were pirated at- or post-release with the average week of piracy being 2.4. Additionally there are 7 movies that had no piracy. Hence a more fair comparison would be to predict the loss in box-office revenue compared to piracy occurring sometime between one week after-release and four weeks after-release. According to Table 8, this comparison would estimate between a 10% and 16% reduction in box-office revenue[13]. Secondly, we point out that our predictions are conditional on advertising spending and screen availability, while Ma et al. (2014) estimates are marginalized over these values. If advertising spending or screen availability is reduced in the presence of piracy then this could also account for our lower estimate. In summary, we believe that our estimates are generally consistent with those in Ma et al. (2014), but the approach in this paper comes with added benefits with respect to counterfactual simulations.

6. Trends in Piracy

Potentially piracy has changed from the 2006-2008 time period that has formed the focus of this study. To understand how piracy has changed we replicate our model for models over the 2011-2013 time period. We collect a similar data as before. First, we identify the wide releases from BoxOfficeMojo during the period of January 2011 through December 2013. We obtained weekly advertising data from Kantar Research. The Nielsen Research box-office data is not available to use for this time period, so we have instead used the box-office revenue and the number of screens as reported by The Numbers website which is maintained by Nash Information Services. Finally, we have movie piracy data from vcdquality.com and a second source of data about piracy from the Motion Picture Association of America (MPAA). We combine these two sources of piracy to identify the first data at which a pirated version of a movie became available anywhere in the world. BoxOfficeMojo identifies a total of 441 wide releases during this three-year period with a total of $30.8b in US box-office revenue. There is missing information for 143 movies so we have

[13] If piracy occurs one week after release then box-office revenue would be $26.96 compared with $24.14 for piracy at eight weeks before release, or a reduction of 10%. If piracy occurs fours week after release then box-office revenue would be $28.75 or a reduction of 16% compared with pre-release piracy occurring eight weeks before release. Our table does not provide a prediction of piracy occurs 2.4 weeks after release but we can interpolate that it falls in this interval.

removed these titles from the analysis, which leaves 298 movies with a total box office of $23.0b (or roughly three quarters of the box office for this period).

Variable	Mean	Std. Dev.	Minimum	Median	Maximum
Box-office Revenue ($ million)	$77.3	$85.8	$1.4	$49.1	$623.3
Advertising ($ millions)	$22.7	$12.1	$0.0	$21.3	$62.0
Opening Screens (at wide-release)	2571	1028	249	2870	4404
Week pirated version available	1.6	5.8	-34.0	0.0	20.0

Table 10. Descriptive Statistics for Movie Characteristics during 2011-2013. There are 298 movies in our dataset that have wide releases between Jan 2011 and Dec 2013. The week for the first week a pirated version is available is expressed relative to the wide release (e.g., 0 represents at release, -1 represents pre-release piracy since the copy was available 2 weeks before release, 1 represents post-release piracy since the copy was available 2 weeks after release). In this dataset all movies show piracy during the US theatrical window.

For comparison the parameter estimates of the movies from 2011-2013 are given in Table 11. Qualitatively there is a high correlation with those given earlier estimates as given in Table 5, although the newer set of movies shows less variability. A focal parameter of our interest is the coefficient on movie piracy, which shows a slight decrease on our logistic transformed scale. To better illustrate the piracy parameter (in its natural units) we plot the histogram across movies in Figure 11. The mean is 7.8% which is a decrease from the 10.0% found previously..

Parameter	*Transform Mean*	*Transform Std. Dev.*	*Mean*	*Std. Dev.*	*Lower*	*Upper*
ν_i	-4.49	0.41	0.017	0.030	0.006	0.061
υ_i	-7.87	0.26	0.0004	0.0001	0.0002	0.0008
λ_i	-1.42	0.27	0.212	0.049	0.152	0.363
τ_i	0.15	0.42	0.531	0.090	0.358	0.712
ϕ_i	1.32	0.46	0.761	0.076	0.589	0.881
μ_i	2.13	0.53	0.849	0.068	0.705	0.933
π_i	-3.31	0.52	0.078	0.051	0.033	0.239
θ_i	0.37	0.55	0.572	0.112	0.346	0.777
ρ_i	1.51	0.53	0.724	0.087	0.494	0.848
σ_i	-10.94	1.07	0.00006	0.00009	0.00000	0.00027

Table 11. Parameter Estimates for Movies during 2011-2013. Distribution of mean of parameter estimates across all movies along with standard deviations. Both the transformed and inverse of the parameters are given. Lower and Upper values correspond with the 2.5% and 97.5% percentiles.

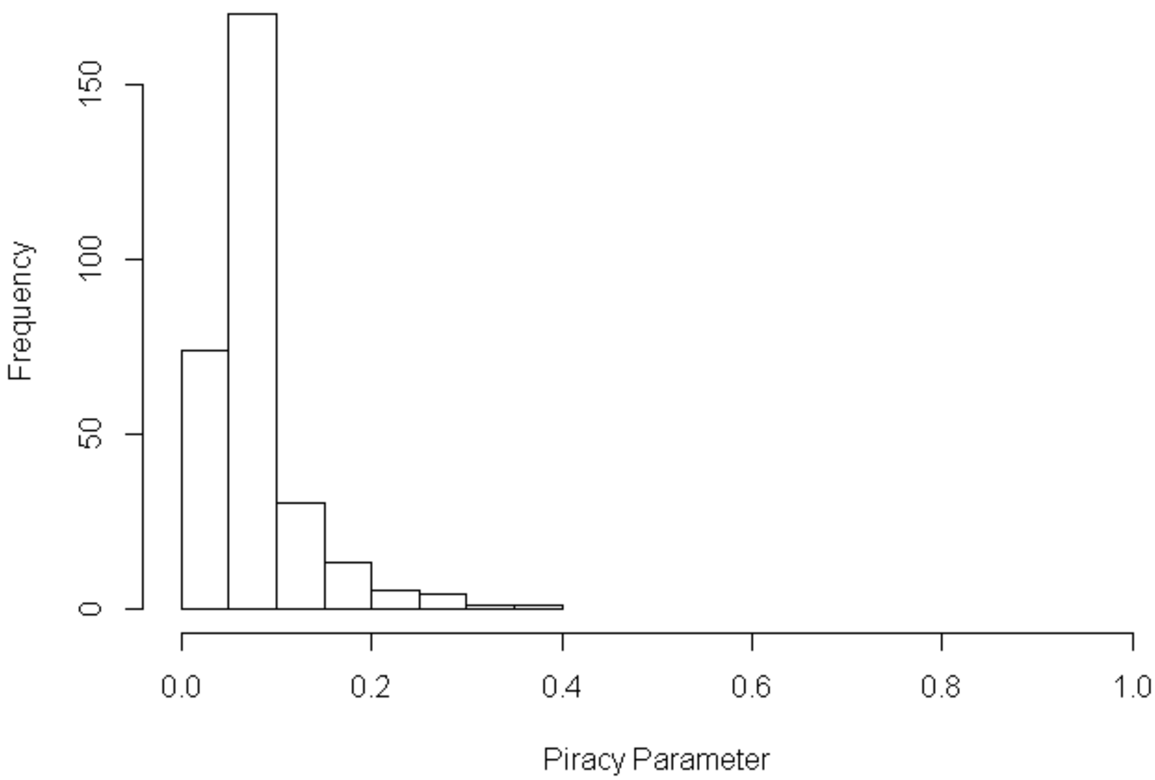

Figure 11. Histogram of the piracy parameter estimates across all movies during 2011-2013.

To more fully assess the effects of piracy we consider the counterfactual exercise that we completed in §5.1. As before we consider the effects of piracy under a number of simulated scenarios which range from no piracy occurring at all to every movie experience pre-release piracy eight weeks before wide release at the box office. These estimates are given in Table 12. If piracy could be eliminated entirely from the theatrical release window in our 2011-2013 time span we predict that revenue would be 14% greater than piracy occurring at release. This is less than the 16% we estimated for the 2006-2008 period. If we consider the potential effect of pre-release piracy, so piracy that occurs one week before release, then we revenues would increase by 18% during our newer time period compared with the 19% predicted during our older time period. Substantively our conclusion is that the effects of piracy are similar in both the earlier and later time periods. Although there is some evidence for a marginal decrease in piracy's relative impact if one accounts for uncertainty in the parameter estimates as well as potential uncertainty in the model it may be difficult to assign a statistical significance to these changes.

Scenario	*Predicted % Increase in Box-Office Revenue during 2006-2008*	*Predicted % Increase in Box-Office Revenue during 2011-2013*
Release-8	22%	21%
Release-4	22%	21%
Release-1	19%	18%
At release	16%	14%
Release+1	10%	8%
Release+4	3%	2%
Release+8	1%	1%

Table 12. Counterfactual Simulation of Piracy Scenarios during 2011-2013. Effects of box-office revenue (as a percentage change from no-piracy occuring) summed over all movies under various scenarios. For example, "Release-4" means that piracy occurs 4 weeks before release, while "Release+4" means that piracy occurs 4 weeks after release.

7. Discussion

Studios have long argued that pre-release piracy is particularly harmful for box-office revenue because it occurs at a point in time where there is no legitimate alternative channel. And, from a policy perspective, the 2005 Family Entertainment Copyright Act in the United States made pre-release distribution of movies a felony offense with up to 5 years of imprisonment for first time offenders. On the other side of the spectrum, some have argued that pre-release piracy might have a promotional effect on box-office revenue by increasing word-of-mouth buzz for the movie. For example, after a copy of the movie "Soul Plane" was leaked prior to its theatrical release, one of its stars said "I don't think a bootleg is going to stop anything…a bootleg is like buzz."[14] After empirically examining both the cannibalization and promotion effects of piracy we find clear support that, for the vast majority of movies, the negative, cannibalization effects far outweigh the positive, promotional effects.

Understanding the impact of piracy on media products has important implications for managers (particularly in the context of revenue forecasts and advertising plans for movie releases) and for policymakers (particularly in the context of balancing the benefits and costs of potential policy interventions).

[14] http://www.blackfilm.com/20040521/features/snoopdogg.shtml

However, while piracy has been studied widely in the academic literature, the literature has not simulated the counterfactual of box-office sales that would occur in the absence of piracy, nor has it disentangled the hypothesized promotional effect of piracy from the cannibalization effect of piracy. In this paper we address these two limitations of the previous literature by adapting the established MOVIEMOD model (Eliashberg et al 2000) to incorporate the impact of piracy on sales and to allow the use of historical data in estimating consumer adoption decisions for movies.

Our results show that piracy lowers box-office revenue by $1.34b per year (15%) as compared to a hypothetical world where piracy was eliminated during the theatrical window. This percentage appears to be consistent between 2006-2009 and 2011-2013. This result provides empirical justification for policy approaches providing strong penalties for consumers guilty of facilitating piracy. However, our results also suggest that for a small number of movies (about 3% of our sample), pre-release piracy may increase sales relative to a world where piracy coincides with release. We wish to emphasize that this does not mean that piracy is better than no piracy — all of the movies in our sample would have higher sales if there were no piracy. Rather, this result suggests that in a small number of cases, earlier piracy can be better than later piracy to the extent that it can increases awareness of a movie early. In other words the promotional lift associated with piracy helps lessen the effects of cannibalization, but does not offset it.

Our work contributes to the literature in several ways. First, it fills a gap in the literature by presenting evidence of the impact of Internet-based movie piracy on important marketing questions regarding box-office revenue forecasts. Second, by directly controlling for advertising, early release, and the timing of piracy, our model addresses several factors that complicate the analysis in most existing studies of piracy, helping us draw causal inference in a more convincing way. Finally, our model allows us to separate the promotional and cannibalizing impacts of piracy, providing the first empirical paper that we are aware of that—for a small number of titles—early piracy can be better than later piracy because of increased word-of-mouth promotional effects.

In general our results show deliberate leaks by studios would not be profitable, but for a few selected movies we show that the promotional effects of piracy could profitably overcome the cannibalization effect when comparing pre-release piracy to pirated copies becoming available at release. We note that for this promotional effect to hold two things must be true. First, the quality of the movie must be high, so that those that watch the pirated version are likely to recommend watching the movie, and second word-of-mouth intensity must also be high, so that those that watch the pirated version actively promote the movie to others. If either factor is missing, the net effect of piracy is negative. Essentially pre-release piracy in this limited context serves as a form of promotion that encourages consumers to watch a movie that they would have not have known about.

We also note several additional major limitations of our study. The first derives from the data. Although we know when piracy has occurred, we do not know the number or quality of the pirated downloads. Having that information could further strengthen the causal inference on the impact of piracy. Secondly, our model is based on historical data. Although our model is quite accurate, it uses all information available. It is well suited to answering post hoc questions about how piracy impacts movies and can be used to set general strategic decisions. But it is an ex ante promotional planning tool. If managers want to set to plan out distribution or promotional budgets in the light of piracy occurring at some predicted point in time then these decisions require knowledge of the parameter estimates before the release of the movie. Being able to reliably predict the parameter estimates before release would be critical to these decisions which we are unable to do. To do this new information that is available pre-release needs to be used. This information could include consumer surveys about their liking of the movie (Eliashberg et al. 2000), movie attributes and critic reviews (Lee et al 2003, Dellarocas et al 2007), pre-release media publicity and advertising (Wang et al 2013), keyword search (Kulkarni, Kannan and Moe 2012) and blogs (Godes and Mayzlin 2004, Liu 2006, Kim and Hanssens 2014, Xiong and Bharadway 2014), stock prices (Joshi and Hanssens 2009), and viewings of movie trailers (Swait and Shugan 2000). Given that the intent of the original MOVIEMOD paper was to be a pre-release planning tool, we believe that our proposed methodology could be readily adapted to combine these new information sources and would be a fruitful direction for new research. Third we have assumed that consumers who pirate spread word-of-mouth similarly to those who do not, but potentially they may be driven by different considerations (Sinha and Mandel 2008). Fourth we have only considered one product category—movies, potentially episodic content like television programs or music recordings where ticket sales from tours would benefit to a greater extent from piracy's promotional effect. Finally we consider only box-office revenue and not subsequent revenue from home video and other licensing. These limitations point to directions for future research on this important topic.

References

Ainslie, Andrew, Xavier Dreze, and Fred Zufryden (2005), "Modeling Movie Life Cycles and Market Share", *Marketing Science*, 24 (3), 508-17.

Bai, J., J. Waldfogel (2012), "Movie Piracy and Sales Displacement in Two Samples of Chinese Consumers," *Information Economics and Policy*, 24(3) 187-196.

Barr, Merrill (2015), "'Supergirl' Pilot Leak: Theft or CBS Marketing Ploy?", *Forbes*, http://www.forbes.com/sites/merrillbarr/2015/05/22/supergirl-pilot-leak/.

Bass, Frank M. (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 15 (January), 215-27.

Bounie, D., M. Bourreau, P. Waelbroeck (2006), "Piracy and the Demand for Films: Analysis of Piracy Behavior in French Universities," *Review of Economic Research on Copyright Issues*, 3(2) 15-27.

Chellappa, R.K., S. Shivendu (2005), "Managing piracy: pricing and sampling strategies for digital experience goods in vertically segmented markets", *Information Systems Research*, 16(4) 400-17.

Chen, Yubo, Yong Liu, and Jurui Zhang (2012), "When Do Third-party Product Reviews Affect Firm Value and What Can Firms Do? The Case of Media Critics and Professional Movie Reviews," *Journal of Marketing*, 76 (2): 116-134.

Cheng, H.K., R. Sims, H. Teegen (1996), "Toward a Profile of Student Software Pirates," *Journal of Business Ethics*, 8, 839-851.

Conner, Kathleen Reavis and Richard Rumelt (1991), "Software Piracy: An Analysis of Protection Strategies", *Management Science*, 37 (2); 125-139.

Conlisk, J. (1976), "Interactive Markov chains", *Journal of Mathematical Sociology*, 4, 2, 157-185.

Crossley, Rob (2009), "Trials HD studio leaked game to torrent sites", http://www.develop-online.net/news/trials-hd-studio-leaked-game-to-torrent-sites/0105690.

Danaher, Brett, Samita Dhanasobhon, Michael D. Smith, and Rahul Telang (2010), "Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy", *Marketing Science*, 29 (6), 1138-1151.

Danaher, Brett, Michael D. Smith, Rahul Telang (2014), "Piracy and Copyright Enforcement Mechanisms", Lerner and Stern, eds. *Innovation Policy and the Economy*, Volume 14, University of Chicago Press, Chicago, Illinois.

de Moraes, Lisa (2015), "CBS' 'Supergirl' Latest Superhero Project to Leak Online", http://deadline.com/2015/05/supergirl-leak-online-piracy-cbs-dc-comics-video-1201431621/.

De Vany, A.S., W.D. Walls (2010), "Estimating the Effects of Movie Piracy on Box-office Revenue", *Review of Industrial Organization* 30:291-301.

Dellarocas, Chrysanthos, Xiaoquan (Michael) Zhang, and Neveen F. Awad (2007), "Exploring the Value of Online Product Review in Forecasting Sales: The Case of Motion Pictures", *Journal of Interactive Marketing*, 21 (4), 23-45.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008), "Do online reviews matter? – An empirical investigation of panel data", *Decision Support Systems*, 45, 4 (November), 1007-1016

Elberse, Anita (2007), "The Power of Stars: Do Star Actors Drive the Success of Movies?", *Journal of Marketing*, 71, 4 (October), 102-120.

Elberse, Anita and Jehoshua Eliashberg (2003), "Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures", *Marketing Science*, 22, 3 (Summer), 329-354.

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

Eliashberg, Jehoshua, Samuel K. Hui, and Z. John Zhang (2007), "From Story Line to Box-office: A New Approach for Green-Lighting Movie Scripts", *Management Science*, 53, 6 (June), 1097-1110.

Eliashberg, Jehoshua, Jedid-Jah Jonker, Mohanbir S. Sawhney, and Berend Wierenga (2000), "MOVIEMOD: An Implementable Decision Support System for Pre-Release Market Evaluation of Motion Pictures", *Marketing Science*, 19, 3 (Summer), 226-243.
Eliashberg, Jehoshua and Steven M. Shugan (1997), "Film Critics: Influencers or Predictors?", *Journal of Marketing*, 61, 2 (April), 68-78.
Givon, Moshe, Vijay Mahajan, and Eitan Muller (1995), "Software Piracy: Estimation of Lost Sales and the Impact on Software Diffusion", *Journal of Marketing*, 59, 1 (January), 29-37.
Godes, David and Dina Mayzlin (2004), "Using online conversations to study word of mouth communication", *Marketing Science*, 23 (4) 545-560.
Hennig-Thurau, T., V. Henning, H. Sattler, (2007), "Consumer File Sharing of Motion Pictures", *Journal of Marketing*. 71(October) 1-18.
Jain, Sanjay, (2008) "Digital piracy: A competitive analysis," *Marketing Science* 27.4: 610-626.
Joshi, Amit M. and Dominique M. Hanssens (2009), "Movie Advertising and Stock Market Valuation of Studios: A Case of 'Great Expectations?'", *Marketing Science*, 28 (2): 239-250.
Kim, Ho and Dominique M. Hanssens (2014), "Paid and Earned Media, Consumer Interest and Motion Picture Revenue", UCLA Working Paper.
Kulkarni, Gauri, P.K. Kannan, and Wendy Moe (2012), "Using online search data to forecast new product sales", *Decision Support Systems*, 52 (3), 604-611.
Krider, Robert E. and Charles B. Weinberg (1998), "Competitive Dynamics and the Introduction of New Products: The Motion Picture Timing Game", *Journal of Marketing Research*, 35, 1 (February), 1-15.
Lee, Jonathan, Peter Boatwright, and Wagner A. Kamakura (2003), "A Bayesian Model for Prelaunch Sales Forecasting of Recorded Music", *Management Science*, 49 (2): 179-196.
Liebowtiz, Stan J. (2006), "File Sharing: Creative Destruction of Just Plain Destruction?", *Journal of Law and Economics*, 49:1 (April) 1-28.
Liu, Yong (2006), "Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (3): 74-89
Liu, Y., H.K. Cheng, Q.C. Tang, E. Eryarsoy (2011), "Optimal software pricing in the presence of piracy and word-of-mouth effect", *Decision Support Systems*, 99-107.
Ma, Liye, Alan L. Montgomery, Param Vir Singh, and Michael D. Smith (2014), "An Empirical Analysis of the Impact of Movie Piracy on Box Office Revenue", *Information Systems Research*, 25 (3): 590-603.
Mahajan, V., E. Muller, R.A. Kerin (1984), "Introduction strategy for new products with positive and negative word-of-mouth", *Management Science*, 30, 1389-1404.
Mick, Jason (2009), "'X-Men Origins: Wolverine' Leaks, Some Speculate Leak Was Intentional", http://www.dailytech.com/XMen+Origins+Wolverine+Leaks+Some+Speculate+Leak+Was+Intentional/article14747.htm
Moraski, Lauren (2013), "What's the story behind song leaks?", CBS News, http://www.cbsnews.com/news/whats-the-story-behind-song-leaks/.
Peitz, M., P. Waelbroeck, (2006), "Why the music industry may gain from free downloading — the role of sampling", *International Journal of Industrial Organization* 24 (5).
Prag, Jay and James Casavant (1994), "An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry", *Journal of Cultural Economics*, 18, 3, 217-235.
Prasad, A and V. Mahajan. (2003), "How many pirates should a software firm tolerate? An analysis of piracy protection on the diffusion of software", International Journal of Research in Marketing, 20, 4, 337-353.
Rob, R., J. Waldfogel (2007), "Piracy on the Silver Screen", *Journal of Industrial Economics*, 55(3) 379-393.
Sandoval, Greg (2010), "Studio didn't report 'Potter' leak to feds", http://www.cnet.com/news/studio-didnt-report-potter-leak-to-feds/.
Sinha, Rajiv K., Fernando S. Machado, and Collin Sellman, (2010), "Don't think twice, it's all right: Music piracy and pricing in a DRM-free environment," *Journal of Marketing* 74.2: 40-54.
Sinha, Rajiv K., and Naomi Mandel, (2008), "Preventing digital music piracy: the carrot or the stick?" *Journal of Marketing* 72.1: 1-15.
Smith, Michael D. and Rahul Telang (2010), "Piracy or promotion? The impact of broadband Internet penetration on DVD sales", *Information Economics and Policy*, 22 (4) 289-298.

Swait, Joffre and Steven Shugan (2000), "Enabling Movie Design and Cumulative Box Office Predictions", *2000 ARF Conference Proceedings*.
Swami, S., Jehoshua Eliashberg, Charles B. Weinberg (1999), "SilverScreener: A Modeling Approach to Movie Screens Management", Marketing Science, 18, 3, 352-372.
Vernik, Dinah A., Devavrat Purohit, and Preyas S. Desai (2011), "Music downloads and the flip side of digital rights management," *Marketing Science* 30.6: 1011-1027.
Wallenstein, Andrew (2010) "'Harry Potter' Leak Seems More Like Publicity Than Piracy", https://gigaom.com/2010/11/16/419-harry-potter-leak-seems-more-like-publicity-than-piracy/
Wang, Feng, Yin Zhang, Xiaoling Li, and Huawei Zhu (2010), "Why Do Moviegoers Go to the Theater? The Role of Prerelease Media Publicity and Online Word of Mouth in Driving Moviegoing Behavior", *Journal of Interactive Advertising*, 11 (1): 50-62.
Waters, J. (2013), "Variable marginal propensities to pirate and the diffusion of computer software", Working Paper, Nottingham University Business School Research Paper No. 2013-06. (available from http://ssrn.com/abstract=2345431, last accessed September 19, 2014.)
Xiong, Guiyang and Sundar Bharadwaj (2014), "Prerelease Buzz Evolution Patterns and New Product Performance", *Marketing Science*, 33 (3): 401-421.
Zufryden, F. (1996), "Linking Advertising to Box-office Performance of New Film Releases-a Marketing Planning Model", *Journal of Advertising Research*, 36, 4, 29-41.

Appendix

A1. Monte Carlo Markov Chain Algorithm

To compute the posterior distribution of the parameters given the data we employ a Monte-Carlo Markov Chain (MCMC) approach. The idea is to sample a series of conditional distributions to yield samples of our target of interest—the posterior distribution. These samples can then be used to estimate any value of interest, such as the mean and standard deviation of the parameters, or to construct the predictive distribution of a new movie. Using the hierarchical structure and conditional independence of our model we sequentially sample from the following conditional distributions:

$$\boldsymbol{\beta}_i \mid \mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\Lambda}, \boldsymbol{\Omega} \tag{A4}$$

$$\boldsymbol{\Lambda} \mid \{\mathbf{x}_i\}, \{\boldsymbol{\beta}_i\}, \boldsymbol{\Omega} \tag{A5}$$

$$\boldsymbol{\Omega} \mid \{\mathbf{x}_i\}, \{\boldsymbol{\beta}_i\}, \boldsymbol{\Lambda} \tag{A6}$$

where z_{ij} is an indicator value if the corresponding x_{ij} is missing. Our notation $[\mathbf{a}]_{-j}$ denotes the removal of the *i*th element from the vector $\mathbf{a}$, $[\mathbf{A}]_{ij}$ denotes the selection of the *i*th row and the *j*th column from the vector $\mathbf{A}$, and if *i* or *j* are negative then the corresponding row or column are removed. Additionally, $[\mathbf{A}]_{i\text{g}}$ denotes the *i*th row and $[\mathbf{A}]_{\text{g}j}$ denotes the *j*th column. For brevity, the prior parameters are suppressed in these conditional statements, since it is understood all prior parameters are known. We define the following matrices:

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_1' \\ \mathrm{M} \\ \boldsymbol{\beta}_n' \end{bmatrix}, \ \mathbf{X} = \begin{bmatrix} \mathbf{x}_1' \\ \mathrm{M} \\ \mathbf{x}_n' \end{bmatrix} \tag{A7}$$

The parameters of the box-office predictions do not have a natural conjugate form. Therefore, we employ a random-walk Metropolis-Hastings algorithm and cycle through each of the parameters given the others:

$$\beta_{ij} \mid \left[\boldsymbol{\beta}_i\right]_{-j}, \mathbf{y}_i, \mathbf{X}_i, \boldsymbol{\Lambda}, \boldsymbol{\Omega} \tag{A8}$$

Specifically, we perturb the *t*th sample with a random walk to generate a proposal draw: $\beta_{ij}{}' \sim N(\beta_{ij}^{(t)}, \varsigma_j^2)$.

The proposal draw is accepted with the following probability:

$$\beta_{ij}^{(t+1)} = \begin{cases} \beta_{ij}{}' & w.p.\ \min\left(1, \dfrac{f(\beta_{ij}{}'; \left[\boldsymbol{\beta}_i\right]_{-j})}{f(\beta_{ij}^{(t)}; \left[\boldsymbol{\beta}_i\right]_{-j})}\right), \\ \beta_{ij}^{(t)} & \textit{otherwise} \end{cases} \quad f(\beta_{ij}; \left[\boldsymbol{\beta}_i\right]_{-j}) = \left(\prod_{t=1}^{T_i} p(y_{it} \mid \beta_{ij}, \left[\boldsymbol{\beta}_i\right]_{-j})\right) p(\beta_{ij} \mid \left[\boldsymbol{\beta}_i\right]_{-j}) \tag{A9}$$

We choose the variance of the random walk (ς_j^2) such that the rejection rate is approximately 50%. The conditional distribution of $\beta_{ij} \mid \left[\boldsymbol{\beta}_i\right]_{-j}$ is normal since the hyper-distribution is conjugate:

$$\beta_{ij} \mid \left[\boldsymbol{\beta}_i\right]_{-j} \sim N\left(\bar{\beta}_{ij} + \boldsymbol{\Omega}_{j,-j} \boldsymbol{\Omega}_{-j}^{-1} (\left[\boldsymbol{\beta}_i\right]_{-j} - \left[\bar{\boldsymbol{\beta}}_i\right]_{-j}), \boldsymbol{\Omega}_{jj} - \boldsymbol{\Omega}_{j,-j} \boldsymbol{\Omega}_{-j,-j}^{-1} \boldsymbol{\Omega}_{-j,j})\right), \text{ where } \bar{\boldsymbol{\beta}}_i = \boldsymbol{\Lambda} \mathbf{x}_i \tag{A10}$$

Our vector of transformed parameters follows a multivariate regression, and we assume a natural conjugate prior: $\text{vec}(\bar{\mathbf{B}}) \sim N(\text{vec}(\bar{\mathbf{B}}), \mathbf{V}_\beta)$. As a consequence the conditional distribution is:

$$\text{vec}(\boldsymbol{\Lambda}) \mid \{\mathbf{x}_i\}, \{\boldsymbol{\beta}_i\}, \boldsymbol{\Omega} \sim N\left((\boldsymbol{\Omega}^{-1} \otimes \mathbf{X}'\mathbf{X} + \mathbf{V}_\beta^{-1})^{-1} (\text{vec}(\mathbf{X}'\mathbf{B}\boldsymbol{\Omega}^{-1}) + \mathbf{V}_\beta^{-1} \text{vec}(\bar{\mathbf{B}})), (\boldsymbol{\Omega}^{-1} \otimes \mathbf{X}'\mathbf{X} + \mathbf{V}_\beta^{-1})^{-1}\right) \tag{A11}$$

The covariance matrix of the multivariate regression is also assumed to have a conjugate inverted Wishart prior: $\boldsymbol{\Omega} \sim IW(\boldsymbol{\Psi}, \nu_{\boldsymbol{\Omega}})$. It follows that the conditional distribution is:

$$\boldsymbol{\Omega} \mid \mathbf{S} \sim IW(\mathbf{S} + \boldsymbol{\Psi}, n + \nu_{\boldsymbol{\Omega}}), \text{ where } \mathbf{S} = \mathbf{E}'\mathbf{E},\ \mathbf{E} = \mathbf{B} - \mathbf{X}\boldsymbol{\Lambda} \tag{A12}$$

Table of Variables and Parameters

Type	*Variable*	*Definition or Range*	*Description*
Derived	α_{it}	$\alpha_{it} = 1 - \exp\{-\nu_i A_{it}\}$	Probability of exposure to advertising, where A_{it} measures advertising expenditures
	ω_{it}	$\omega_{it} = \lambda S^{+}_{i,t-1}$	Probability of exposure to positive WOM
	ψ_{it}	$\psi_{it} = \lambda S^{-}_{i,t-1}$	Probability of exposure to negative WOM
	γ_{it}	$\gamma_{it} = 1 - \exp\{-\upsilon_i S_{it}\}$	Probability of release being convenient at theater, where S_{it} measures the number of screens showing movie *i*
	$\mathcal{E}_{it}$	$\mathcal{E}_{it} = 1 - (1-\alpha_{it})(1-\omega_{it})(1-\psi_{it})$	Probability of no exposure
	$\mathcal{P}_{it}$	$\mathcal{P}_{it} = \sum_{l,k,q\in\{0,1\}} w_{ji} \alpha_{it}{}^{l} (1-\alpha_{it})^{1-l} \omega_{it}{}^{k} (1-\omega_{it})^{1-k} \psi_{it}{}^{q} (1-\psi_{it})^{1-q}$	
	$\mathcal{L}_{it}$	$\mathcal{L}_{it} = \mathcal{P}_{it}\theta_i$	Probability of accepting the theme given a positive exposure
	$\mathcal{V}_{it}$	$\mathcal{V}_{it} = \delta^{*}_{it}\pi_i\tau_i + (1-\delta^{*}_{it}\pi_i)\tau_i\gamma_{it}$	Probability of immediately watching
Estimated	ν_i	$(0,\infty)$	Advertising Effectiveness Coefficient
	υ_i	$(0,\infty)$	Immediacy Coefficient
	λ_i	$(0,1)$	Probability of exposure to WOM
	θ_i	$(0,1)$	Probability of movie theme being acceptable
	ρ_i	$(0,1)$	Probability of a positive movie evaluation
	μ_i	$(0,1)$	Probability of remaining an active spreader
	ϕ_i	$(0,1)$	Probability of not forgetting
	τ_i	$(0,1)$	Probability of movie being watched now
	π_i	$(0,1)$	Probability of watching pirated version
	σ_i	$(0, \infty)$	Scale parameter of the Gamma distribution

Table B. Notation for variables and parameters used in the model. Note: all probabilities are conditional probabilities, where the conditioning information is defined in the text.