Technical Appendix to 'Modeling Movie Lifecycles and Market Share

1 Demand Model

All our models were estimated using Markov Chain Monte Carlo simulation (MCMC). This method is widely used in the marketing literature and is described in detail in Gelman, Carlin, Stern, and Rubin (1995). Given the vast literature on the subject, we limit our discussion to the specification of the various distributions used. In paragraph 1.1, we describe the conditional distributions of $\eta_{i,d}, \gamma_{i,d}, \beta_{i,d}$. These distributions are non-conjugate and thus are handled through a Metropolis step. In paragraphs 1.2 to 1.7 we describe the distributions used for the error terms (σ_d *and* V_d) and the parameters of the hierarchical regression ($\Delta_d, \theta_d^{Studio}$, *and* θ_d^{Genre}). As these distributions are conjugate, we handle them with a Gibbs step. For notational convenience, we define *m* as the number of movies in the dataset (*m* = 404), we use *i* to index movies, *t* to index time (162 weeks), and *d* to refer to the demand model.

1.1 Conditional distribution of $\eta_{i,d}, \gamma_{i,d}, \beta_{i,d}$ *(Metropolis Step)*

In section 3 of the paper, we defined the demand model (equation 7) as:

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

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

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

where:

$$\hat{S}_{it} = \eta_{i,d} w_{it}^{\gamma_{i,d}/\beta_{i,d}} e^{(1-w_{it})/\beta_{i,d}} ,$$

$$\eta_{i,d}^{*} = \ln(\eta_{i,d}) ,$$

$$\gamma_{i,d}^{*} = \gamma_{i,d} ,$$

$$\beta_{i,d}^{*} = \ln(\beta_{i,d}) ,$$

$$\varepsilon_{it,d} \sim N(0, \sigma_d^2) \text{ and}$$

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

1.2 Conditional distribution of σ_d (Gibbs Step)

σ_d is the standard deviation of the error term in the demand model. It is distributed Inverse Gamma as follows:

$$\sigma_d \sim IG\left(\nu + n, \sqrt{\left(\nu v + ns^2\right)/\left(\nu + n\right)}\right),$$

where:

ν=3; $v = 1$ (Priors chosen to be reasonable uninformative),

n is the total number of movie/week observation,

$$s^2 = \sum_{i} \sum_{it} (S_{it} - \hat{S}_{it})^2 .$$

1.3 Conditional distribution of V_d (Gibbs Step)

V_d is the variance of the vector of error terms (dimension 3) in the hierarchical regression of the parameters of the demand model. The prior on V_d is distributed Inverse Wishart(ν_0, V_0).

Consequently, the posterior conditional on V_d is distributed Inverse Wishart:

$$V_d \sim IW\left(\nu_0 + m, V_0 + \sum_i \left(\begin{bmatrix} \hat{\eta}^*_{i,d} \\ \hat{\gamma}^*_{i,d} \\ \hat{\beta}^*_{i,d} \end{bmatrix} - z_i \Delta_d\right)\left(\begin{bmatrix} \hat{\eta}^*_{i,d} \\ \hat{\gamma}^*_{i,d} \\ \hat{\beta}^*_{i,d} \end{bmatrix} - z_i \Delta_d\right)'\right).$$

Where m represents the total number of movies (i.e., 404). We chose $\nu_0 = 7, V_0 = I(3) \times 7$ as a locally diffuse prior ($I(3)$ is a 3x3 identity matrix).

1.4 Conditional distribution of Δ_d (Gibbs Step)

Let $\beta^\Delta_{i,d} = \begin{bmatrix} \eta^*_{i,d} \\ \gamma^*_{i,d} \\ \beta^*_{i,d} \end{bmatrix} - \theta^{Genre}_{i,d} - \theta^{Studio}_{i,d}$ be that component not attributable to studio or genre. And let

β^Δ be a stacked matrix of $\beta^{\Delta\prime}_{id}$ across all movies.

Then $\delta = \text{vec}(\Delta_d) \sim N\left(\tilde{d}, \left(V_d \otimes \left(Z'Z + A^\Delta_d\right)^{-1}\right)\right)$,

where:

$$\tilde{d} = vec\left(\tilde{D}\right),$$

$$\tilde{D} = \left(Z'Z + A_d\right)^{-1}\left(Z'Z\hat{D} + A_{d^\beta}\bar{D}\right),$$

$$\hat{D} = \left(Z'Z\right)^{-1} Z'\beta,$$

$\bar{D} = \text{stack}(\bar{d})$ is the prior on Δ_d, Z is a stacked matrix of the individual movies z_i's.

The prior on Δ_d is diffuse, given by:

$$\bar{d} = \text{vec}(\bar{D}) \sim N(0, A_d^{-1})$$
$$A_d = I(1) \times 0.01$$

1.5 Conditional distribution of θ_d^{Studio} (Gibbs Step)

We use a multi-dimensional version of the massively categorical approach suggested in Steenburgh *et al* (2003) to allow for an intercept shift (θ_d^{Studio}) on each parameter at the studio level.

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

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

1.6 Conditional distribution of $V_{d,Studio}$ (Gibbs Step)

The prior on $V_{d,Studio}$ is $IW(\nu_{Studio,0}, V_{Studio,0})$. Thus, the posterior conditional on $V_{d,Studio}$ is Inverse Wishart:

$$V_{d,Studio} \sim IW\left(\nu_{Studio,0} + m, V_{Studio,0} + \sum_i \left(\beta_{i,d}^{Studio} - \theta_{s_i,d}^{Studio}\right)\left(\beta_{i,d}^{Studio} - \theta_{s_i,d}^{Studio}\right)'\right)$$

Where $\theta_{s_i}^{Studio}$ is the vector of coefficients for the studio that represents movie *i,* and *m* once more represents the total number of movies. We chose $\nu_{Studio,0} = 7, V_{Studio,0} = I(3)\times 7$ as a locally diffuse prior.

1.7 Conditional distribution of θ_d^{Genre} and $V_{d,Genre}$ (Gibbs Step)

This proceeds exactly as for the studio coefficients in 1.5 and 1.6, except that the indexing is performed across movies that belong to each genre instead of movies represented by each studio.

2 Estimation for the Market Share Model

The estimation of the market share model is similar in nature to the estimation of the demand model. The hierarchical structure on the parameters $\eta_{i,ms}, \gamma_{i,ms}, \beta_{i,ms}$ is identical to the structure put on $\eta_{i,d}, \gamma_{i,d}, \beta_{i,d}$ except that we restrict $\eta_{i,ms}$ to be between 0 and 1 by using a Logit transform rather than log since it can be interpreted in terms of expected market share.

Thus, the steps described in sections 1.3 to 1.7 are identical and will not be repeated here. What is different here is that we model *market share* rather than demand; we introduce an *outside good*; and we use a *Poisson* approximation for the logit specification. These three changes are described in the next three sections.

2.1 Conditional distribution of $\eta_{i,ms}, \gamma_{i,ms}, \beta_{i,ms}$ (Metropolis Step)

Our choice model is defined by equation 1 as:

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

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

$$L(\eta_{i,ms},\gamma_{i,ms},\beta_{i,ms}) \sim \left(\prod_{it} \left(\frac{\left(e^{V_{it}}\right)^{S_{it}}}{\left(e^{V_{Ot}} + \sum_{j} e^{V_{jt}} I_{jt}\right)^{S_t}} \right)^{I_{it}} \right) \exp\left(-\frac{1}{2} \left(\begin{bmatrix} \eta^*_{i,ms} \\ \gamma^*_{i,ms} \\ \beta^*_{i,ms} \end{bmatrix} - \begin{bmatrix} \hat{\eta}^*_{i,ms} \\ \hat{\gamma}^*_{i,ms} \\ \hat{\beta}^*_{i,ms} \end{bmatrix} \right)' V_c^{-1} \left(\begin{bmatrix} \eta^*_{i,ms} \\ \gamma^*_{i,ms} \\ \beta^*_{i,ms} \end{bmatrix} - \begin{bmatrix} \hat{\eta}^*_{i,ms} \\ \hat{\gamma}^*_{i,ms} \\ \hat{\beta}^*_{i,ms} \end{bmatrix} \right) \right), \quad \text{(C-1)}$$

where:

$$V_{it} = \eta_{i,ms} w_{it}^{\gamma_{i,ms}/\beta_{i,ms}} e^{(1-w_{it})/\beta_{i,ms}};$$

$S_t = S_{Ot} + \sum_{j=1}^{J} S_{jt}$ (we discuss this term in more details in 2.3)

$$\eta^*_{i,ms} = \log\left(\eta_{i,ms}/(1-\eta_{i,ms})\right);$$

$$\gamma^*_{i,ms} = \gamma_{i,ms};$$

$\beta^*_{i,ms} = \ln(\beta_{i,ms});$ and

$$\begin{bmatrix} \hat{\eta}^*_{i,ms} \\ \hat{\gamma}^*_{i,ms} \\ \hat{\beta}^*_{i,ms} \end{bmatrix} = X_i \Delta_{ms} + \theta^{Studio}_{i,ms} + \theta^{Genre}_{i,ms}.$$

As in any logit specification, $\eta_{i,ms}$ is identifiable only to an arbitrary constant. We show in the next section how we set this constant such that $\eta_{i,ms}$ can be interpreted as the expected market share for movie *i* in its opening week.

2.2 Poisson Transform

The likelihood (C-1) above can be messy to evaluate directly, and indeed one encounters machine precision problems as S_{it} is frequently in the millions. A common transform used to make estimation more efficient is to use a Poisson approximation (Baker (1995); see also BUGS Manual p. 51; BUGS Examples, Vol. 2, p. 51, which we follow closely in our discussion).

To illustrate how this approximation works, let us look at a simplified likelihood formulation where we ignore the prior and the outside good (we will re-introduce the outside good in the next section). The likelihood for this simple case is given by:

$$L(Data \mid \eta_{ms}, \gamma_{ms}, \beta_{ms}) \sim \left(\prod_{i=1}^{I} \prod_{t=1}^{T} \frac{\left(e^{V_{it}}\right)^{I_{it} S_{it}}}{\left(\sum_{j=1}^{J} \left(e^{V_{jt}} I_{jt} \right) \right)^{S_t}} \right)$$

Suppose we assume that the data is actually generated by

$$S_{it} \sim Poisson(\omega_{it})$$
$$\omega_{it} = \bar{\omega}_t e^{V_{it}}$$

Then the likelihood is given by:

$$L(Data \mid \eta_{ms}, \gamma_{ms}, \beta_{ms}) \sim \prod_{t=1}^{T} \left(\left(\prod_{i=1}^{I} \omega_{it}{}^{S_{it}} e^{-\omega_{it}} I_{it} \right) \right)$$
$$= \prod_{t=1}^{T} \bar{\omega}_t{}^{S_t} \left(\exp\left(\sum_{i=1}^{I} S_{it} * V_{it} * I_{it} \right) \exp\left(-\bar{\omega}_t \sum_{i=1}^{I} e^{V_{it}} * I_{it} \right) \right)$$

Let the $\bar{\omega}_t$'s have independent gamma (a,b) priors; integrating them out gives a marginal likelihood of *V*'s:

$$L(Data \mid \eta_{i,ms}, \gamma_{i,ms}, \beta_{i,ms}) \propto \prod_{t=1}^{T} \exp\left(\sum_{i=1}^{I} S_{it} * V_{it} * I_{it} \right) \int \bar{\omega}_t^{S_t} \exp\left(-\bar{\omega}_t \sum_{i=1}^{I} e^{V_{it}} * I_{it} \right) \quad \bar{\omega}_t{}^{a-1} e^{-b\bar{\omega}} d\bar{\omega}_t$$
$$\Rightarrow L(Data \mid \eta_{i,ms}, \gamma_{i,ms}, \beta_{i,ms}) \propto \prod_{t=1}^{T} \frac{\exp\left(\sum_{i=1}^{I} S_{it} * V_{it} * I_{it} \right)}{\left[\sum_{i=1}^{I} \left(e^{V_{it}} + b \right) \right]^{S_t + a}}$$

We note that as $a, b \rightarrow 0$, the likelihoods become the same.

The inclusion of an outside good and the assumption that the total demand ($\bar{S}$) is constant allows us to make $\bar{\omega}_t$ invariant to time (i.e., $\bar{\omega}_t = \bar{\omega}$). Since this constant is unidentified for our logit model, we arbitrarily set it to $\bar{\varpi} = -e^{\bar{S}}$. As an important side effect, by adding this further constraint, η_i can now be directly interpreted as the expected of movie *i* market share in its opening week.

2.3 The "Outside Good" (Metropolis Step)

Demand for movies is highly seasonal. A useful way to directly account for seasonality is to imagine non-demand for movies. That is, in every period there is a percentage of the total potential market that chooses not to attend movies in that period. Those who chose not to go to the movies are said to consume the "outside good." This is analogous to the no-purchase option in the choice literature. We define this demand for the outside good as S_{Ot} .

If we assume that total demand is constant over time ($S_t = \bar{S}$) and is allocated every week between the movies in theater and the outside good as a function of the quality of the movies, and the quality of the other options available to consumers, we can compute S_{Ot} as:

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

The total demand can be arbitrarily set to a large number. We set it such that total sales in each period is 1.2* total tickets sold in the highest demand week within the dataset, to ensure that there is positive demand for the outside good in each period. We tested values of 1.1 and 1.4; there were no effects other than an intercept shift.

Next, to account for seasonal change in the demand for movies, we model the demand for the outside good as a function of the demand for the outside good in past years. Using a simple auto-regressive model, the market share of the outside good is given as:

$$V_{Ot} = \alpha + \sum_{k=1}^{K} \phi_k \ln\left(\frac{S_{Ot-(k*52)}}{\bar{S}}\right), \qquad \text{(C-2)}$$

where $S_{Ot-52*k}$ is ($\bar{S}$ -total demand for movies in week *t*, *k* years previously.

In (C-2), we divide the lagged demand by $\bar{S}$ so that this can be thought of as market share for the outside good in week *t*, *k* years previously. With this outside good specification, the conditional likelihood for the parameters for each movie becomes:

$$L(\eta_{i,ms},\gamma_{i,ms},\beta_{i,ms}) \sim \left(\prod_{t=1}^{T} I_{it} e^{V_{it}*S_{it}/\bar{S}} * e^{-e^{V_{it}/\bar{S}}}\right)\exp\left(-\frac{1}{2}\left(\begin{bmatrix}\eta^*_{i,ms}\\ \gamma^*_{i,ms}\\ \beta^*_{i,ms}\end{bmatrix}-\begin{bmatrix}\hat{\eta}^*_{i,ms}\\ \hat{\gamma}^*_{i,ms}\\ \hat{\beta}^*_{i,ms}\end{bmatrix}\right)' V_c^{-1}\left(\begin{bmatrix}\eta^*_{i,ms}\\ \gamma^*_{i,ms}\\ \beta^*_{i,ms}\end{bmatrix}-\begin{bmatrix}\hat{\eta}^*_{i,ms}\\ \hat{\gamma}^*_{i,ms}\\ \hat{\beta}^*_{i,ms}\end{bmatrix}\right)\right),$$

and for the outside good it is:

$$L(\alpha,\phi_1..\phi_k) \sim \left(\prod_{t=1}^{T} e^{V_{Ot}*S_{it}/\bar{S}} * e^{-e^{V_{Ot}/\bar{S}}}\right)\exp\left(-\frac{1}{2}\left(\begin{bmatrix}\alpha\\ \phi_1\\ \vdots\\ \phi_k\end{bmatrix}\right)' V_O^{-1}\left(\begin{bmatrix}\alpha\\ \phi_1\\ \vdots\\ \phi_k\end{bmatrix}\right)\right),$$

where the prior on these parameters is chosen to be diffuse, N(0,V_O), where $V_O = I*10^6$. In practice we found k=3 to work best (i.e., we used lagged demand for the outside good in the previous 3 years to predict current year sales).

Reference

Baker, S. G. (1995), "The multinomial-poisson transformation," *The Statistician*, Vol. 43, 495-504.

BUGS Manual, http://www.mrc-bsu.cam.ac.uk/bugs/documentation/contents.shtml, version 0.5, 51.

BUGS Examples, http://www.mrc-bsu.cam.ac.uk/bugs/documentation/contents.shtml, Vol. 2, Version 0.5, 51.

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