FORECASTING MOVIE ATTENDANCE OF INDIVIDUAL MOVIE SHOWINGS: A HIERARCHICAL BAYES APPROACH

by

Jin Hee Jinny Lim

B.Sc in Statistics, Simon Fraser University, 2010

a Thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Science

in the
Department of Statistics and Actuarial Science
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2012

APPROVAL

Name: Jin Hee Jinny Lim

Degree: Master of Science

Title of Thesis: Forecasting Movie Attendance of Individual Movie Showings: A Hierarchical Bayes Approach

Examining Committee:

Richard Lockhart
Professor
Chair

____

David Alexander Campbell
Senior Supervisor
Assistant Professor

____

Lawrence McCandless
Supervisor
Assistant Professor
Health Science, Simon Fraser University

____

Jason Ho
Examiner
Assistant Professor
Business School, Simon Fraser University

Date Approved: ______________________________

Abstract

Despite the availability of transaction data, most movie theaters nowadays still rely on managers' gut feeling to decide how many times and when a certain movie will be screened. Eliashberg et al. (2009) suggest that movie theaters could improve their profits by a more data-driven approach such as a movie attendance forecasting model. However, there are two limitations in the model. First, it does not capture both cannibalization and demand expansion effects. Second, it does not accurately access the uncertainty when making predictions for new movies. To address the limitations in Eliashberg et al. (2009), three hierarchical Bayes models of movie attendance are investigated and compared: linear regression model, standard logit model and nested logit model. Hierarchical linear regression model extends Eliashberg et al's model by accurately assessing the uncertainty in the predicted admissions. The standard logit model captures both the cannibalization and demand expansion effects in a relatively restrictive manner because of the property called independence from irrelevant alternatives, IIA. The nested logit model relaxes the restrictive IIA property and thus better captures the cannibalization and demand expansion effects.

Keywords: Demand Expansion; Cannibalization; Linear Regression; Standard Logit Model; Nested Logit Model; Hierarchical Bayesian Approach; MCMC

To my family.

"When one door closes another door opens, but we so often look so long and so regretfully upon the closed door, that we do not see the ones which open for us."

— Alexander Graham Bell

Acknowledgments

I would like to express my deep appreciation for my supervisor, Dr. Dave Campbell, who made my Masters degree possible. Because of his endless support, clear guidance and encouragement, I was able to finish this thesis. I always appreciate his insights about problems and suggestions on how to improve. Without his help, I would not have been able to get to where I am now. I would also like to thank Dr. Jason Ho for giving me this interesting and exciting project. I really appreciate our Sunday meetings where we discussed this project and where he taught me business analytics and marketing theory. Last, but not least, I would like to thank Dr. Lawrence McCandless for being my commitee member.

Thank you to all of the faculty and staff in the Department of Statistics and Actuarial Science for their kindness and support in helping me pursue my Masters degree. Thank you also to my classmates for their smart ideas and suggestions.

Finally, I would like to thank my family and Wesley for their endless love and support.

Contents

Approval **ii**

Abstract **iii**

Dedication **iv**

Quotation **v**

Acknowledgments **vi**

Contents **vii**

List of Tables **x**

List of Figures **xi**

1 Introduction **1**

1.1 Motivation 1

1.2 Project Outline 3

2 Linear Regression Model **4**

2.1 Model 4

2.2 Bayesian Approach: Gibbs Sampling 6

2.2.1 Likelihood Function, Prior and Hyper Prior Distributions 6

2.2.2 Derivation of Full Conditional Distributions 7

2.2.2.1 Full Conditional Distribution of $[\alpha_{jk} \,|\, \cdot\,]$ 8

2.2.2.2 Full Conditional Distribution of $[\omega \,|\, \cdot\,]$ 9

2.2.2.3 Full Conditional Distribution of $\left[\sigma^2 \,|\, \cdot\right]$ . . . 9
2.2.2.4 Full Conditional Distribution of $[\eta \,|\, \cdot]$ . . . 10
2.2.2.5 Full Conditional Distribution of $[\Sigma \,|\, \cdot]$ . . . 11

3 Standard Logit Model **12**
3.1 IIA Property . . . 13
3.2 Movie Forecasting Model . . . 14
3.3 Posterior Distributions . . . 15
3.3.1 Posterior Distribution of $[\alpha_g \,|\, \cdot]$ . . . 17
3.3.2 Posterior Distribution of $[\omega \,|\, \cdot]$ . . . 18
3.3.3 Full Conditional Distribution of $[\eta^* \,|\, \cdot]$ . . . 19
3.3.4 Full Conditional Distribution of $[\Sigma \,|\, \cdot]$ . . . 19

4 Nested Logit Model **20**
4.1 Movie Forecasting Model . . . 20
4.2 Posterior Distributions . . . 23
4.2.1 Posterior Distribution of $[\alpha_g \,|\, \cdot]$ . . . 24
4.2.2 Posterior Distribution of $[\omega \,|\, \cdot]$ . . . 25
4.2.3 Posterior Distribution of $[\sigma_{jk} \,|\, \cdot]$ . . . 25
4.2.4 Posterior Distribution of $[\sigma \,|\, \cdot]$ . . . 26
4.2.5 Full Conditional Distribution of $[\eta^* \,|\, \cdot]$ . . . 26
4.2.6 Full Conditional Distribution of $[\Sigma \,|\, \cdot]$ . . . 27

5 Data **28**
5.1 Description of Data . . . 28
5.2 Implementation . . . 30

6 Result **31**
6.1 Actual Movie Ticket Sales vs. Predicted Median Movie Ticket Sales . . . 31
6.2 Posterior Predictive Distribution for Existing Movies . . . 34
6.3 Posterior Predictive Distributions for New Movies . . . 38

7 Conclusion and Recommendation **41**

Appendix A Posterior Distribution of σ_{jk} **43**

A.1 σ_{jk} where genre and age restriction are used to construct a hierarchical layer 43

A.2 Posterior for σ_g 44

A.3 Posterior for γ_g 45

A.4 Posterior for δ_g 45

A.5 Posterior for β_{γ_g} 45

A.6 Posterior for β_{δ_g} 46

Appendix B Poisson Model For Counts of Ticket Sales **47**

B.1 Likelihood $p(S \mid \mu)$ 47

B.2 Prior $p(\mu)$ 48

B.3 Full Conditional Distribution 48

Bibliography **49**

List of Tables

5.1 Sample of Data 29

6.1 R^2 of three models on existing, new and all movies. By looking at R^2 for all movies, linear regression model and standard logit model are much better than nested logit model. However, when the movies are broken down into new and existing movies, the nested logit model is the clear winner. A more detailed discussion of the results is in Chapter 7. 33

List of Figures

3.1 Example of standard logit structure on a given day d. On a given day d, there are choice sets of movie and hour combinations as well as not watching movie, which is the outside option. 14

4.1 Example of nested logit structure on a given day d. On a given day d, there are two choice sets: (1) watching a movie at a theater or (2) doing other activity. For the choice set of movie (1), it can be further broken down into first, different kinds of movies and second, different showings of the particular movie that is chosen. 21

6.1 Scatter plot of actual movie ticket sales vs predicted median of movie ticket sales in the period of January 24, 2008 to Jan 30, 2008. Top graphs are actual scale of movie ticket sales and the bottom graphs are log scales of movie ticket sales. Black dots are predictions for existing movies and red plus signs are predictions for new movies. The blue line is the reference line for the perfect match of the actual and the predicted ticket sales. 32

6.2 Posterior predictive distributions of Bee Movie on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales. 35

6.3 Posterior predictive distributions of The Nanny Diaries on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales. 36

6.4 Posterior predictive distributions of Moordwijven on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales. 37

6.5 Posterior predictive distributions of Cloverfield on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales. 39
6.6 Posterior predictive distributions of We Own the Night on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales. 40

Chapter 1

Introduction

1.1 Motivation

As technology advances, large amounts and varieties of data are being created at an increasing rate every day. Such abundance of data is creating opportunities for businesses and organizations to improve their decision making process [3]. Movie scheduling at multiplex movie theaters is one of the contexts, where data could be used to improve the effectiveness and efficiency of the decision. Despite the availability of transaction data, most movie theaters nowadays still rely on managers' gut feelings to decide how many times and when a certain movie will be screened. For example, by every Monday, movie theater managers use only their subjective judgements to finalize the movie showing schedules for the coming week. However, as shown by Eliashberg et al. (2009), movie theaters could improve their profits by a more data-driven approach.

The data-driven approach consists of two components namely a forecasting model based on the transaction data, and an optimization algorithm to schedule individual movie showings based on the forecasts. Focusing on the forecasting model, the goal of this project is to estimate three alternative models using the hierarchical Bayes approach [5]. If a movie theater can more accurately forecast movie attendances of individual showings, it can formulate a movie screening schedules, which would maximize the admissions subject to its current capacity. Given the size of the data involved and complexity of the models, the model estimation is a non-trivial exercise. The three alternative models estimated in this project extend the forecasting model in Eliashberg et al. (2009). Eliashberg et als movie attendance forecasting model has two major limitations. First, their forecasting model is essentially a linear regression model which assumes each movie showing (i.e. each observation) to be independent of one another. Such an independence assumption implies that a movie showing would be predicted to generate the same number of admissions, no matter whether it is scheduled with two or twenty other showings of the same movie.

As explained in Ho (2005), a model of movie attendance needs to be able to capture two effects of movie schedule change. To illustrate the two effects, let us consider a scenario, where a new movie showing is added to an existing movie schedule. The new movie showing would create two effects: (1) it would crowd out or cannibalize the existing showings in the movie schedule by simply encouraging moviegoers, who would go to other showings, to switch to this new movie showing, and (2) it would expand the total admission at the theater by providing one more starting time, which adds to the important value of convenience to the moviegoers, who originally would not visit the particular movie theater at all. However, in a linear regression model like that in Eliashberg et al. (2009), adding extra movie showing would not affect any of the existing movies in the movie schedule and the total demand will always increase. To address this limitation, this project will estimate two choice models, which explicitly characterize both the cannibalization effect and demand expansion effects. The second limitation in Eliahberg et al. (2009) is its inefficient way to generate predictions for new movies, which are yet to be screened in the theater and thus have no historical data.

A key specification in Eliashberg et als forecasting model is the two movie-specific parameters: θ_j and λ_j where θ_j captures the time-invariant base attractiveness of movie j, while λ_j characterize the weekly attendance decay rate of movie j. While these two movie-specific parameters provide a very flexible structure to fit the time trend of any one of the currently running movies, they create a non-trivial problem when the model needs to be used to make prediction for new movies, which we do not have any data to estimate the corresponding movie-specific parameters. To address this problem, Eliashberg et al. (2009) consider all existing movies' parameters, θ_j and λ_j for all j to be "data" and regress these data on covariates like genres, which are then plugged into the original admission prediction model. This approach is suboptimal by ignoring all the uncertainty surrounding the movie-specific parameters. This project explicitly considers the uncertainty surrounding the movie-specific parameters and relate them to the covariates in a hierarchical Bayes structure [5]. When making prediction for a new movie, the hierarchical Bayes model would still start from covariates like genre and age restriction but would more accurately represent the uncertainty on the predicted admissions.

1.2 Project Outline

To address the limitations in Eliashberg et al. (2009), a forecasting model is needed which can capture both cannibalization and demand expansion effects as well as can accurately assess the uncertainty when making predictions for new movies. In this project, three hierarchical Bayes models are investigated and compared: linear regression model (Chapter 2), standard logit model (Chapter 3), and nested logit model (Chapter 4). First, the linear regression model extends Eliashberg et al.'s model by accurately assessing the uncertainty in the predicted admissions. The standard logit model then capture both the cannibalization and demand expansion effects in a relatively restrictive manner, because of the property called independence from irrelevant alternatives, IIA (Chapter 3). The nested logit model relaxes the restrictive IIA property and thus better captures the cannibalization and demand expansion effects. Chapter 5 describes the data, Chapter 6 compares the predictions of movie ticket and Chapter 7 discusses the results.

Chapter 2

Linear Regression Model

This chapter describes the hierarchical Bayes linear regression model on the number of movie ticket sales. The main purpose is to predict the number of movie ticket sales by estimating parameters using a Bayesian approach. Note that, the linear regression model does not capture any substitution effects because it assumes each movie showing is independent of one another. That is, adding extra movie would not affect the admissions to existing movies and as a result, the total demand will always increase. Since the linear regression function forms the basis for the utility function, V_{in}, in the standard logit model in chapter 3 and nested logit model in chapter 4, it should be a good benchmark for comparing models. As the linear regression model is of a hierarchical Bayes structure, it would better capture the uncertainty of predicted admissions than that in Eliashberg et al. (2009).

2.1 Model

Following Eliashberg et al. (2009), the regression model assumes a log relation between the response variable, the number of ticket sales for a specific movie showing and the predictor variables.

$$
\begin{aligned}
\ln(S_{jkhd}) &= \alpha_{jk}\, X_{jkhd} + \omega\, Y_d + \epsilon_{jkhd} \qquad (2.1)\\
&\text{where,}\\
&\alpha_{jk} = \left[\theta_{jk},\, \lambda_{jk},\, \beta_{jk_{hour}},\, \beta_{jk_{hour^2}},\, \beta_{jk_{hour^3}},\, \beta_{jk_{hour^4}}\right]'\\
&\omega = [\omega_{dw},\, \omega_{hw}]'\\
&\epsilon_{jkhd} \overset{\text{iid}}{\sim} N(0, \sigma^2)
\end{aligned}
$$

The response variable, the number of ticket sales for the version k of movie j in hour h of date d is denoted as S_{jkhd}. The predictor variables can be decomposed into two components X_{jkhd} and Y_d. X_{jkhd} is a vector containing variables including (1) an indicator variable of version k of movie j, (2) age of movie and (3) starting time of the movie showings. α_{jk} is a parameter corresponding the above data. θ_{jk} is a coefficient of the indicator variable and the parameter capturing the weekly attractiveness of version k of movie j. λ_{jk} is a coefficient of age of movie j of version k which is a parameter capturing the weekly attractiveness decay of movie. Note that starting time is treated as a continuous variable. For maximum flexibility, a polynomial of degree four is used and $\left[\beta_{jk_{hour}},\, \beta_{jk_{hour^2}},\, \beta_{jk_{hour^3}},\, \beta_{jk_{hour^4}}\right]$ are coefficients for the linear, quadratic, cubic and quartic terms relatively for version k of movie j.

Y_d is a vector containing the dummy variables of day of week and holidays. ω is a vector containing all the parameters corresponding the above variables. ω_{DW} are set of parameters capturing the effects of different days of the week, namely Monday, Tuesday, etc and ω_{HW} are set of parameters capturing the effects of the holidays if date d falls on a holiday.

As a hierarchical structure, α_{jk} is assumed to vary with a set of predictor variables, Z_{jk}. which include genres and age restriction for the version k of movie j.

$$
\begin{aligned}
\alpha_{jk} &= \eta\, Z_{jk} + \epsilon_{\alpha_{jk}} \qquad (2.2)\\
&\text{where } \epsilon_{\alpha_{jk}} \overset{\text{iid}}{\sim} N(0, \Sigma)
\end{aligned}
$$

The η is a vector of parameters capturing the effects of genre and age restriction of movies.

2.2 Bayesian Approach: Gibbs Sampling

In order to assess the posterior distribution of parameters (target distribution), Markov Chain Monte Carlo method, MCMC, is used. MCMC is a general iterative algorithm that in each iteration draws a sample of parameters from a proposal distribution based on a previous value. Then it accepts or rejects the proposed value according to the acceptance ratio. As the number of iterations increase, the samples would converge to target distribution which is the posterior distribution [5]. Gibbs sampling is one of the techniques in MCMC where all the samples of parameters are drawn from the proposal distribution which is the full conditional and consequently, the proposed values are always accepted [5].

2.2.1 Likelihood Function, Prior and Hyper Prior Distributions

$$[\, S \,|\, \alpha_{jk},\, \omega,\, \sigma^2] \;=\; \prod_{jk=1}^{n} \text{Normal}(\alpha_{jk}\, X_{jkhd} + \omega\, Y_d \;,\, \sigma^2) \tag{2.3}$$

$$[\alpha_{jk} \,|\, \eta,\, \Sigma] \;\sim\; \text{Multivariate Normal}(\eta\, Z_{jk} \;,\, \Sigma) \tag{2.4}$$

$$[\omega] \;\sim\; \text{Multivariate Normal}(0 \;,\, \text{I}) \tag{2.5}$$

$$[\sigma^2] \;\sim\; \text{Inverse Gamma}(1, 1) \tag{2.6}$$

$$[\eta] \;\sim\; \text{Multivariate Normal}(0 \;,\, \text{I}) \tag{2.7}$$

$$[\Sigma] \;\sim\; \text{Inverse Wishart}(7 \;,\, \text{M}) \tag{2.8}$$

The likelihood function of S is defined in (2.3) where jk is the movie index and N is the total number of movies. The distributions from (2.4) to (2.6) are prior distributions and from (2.7) to (2.8) are hyper prior distributions. Note that the prior distributions are chosen to be conjugate to the likelihood in order to take advantage of having full conditional and using the Gibbs sampling method. For Σ, the inverse wishart distribution is chosen with 7 degree of freedom and with scale matrix to be 3 in diagonal and 1.5 off diagonal. The reason why the degree of freedom is set to be 7 is that it should be larger than the number of parameters and Σ is 6 by 6 matrix.

2.2.2 Derivation of Full Conditional Distributions

As stated earlier, the likelihood is normal and the priors are conjugate distributions. As such the full conditional distribution (derived next), which are of closed forms of known distributions. The posterior distribution of α_{jk}, the vector containing movie specific multivariate parameters capturing the opening week attractiveness, age decay and time of movie showing effects, is normally distributed. The posterior distribution of ω, the vector containing parameters capturing the seasonality effects and day of week effects, is normally distributed. The posterior distribution of σ^2, the variance of linear regression model, is inverse gamma distribution. The posterior distribution of hyper parameter vector η, capturing the effects of genre and age restriction, is normally distributed and Σ, the variance covariance matrix of α_{jk}, is inverse wishart distribution. The detailed derivations of the posterior distributions are in the following. Note that the results of the predictions are left for Chapter 6 so as to make a comparison between models.

2.2.2.1 Full Conditional Distribution of $[\alpha_{jk} \mid \cdot]$

$$
\begin{aligned}
[\alpha_{jk} \mid \cdot] &= [\, S \mid \alpha_{jk},\, x_{jkhd},\, y,\, \omega \,] \times [\, \alpha_{jk} \mid \eta, z_{jk}, \Sigma \,] \\
&= (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right]' \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right] \right) \\
&\quad \cdot (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \left[\alpha_{jk} - \eta z_{jk} \right]' \Sigma^{-1} \left[\alpha_{jk} - \eta z_{jk} \right] \right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2} \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right]' \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right] \right) \\
&\quad \cdot \exp\left(-\frac{1}{2} \left[\alpha_{jk} - z'_{jk}\eta'\Sigma^{-1} \right] \left[\alpha_{jk} - \eta z_{jk} \right] \right) \\
&\propto \exp\left(-\frac{1}{2\sigma^2} \left[-2\alpha'_{jk} x_{jkhd} s_{jkhd} - \alpha'_{jk} x'_{jkhd} x_{jkhd} \alpha_{jk} + 2\alpha'_{jk} x_{jkhd} y\omega \right] \right) \\
&\quad \cdot \exp\left(-\frac{1}{2} \left[\alpha'_{jk}\Sigma^{-1}\alpha_{jk} - \alpha'_{jk}\Sigma^{-1}\eta z_{jk} - z'_{jk}\eta'\Sigma^{-1}\alpha_{jk} \right] \right) \\
&= \exp\left(-\frac{1}{2} \left[\frac{-2\alpha'_{jk} x_{jkhd} s_{jkhd}}{\sigma^2} + \frac{\alpha'_{jk} x'_{jkhd} x_{jkhd} \alpha_{jk}}{\sigma^2} + \frac{2\alpha'_{jk} x_{jkhd} y\omega}{\sigma^2} \right] \right) \\
&\quad \cdot \exp\left(-\frac{1}{2} \left[\alpha'_{jk}\Sigma^{-1}\alpha_{jk} - 2\alpha'_{jk}\Sigma^{-1}\eta z_{jk} \right] \right) \\
&\propto \exp\left(-\frac{1}{2} \left(\alpha'_{jk} \left[\frac{x'_{jkhd} x_{jkhd}}{\sigma^2} + \Sigma^{-1} \right] \alpha_{jk} - 2\alpha'_{jk} \left[\frac{x'_{jkhd} s_{jkhd}}{\sigma^2} - \frac{x_{jkd} y\omega}{\sigma^2} + \Sigma^{-1}\eta z_{jk} \right] \right) \right) \\
&\quad \text{Let Q} = \left[\alpha_{jk} - \left(\frac{x'_{jk} x_{jk}}{\sigma^2} + \Sigma^{-1} \right)^{-1} \left(\frac{x'_{jkhd} s_{jkhd}}{\sigma^2} - \frac{x_{jkd} y\omega}{\sigma^2} + \Sigma^{-1}\eta z_{jk} \right) \right] \\
&\quad \text{Let R} = \left[\frac{x'_{jk} x_{jk}}{\sigma^2} + \Sigma^{-1} \right] \\
&\propto \exp(-\frac{1}{2}\text{Q}'\text{RQ}) \\
&\sim \text{MVN}(\text{R}^{-1} \left[\frac{x'_{jkhd} s_{jkhd}}{\sigma^2} - \frac{x_{jkd} y\omega}{\sigma^2} + \Sigma^{-1}\eta z_{jk} \right], \text{R}^{-1})
\end{aligned} \tag{2.9}
$$

2.2.2.2 Full Conditional Distribution of $[\omega \,|\, \cdot]$

$$\begin{aligned}
[\omega \,|\, \cdot] &\propto [\, S \,|\, \alpha_{jk},\, x_{jkhd},\, y,\, \omega \,] \times [\, \omega \,] \\
&\propto \exp\left(-\frac{1}{2\sigma^2} \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right]' \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right] \right) \\
&\quad \cdot \exp\left(-\frac{1}{2} \left(\omega - 0 \right)' I^{-1} \left(\omega - 0 \right) \right) \\
&\propto \exp\left(-\frac{1}{2} \left[\frac{1}{\sigma^2} \left(-2\omega' y' s_{jkhd} + 2\omega' y' x_{jkhd}\alpha_{jk} + \omega' y' y \omega \right) + \omega'\omega \right] \right) \\
&= \exp\left(-\frac{1}{\sigma^2} \left[\omega' \left(y'y + \sigma^2 \right) \omega - 2\omega' \left(\frac{y' s_{jkhd} - y' x_{jkhd}\alpha_{jk}}{1+\sigma^2} \right) \right] \right) \\
&\sim \mathrm{MVN}\left(\left[y'y + \sigma^2 I \right]^{-1} \left[y' s_{jkhd} - y' x_{jkhd}\alpha_{jjk} \right],\, \left[y'y + \sigma^2 I \right]^{-1} \right) \qquad (2.10)
\end{aligned}$$

2.2.2.3 Full Conditional Distribution of $\left[\sigma^2 \,|\, \cdot\right]$

$$\begin{aligned}
\left[\sigma^2 \,|\, \cdot \right] &= [\, S \,|\, \alpha_{jk},\, x_{jkhd},\, y,\, \omega \,] \times \left[\, \sigma^2 \,\right] \\
&= \prod_{jk=1}^{n} \left[\left(2\pi\sigma^2 \right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \left(s_{jkhd} - \left(x_{jkhd}\alpha_{jk} + y\omega \right) \right)^2 \right) \right] \\
&\quad \cdot \frac{b^a}{\gamma(a)} \left(\sigma^2 \right)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2} \right) \\
&\propto \left(\sigma^2 \right)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} \left[s_{jkhd} - \left(x_{jkhd}\alpha_{jk} + y\omega \right) \right]' \left[s_{jkhd} - \left(x_{jkhd}\alpha_{jk} + y\omega \right) \right] \right) \\
&\quad \cdot \left(\sigma^2 \right)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2} \right) \\
&= \left(\sigma^2 \right)^{-\left(\frac{N}{2} + a + 1 \right)} \exp\left[-\frac{1}{2\sigma^2} \left(b + \frac{1}{2} \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right]' \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right] \right) \right] \\
&\sim \mathrm{IG}\left(\frac{N}{2} + a,\, b + \frac{1}{2} \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right]' \left[s_{jkhd} - x_{jkhd}\alpha_{jk} - y\omega \right] \right) \qquad (2.11)
\end{aligned}$$

2.2.2.4 Full Conditional Distribution of $[\eta \,|\, \cdot\,]$

$$
\begin{aligned}
[\eta \,|\, \cdot\,] &\propto [\,\alpha_{jk} \,|\, \eta, z_{jk}, \Sigma\,] \times [\,\eta\,] \;\propto\; \left[\,\alpha^*_{jk} \,|\, \eta^*, z_{jk}, \Sigma\,\right] \times [\,\eta^*\,] \\
&\propto |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\left[\alpha^*_{jk} - (z_{jk} \otimes I_{15})\eta^*\right]' (I_m \otimes \Sigma^{-1}) \left[\alpha^*_{jk} - (z_{jk} \otimes I_{15})\eta^*\right]\right) \\
&\quad \cdot \exp\left(-\frac{1}{2}\eta'^* I_{15}^{-1} \eta'^*\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\left[\alpha^*_{jk} - (z_{jk} \otimes I_{15})\eta^*\right]' (I_m \otimes \Sigma^{-1}) \left[\alpha^*_{jk} - (z_{jk} \otimes I_{15})\eta^*\right] + \eta'^*\eta^*\right)\right) \\
&= \exp\left(-\frac{1}{2}\left(\left[\alpha'^*_{jk}(I_m \otimes \Sigma^{-1}) - \eta'^*(z_{jk} \otimes I_{15})'(I_m \otimes \Sigma^{-1})\right] \left[\alpha^*_{jk} - (\eta \otimes I_{15})\eta'^*\right] + \eta'^*\eta^*\right)\right) \\
&\quad \text{Note: } (z_{jk} \otimes I_{15})'(I_m \otimes \Sigma^{-1}) = z_{jk} \otimes \Sigma^{-1} \\
&\quad \text{Note: } (z' \otimes I'_m)(I_m \otimes \Sigma^{-1})(z_{jk} \otimes I_{15}) = z_{jk} \otimes \Sigma^{-1} \\
&\propto \exp\left(-\frac{1}{2}\left[(I_m \otimes \Sigma^{-1})(z_{jk} \otimes \Sigma^{-1})\alpha'_{jk} + \eta^* z'_{jk} z_{jk} \otimes \Sigma^{-1}\eta^* + \eta'^*\eta^*\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[-2\eta'^* z_{jk} \otimes \Sigma^{-1}\alpha^*_{jk} + \eta^* z'_{jk} z_{jk} \otimes \Sigma^{-1}\eta^* + \eta'^*\eta^*\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\eta'^*(z'_{jk} z_{jk} \otimes \Sigma^{-1} + I_{15})\eta'^* - 2\eta'^*(z_{jk} \otimes \Sigma^{-1}\alpha^*_{jk})\right]\right) \\
&\sim \text{MVN}(\eta^* \,|\, \text{U, V}) \\
&\quad \text{where V} = [z'_{jk} z_{jk} \otimes \Sigma^{-1} + I_{15}]^{-1} \text{ and U} = \text{V}[(z'_{jk} \otimes \Sigma^{-1})\alpha^*_{jk}] \qquad (2.12)
\end{aligned}
$$

2.2.2.5 Full Conditional Distribution of $[\Sigma \,|\, \cdot\,]$

$$
\begin{aligned}
[\Sigma \,|\, \cdot\,] &\propto [\,\alpha_{jk} \,|\, \eta, z_{jk}, \Sigma\,] \times [\,\Sigma\,] \\
&\propto \prod_{jk=1}^{n} (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left[\alpha_{jk} - \eta z_{jk}\right]' \Sigma^{-1} \left[\alpha_{jk} - \eta z_{jk}\right]\right) \\
&\quad \cdot |\Sigma|^{\frac{-a+k+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1} b\right)\right) \\
&\propto |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{jk=1}^{n}\left[\alpha_{jk} - \eta z_{jk}\right]' \Sigma^{-1} \left[\alpha_{jk} - \eta z_{jk}\right]\right) \\
&\quad \cdot |\Sigma|^{\frac{-a+k+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1} b\right)\right) \\
&\quad \text{Note: } \sum_{i=1}^{n} y_i' \Sigma y_i = \mathrm{tr}\left[\Sigma \sum_{i=1}^{n} y_i' y_i\right] \\
&= |\Sigma|^{\frac{-n-a+k+1}{2}} \exp\left(\mathrm{tr}\left(-\frac{1}{2}\sum_{jk=1}^{n}\left[\alpha_{jk} - \eta z_{jk}\right]' \Sigma^{-1} \left[\alpha_{jk} - \eta z_{jk}\right]\right)\right) \\
&\quad \cdot \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1} b\right)\right) \\
&= |\Sigma|^{-\frac{1}{2}(n+a-k-1)} \exp\left(-\frac{1}{2}\mathrm{tr}\left[\Sigma^{-1}\left(b + \sum_{jk=1}^{n}\left[\alpha_{jk} - \eta z_{jk}\right]' \left[\alpha_{jk} - \eta z_{jk}\right]\right)\right]\right) \\
&\sim \mathrm{IW}\left(n + a, b + \sum_{jk=1}^{n}\left[\alpha_{jk} - \eta z_{jk}\right]' \left[\alpha_{jk} - \eta z_{jk}\right]\right) \qquad (2.13)
\end{aligned}
$$

Chapter 3

Standard Logit Model

As mentioned earlier, the linear regression model does not capture properly both the demand expansion and cannibalization effects. To address these shortcomings, the choice models are good candidates [1]. In the choice model, potential customers are exposed to multiple choice alternatives. In the movie attendance contexts, choice alternatives are watching a specific movie showing starting at a specific time. But there is alternative called outside option where customers choose to not watch a movie in the theater at all.

Each choice occasion in multinomial choice model involves choice alternatives and each of the choice alternatives is represented by a utility function. The utility function for choice j for individual n, U_{jn}, consists of two components: systematic component, V_{jn}, and random component, ϵ_{in}

$$U_{jn} = V_{jn} + \epsilon_{jn} \tag{3.1}$$

Depending on the different assumptions on the disturbance, ϵ_{in}, different models are derived: standard logit model or nested logit model [1]. In this chapter, standard logit model with its restrictive IIA assumption is discussed.

3.1 IIA Property

When all the disturbances, ϵ_{jn}, for all $j \in C_n$, where C_n is individual n's all possible choice alternatives, are independently and identically distributed (iid) and Gumbel distributed with a location parameter a, and scale parameter $b > 0$, the standard logit model of choice i of individual n is defined as

$$P_n(i) = \frac{\exp(V_{in})}{\sum_{j \in C_n} \exp(V_{jn})} \tag{3.2}$$

Because of the iid assumption imposed on the disturbance, the standard logit model has a property called independence from irrelevant alternatives (IIA) [1]. The IIA property is the "ratio of choice probabilities of any two alternatives is unaffected when another alternative is added in the choice set" (Ben-Akiva and Lerman, 1985). The odds for two choices i and j facing person n are as follow.

$$\frac{P_n(i)}{P_n(j)} = f(V_{in}, V_{jn}) \tag{3.3}$$

IIA property is such a strong restriction that it would yield unintuitive substitution patterns. To understand the IIA property, let's consider a choice occasion: a showing of Harry Potter starting at 8pm (Harry Potter$_{8pm}$) and Transformer starting at 9pm (Transformer$_{9pm}$). Without loss of generality, let's assume that both movies have the same systematic utility, V, and thus each choice has equal probability being chosen (ie. P(Harry Potter$_{8pm}$) $= P$(Transformer$_{\text{9pm}}$) $= 0.5$). In other words, the odds of the choice probability of the Transformer$_{9pm}$ and the Harry Potter$_{8pm}$ is 1. Now, assume that another showing of Harry Potter starting at 10pm (Harry Potter$_{10pm}$) is introduced to the choice set. Note that Harry Potter$_{10pm}$ is exactly identical to the Harry Potter$_{8pm}$ with only the time difference. Should the choices of Harry Potter$_{8pm}$ and Transformer$_{9pm}$ be equally affected, meaning that the the probability of choice alternatives of both Harry Potter$_{8pm}$ and Transformer$_{9pm}$ now be reduced to 0.33? Intuitively, the answer is no. In fact, it would make more sense the probability of Transformer stays the same at 0.5 and the probabilities of Harry Potter$_{8pm}$ and Harry Potter$_{9pm}$ are 0.25. However, because of the

IIA property, the odds of Harry Potter$_{8pm}$ and Transformer$_{9pm}$ is supposed to remain unaffected, implying that the probabilities of Harry Potter$_{8pm}$ and Transformer$_{9pm}$ are affected equally by Harry Potter$_{10pm}$.

In general, IIA property fails to provide an intuitive substitution pattern. The multinomial nested logit model, in chapter 4, is one of the models which relaxes the IIA property by imposing some hierarchical nesting structure on the choice.

3.2 Movie Forecasting Model

In the current project, choice at the individual level is not observed. Instead, aggregated choice outcomes at individual showing levels are observed. On the other hands, multiple choice occasions are observed where each choice occassion is defined by a day, d. Therefore, in estimation, a subscription d should be added. Figure 3.1 is an example of the structure of standard logit model on a given date d. On a given date d, there are three movies showing: A, B, C, and the outside option, O (i.e. not watching any of the movie showings). Subscript of movies A, B and C represent the hour at each movie start. When one extra movie is squeesed in on a given day d, say D_{2pm}, it will equally attract people from the outside option and the other alternatives, A_{6pm}, A_{8pm}, B_{6pm}, B_{9pm}, C_{4pm} and O because of IIA property. Note that the substitution from outside option is what is called demand expansion and the other alternatives are the cannibalization. However, IIA makes these two effects too restrictive.

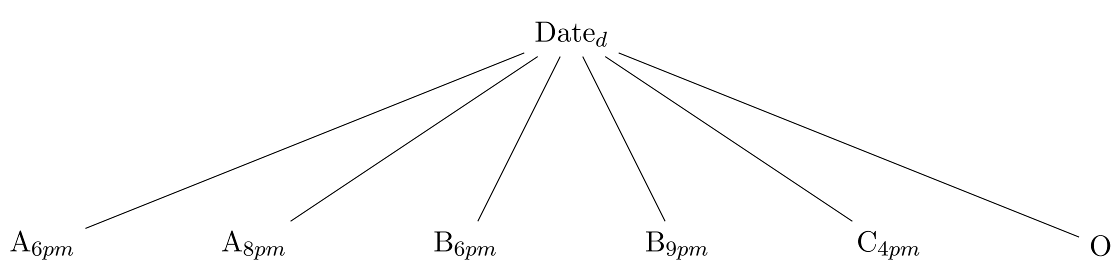

Figure 3.1: Example of standard logit structure on a given day d. On a given day d, there are choice sets of movie and hour combinations as well as not watching movie, which is the outside option.

The standard logit model, P_g, of the probability of choosing choice alternative g in the set containing all movie showings plus the outside option on date d, C_d, is defined as

$$P_g = \frac{\exp\left(\alpha_g X_g + \omega Y\right)}{\sum_{g \in C_d} \exp\left(\alpha_g X_g + \omega Y\right)} \tag{3.4}$$

There are some constraints in this model.

1. $\sum_{g \in C_d} P_g = 1$: the sum of probability of choice alternatives including outside options on a given day are 1.

2. $P_0 = 1 - \sum_{g \in C_d^*} P_g$: the probability of choosing outside options is 1 minus the sum of probability of choice alternatives excluding the outside options on a given day.

3. $\sum_{g \in C_d} S_g = M$: The total number of movie ticket sales and the number of people choosing outside option in a given date d is the population of the market which is Amsterdam, M and it is the market capacity.

3.3 Posterior Distributions

Since there are a finite number of movie showings at each day, the multinomial distribution is appropriate to represent the likelihood function.

Likelihood function

$$[S_g \,,\, g \in C_d \,|\, \alpha_g \,,\, \omega] = \prod_d \frac{M!}{\prod_{g \in C_d} S_g!} \prod_{g \in C_d} P_g^{\;S_g} \tag{3.5}$$

Prior and Hyper Prior Distributions

$$[\alpha_{jk} \,|\, \eta,\, \Sigma] \sim \text{Multivariate Normal}(\eta\, Z_{jk}\,,\, \Sigma) \tag{3.6}$$

$$[\omega] \sim \text{Multivariate Normal}(0\,,\, \text{I}) \tag{3.7}$$

$$[\eta] \sim \text{Multivariate Normal}(0\,,\, \text{I}) \tag{3.8}$$

$$[\Sigma] \sim \text{Inverse Wishart}(16\,,\, \text{M}) \tag{3.9}$$

where M is defined as in linear regression model in Chapter 2

Because of the prior distribution (3.6) to (3.9) are not conjugate to multinomial distribution, the full conditional posterior distribution cannot be derived and hence Metropolis Hastings algorithm is used to estimate α_{jk} and ω. However, Gibbs sampling can be still applied for the hyper parameters η and Σ. The detailed derivations of the posterior distributions follow. Note that the results of the predictions are left for Chapter 6 so as to make a comparison between models.

3.3.1 Posterior Distribution of $[\alpha_g \mid \cdot]$

$$
\begin{aligned}
[\alpha_g \mid \cdot] &\propto [S_g \mid \alpha_g, \omega] \cdot [\alpha_g \mid \eta, \Sigma] \\
&\propto \prod_d \left[\frac{M!}{\prod_{g \in C_d} S_g!} \prod_{g \in C_d} \left(\frac{\exp(\alpha_g X_g + \omega Y}{1 + \sum_{g \in C_d} \exp(\alpha_g X_g + \omega Y)} \right)^{S_g} \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g]) \\
&\propto \prod_d \left[\prod_{g \in C_d} \left(\frac{\exp(\alpha_g X_g + \omega Y}{1 + \sum_{g \in C_d} \exp(\alpha_g X_g + \omega Y)} \right)^{S_g} \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g]) \\
&\propto \prod_d \left[\prod_{g \in C_d} \left(\frac{\exp(S_g(\alpha_g X_g + \omega Y))}{\left(1 + \sum_{g \in C_d} \exp(\alpha_g X_g + \omega Y)\right)^{S_g}} \right) \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g]) \\
&= \prod_d \left[\frac{\exp\left(\sum_{g \in C_d} S_g(\alpha_g X_g + \omega Y) \right)}{\left(1 + \sum_{g \in C_d} \exp(\alpha_g X_g + \omega Y)\right)^{\sum_{g \in C_d} S_g}} \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g]) \\
&= \frac{\exp\left(\sum_d \sum_d S_g(\alpha_g X_g + \omega Y) - \frac{1}{2} \alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g] \right)}{\left(1 + \sum_{g \in C_d} \exp\left(\alpha_g X_g + \omega Y\right)\right)^{\sum_{g \in C_d} S_g}} \qquad (3.10)
\end{aligned}
$$

3.3.2 Posterior Distribution of $[\omega \,|\, \cdot\,]$

$$
\begin{aligned}
[\omega \,|\, \cdot\,] &\propto [S_g \,|\, \alpha_g, \omega] \cdot [\omega] \\
&\propto \prod_d \left[\frac{M!}{\prod\limits_{g \in C_d} S_g!} \prod_{g \in C_d} \left(\frac{\exp(\alpha_g X_g + \omega Y}{1 + \sum\limits_{g \in C_d} \exp(\alpha_g X_g + \omega Y)} \right)^{S_g} \right] \exp(-\frac{1}{2}\omega' I^{-1}\omega) \\
&\propto \prod_d \left[\prod_{g \in C_d} \left(\frac{\exp(\alpha_g X_g + \omega Y}{1 + \sum\limits_{g \in C_d} \exp(\alpha_g X_g + \omega Y)} \right)^{S_g} \right] \exp(-\frac{1}{2}\omega' I^{-1}\omega) \\
&\propto \prod_d \left[\prod_{g \in C_d} \left(\frac{\exp(S_g(\alpha_g X_g + \omega Y))}{\left(1 + \sum\limits_{g \in C_d} \exp(\alpha_g X_g + \omega Y)\right)^{S_g}} \right) \right] \exp(-\frac{1}{2}\omega' I^{-1}\omega) \\
&= \prod_d \left[\frac{\exp\left(\sum\limits_{g \in C_d} S_g(\alpha_g X_g + \omega Y) \right)}{\left(1 + \sum\limits_{g \in C_d} \exp(\alpha_g X_g + \omega Y)\right)^{\sum\limits_{g \in C_d} S_g}} \right] \exp(-\frac{1}{2}\omega' I^{-1}\omega) \\
&= \frac{\exp\left(\sum\limits_d \sum\limits_d S_g(\alpha_g X_g + \omega Y) - \frac{1}{2}\omega' I^{-1}\omega \right)}{\left(1 + \sum\limits_{g \in C_d} \exp\left(\alpha_g X_g + \omega Y\right)\right)^{\sum\limits_{g \in C_d} S_g}} \qquad (3.11)
\end{aligned}
$$

3.3.3 Full Conditional Distribution of $[\eta^* \,|\, \cdot\,]$

$$\begin{aligned} [\eta^* \,|\, \cdot\,] &\propto [\alpha_g^* \,|\, \eta^*, \Sigma] \cdot [\eta^*] \\ &\sim \text{MVN}_{15\cdot 6}([Z'Z \otimes \Sigma^{-1} + I]^{-1}[(Z' \otimes \Sigma^{-1}) \cdot \alpha_g^*], [Z'Z \otimes \Sigma^{-1} + I]^{-1}) \end{aligned} \quad (3.12)$$

Note: the derivation for η^* is in (2.12)

3.3.4 Full Conditional Distribution of $[\Sigma \,|\, \cdot\,]$

$$\begin{aligned} [\Sigma \,|\, \cdot\,] &\propto [\alpha_g^* \,|\, \eta^*, \Sigma] \cdot [\Sigma] \\ &\sim \text{IW}\left(N + a\,, [X_g' X_g]^{-1} I + \sum_{g \in C_d} [\alpha_g - \eta Z_g][\alpha_g - \eta Z_g]'\right) \end{aligned} \quad (3.13)$$

Note: the derivation for η^* is in (2.13)

Chapter 4

Nested Logit Model

Since the standard logit model has unintuitive substitution pattern due to IIA property, as discussed earlier, a nested logit model is used. Nested logit model relaxes the IIA property by grouping (or nesting) similar alternatives together so that within the nest, IIA property holds but not across the nest [1].

4.1 Movie Forecasting Model

The nested logit model is applied when the choice set can be sub-divided into several subsets where elements in each subset are relatively homogeneous [1]. Continuing the same example from 3.2, the choice set, A_{6pm}, A_{8pm}, B_{6pm}, B_{9pm}, C_{4pm}, and O, can be derived into four subsets: C_n^A, C_n^B, C_n^C and C_n^O. C_n^A is choice set of movie A for individual n. C_n^B is choice set of movie B for individual n. C_n^C is choice set of movie C for individual n. C_n^O is choice set of outside option O for individual n. For each subset of movies, there are different showings which have different starting time. Figure 4.1 is an example of the nested logit structure on a given day d. Contrast to Figure 3.1 in the standard logit model, the nested logit model has one more level and it is grouping by movies.

In the nested logit model, P_g as the probability of choosing choice alternative g which is in a set, C_d, containing all movie showings plus the outside option on day d can be decomposed into three components:

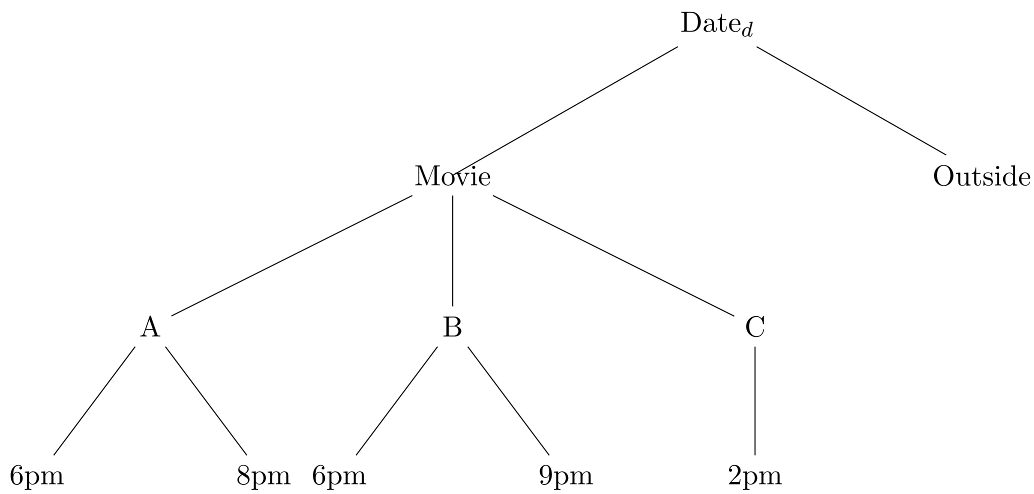

Figure 4.1: Example of nested logit structure on a given day d. On a given day d, there are two choice sets: (1) watching a movie at a theater or (2) doing other activity. For the choice set of movie (1), it can be further broken down into first, different kinds of movies and second, different showings of the particular movie that is chosen.

$$
\begin{aligned}
P_g \quad = \quad & P(\text{hour } h \,|\text{movie } j, \text{ version } k, \text{ date } d) \\
& \times P(\text{movie } j, \text{version } k \,|\, \text{any movie}, \text{ date } d) \\
& \times P(\text{any movie} \,|\, \text{ date } d)
\end{aligned} \tag{4.1}
$$

$$
P(\text{hour } h \,|\text{movie } j, \text{ version } k, \text{ date } d) = \frac{\exp(\sigma_{jk} V_{jkhd})}{\exp(IV_{jkd})} \tag{4.2}
$$

$$
P(\text{movie } j, \text{version } k \,|\, \text{any movie}, \text{ date } d) = \frac{\exp\left(\frac{\sigma}{\sigma_{jk}} IV_{jkd}\right)}{\exp(IV_d^{mv})} \tag{4.3}
$$

$$
P(\text{any movie} \,|\, \text{ date } d) = \frac{\exp(\frac{1}{d} IV_d^{mv})}{1 + \frac{1}{\sigma} IV_d^{mv}} \tag{4.4}
$$

where,

$$
\begin{aligned}
\sigma_{jk} &= \text{parameter capturing the substitution of alternatives of the same version } k \text{ of movie } j \\
\sigma &= \text{parameter capturing the substitution of all movie alternatives} \\
C_{h|jkd} &= \text{set containing all hours available on version } k \text{ of movie } j \text{ on date } d \\
C_{jk|d} &= \text{set containing all movie version } k \text{ of movie } j \text{ on date } d \\
V_{jkhd} &= \alpha_{jk} X_{jkhd} + \omega Y_{jkd} \\
IV_{jkhd} &= \log\left(\sum_{g \in C_{h|jkd}} \exp(\sigma_{g_{jk}} V_{g_{jkhd}})\right) \\
IV_d^{mv} &= \log\left(\sum_{g \in C_{jk|d}} \exp\left(\frac{\sigma}{\sigma_{g_{jk}}} IV_{g_{jkhd}}\right)\right)
\end{aligned}
$$

The major difference between the nested logit model and the standard logit model is that imposing nesting structure would reduce the unintuitive substitution implied by IIA. In other words, the utility of alternatives in a nested logit model is no longer uncorrelated. While $P(\text{hour } h \,|\text{movie } j,\ \text{version } k,\ \text{date } d)$ in (4.2) has the IIA property, $P(\text{movie } j, \text{version } k \,|\, \text{any movie, date } d)$ in (4.3), captures the correlation of the same movie with different hours. Therefore, the nested logit model has IIA property within nests but not across nests and thus has more intuitive substitution effects than the standard logit model.

In the nested logit model, some new parameters are introduced: σ_{jk} and σ. Parameters σ_{jk} capture the substitution of alternatives of the same version k of movie j, while σ captures the substitution of all movie alternatives with reference to the outside option. However, their interpretation is opposite to the traditional correlation coefficient. While the value of these substitution parameters are between 0 and 1, when σ_{jk} is equal to 0, it means that each nest has a perfect substitution. Let's use the same example illustrated in Figure 4.1. When extra movie A at 10pm, A_{10pm}, is squeezed in on a given day d and if σ_{jk} is close to 0, it would not affect movie B and movie C at all. However, when σ_{jk} close to 1, it gets back to IIA property. That is, adding extra movie A_{10pm} affects movie A, movie B and movie

C equally. The same idea applies for σ as when σ equals to 1, the structure in Figure 4.1 reduces to a two-level structure which has the IIA property in Figure 3.1.

4.2 Posterior Distributions

Likelihood Function

$$[S_g, g \in C_d \,|\, \alpha_g, \omega, \sigma_g, \sigma] = \prod_d \left[\frac{M!}{S_g!} \prod_{g \in C_d} \left[\frac{\exp(\sigma_g V_{g_{jkhd}})}{\exp(IV_{g_{jkd}})} \times \frac{\exp\left(\frac{\sigma}{\sigma_g} IV_{g_{jkd}}\right)}{\exp\left(IV_{g_d}^{mv}\right)} \times \frac{\exp\left(\frac{1}{\sigma} IV_{g_d}^{mv}\right)}{1 + \exp\left(\frac{1}{\sigma} IV_{g_d}^{mv}\right)} \right]^{S_g} \right] \tag{4.5}$$

Prior Distributions

$$\begin{aligned} \alpha_g &\sim \text{MVN}_6(\eta Z_g, \Sigma) \\ \omega &\sim \text{MVN}_{34}(0, I) \\ \sigma_g &\sim \text{Uniform}(0,1) \\ \sigma &\sim \text{Uniform}(0,1) \end{aligned}$$

4.2.1 Posterior Distribution of $[\alpha_g \,|\, \cdot]$

$$
\begin{aligned}
[\alpha_g \,|\, \cdot] &\propto [S_g \,|\, \alpha_g, \omega, \sigma_g, \sigma] \cdot [\alpha_g \,|\, \eta, \Sigma] \\
&= \prod_d \left[\frac{M!}{S_g!} \prod_{g \in C_d} \left[\frac{\exp(\sigma_g V_{g_{jkhd}})}{\exp(IV_{g_{jkd}})} \cdot \frac{\exp\left(\frac{\sigma}{\sigma_g} IV_{g_{jkd}}\right)}{\exp\left(IV^{mv}_{g_d}\right)} \cdot \frac{\exp\left(\frac{1}{\sigma} IV^{mv}_{g_d}\right)}{1 + \exp\left(\frac{1}{\sigma} IV^{mv}_{g_d}\right)} \right]^{S_g} \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g]) \\
&\propto \prod_d \left[\prod_{g \in C_d} \left[\frac{\exp\left(\sigma_g V_{g_{jkhd}} + \frac{\sigma}{\sigma_g} IV_{g_{jkd}} + \frac{1}{\sigma} IV^{mv}_{g_d} - IV_{g_{jkd}} - IV^{mv}_{g_d}\right)}{1 + \exp(\frac{1}{\sigma} IV^{mv}_{g_d})} \right]^{S_g} \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g]) \\
&= \prod_d \left[\prod_{g \in C_d} \left[\frac{\exp\left(S_g \left(\sigma_g V_{g_{jkhd}} + \frac{\sigma}{\sigma_g} IV_{g_{jkd}} + \frac{1}{\sigma} IV^{mv}_{g_d} - IV_{g_{jkd}} - IV^{mv}_{g_d}\right)\right)}{1 + \exp\left(\frac{1}{\sigma} IV^{mv}_{g_d}\right)^{S_g}} \right] \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g]) \\
&= \prod_d \left[\frac{\exp\left[\sum_{g \in C_d} S_g \left(\sigma_g V_{g_{jkhd}} + \frac{\sigma}{\sigma_g} IV_{g_{jkd}} + \frac{1}{\sigma} IV^{mv}_{g_d} - IV_{g_{jkd}} - IV^{mv}_{g_d}\right) \right]}{\left[1 + \exp\left(\frac{1}{\sigma} IV^{mv}_{g_d}\right)\right]^{\sum_{g \in C_d} S_g}} \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\alpha_g - \eta Z_g]' \Sigma^{-1} [\alpha_g - \eta Z_g])
\end{aligned}
$$

Note: let $S_g \left(\sigma_g V_{g_{jkhd}} + \frac{\sigma}{\sigma_g} IV_{g_{jkd}} + \frac{1}{\sigma} IV^{mv}_{g_d} - IV_{g_{jkd}} - IV^{mv}_{g_d}\right)$ be A

$$
\propto \frac{\exp\left[\left(\sum_d \sum_{g \in C_d} A \right) - \frac{1}{2}[\alpha_g' \Sigma^{-1} \alpha_g - 2\alpha_g' \Sigma^{-1} \eta Z_g] \right]}{\prod_d \left[1 + \exp\left(\frac{1}{\sigma} IV^{mv}_{g_d} \right) \right]^{\sum_{g \in C_d} S_g}} \tag{4.6}
$$

4.2.2 Posterior Distribution of $[\omega \,|\, \cdot\,]$

$$
\begin{aligned}
[\omega \,|\, \cdot\,] &\propto [S_g \,|\, \alpha_g, \omega, \sigma_g, \sigma] \cdot [\omega] \\
&= \prod_d \left[\frac{M!}{S_g!} \prod_{g \in C_d} \left[\frac{\exp(\sigma_g V_{g_{jkhd}})}{\exp(IV_{g_{jkd}})} \cdot \frac{\exp\left(\frac{\sigma}{\sigma_g} IV_{g_{jkd}}\right)}{\exp\left(IV_{g_d}^{mv}\right)} \cdot \frac{\exp\left(\frac{1}{\sigma} IV_{g_d}^{mv}\right)}{1+\exp\left(\frac{1}{\sigma} IV_{g_d}^{mv}\right)} \right]^{S_g} \right] \\
&\quad \cdot \exp(-\frac{1}{2}[\omega]' I^{-1} [\omega])
\end{aligned}
$$

Note: same derivation of likelihood fn is applied as in $[\alpha_g \,|\, \cdot\,]$

$$
\propto \frac{\exp\left[\left(\sum_d \sum_{g \in C_d} A\right) - \frac{1}{2}(\omega' I^{-1} \omega)\right]}{\prod_d \left[1 + \exp\left(\frac{1}{\sigma} IV_{g_d}^{mv}\right)\right]^{\sum_{g \in C_d} S_g}} \tag{4.7}
$$

4.2.3 Posterior Distribution of $[\sigma_{jk} \,|\, \cdot\,]$

$$
\begin{aligned}
[\sigma_{jk} \,|\, \cdot\,] &\propto [S_g \,|\, \alpha_g, \omega, \sigma_{jk}, \sigma] \cdot [\sigma_{jk}] \\
&= \prod_d \left[\frac{M!}{S_g!} \prod_{g \in C_d} \left[\frac{\exp(\sigma_{g_{jk}} V_{g_{jkhd}})}{\exp(IV_{g_{jkd}})} \cdot \frac{\exp\left(\frac{\sigma}{\sigma_{g_{jk}}} IV_{g_{jkd}}\right)}{\exp\left(IV_{g_d}^{mv}\right)} \cdot \frac{\exp\left(\frac{1}{\sigma} IV_{g_d}^{mv}\right)}{1+\exp\left(\frac{1}{\sigma} IV_{g_d}^{mv}\right)} \right]^{S_g} \right] \\
&\quad \cdot I(0 < \sigma_{jk} < 1)
\end{aligned}
$$

Note: same derivation of likelihood fn is applied as in $[\alpha_g \,|\, \cdot\,]$

$$
\propto \frac{\exp\left[\sum_d \sum_{g \in C_d} A\right]}{\prod_d \left[1 + \exp\left(\frac{1}{\sigma} IV_{g_d}^{mv}\right)\right]^{\sum_{g \in C_d} S_g}} \tag{4.8}
$$

4.2.4 Posterior Distribution of $[\sigma \mid \cdot\,]$

$$\begin{aligned}[\sigma \mid \cdot\,] &\propto [S_g \mid \alpha_g, \omega, \sigma_{jk}, \sigma] \cdot [\sigma] \\ &= \prod_d \left[\frac{M!}{S_g!} \prod_{g \in C_d} \left[\frac{\exp(\sigma_{g_{jk}} V_{g_{jkhd}})}{\exp(IV_{g_{jkd}})} \cdot \frac{\exp\left(\frac{\sigma}{\sigma_{g_{jk}}} IV_{g_{jkd}}\right)}{\exp\left(IV^{mv}_{g_d}\right)} \cdot \frac{\exp\left(\frac{1}{\sigma} IV^{mv}_{g_d}\right)}{1 + \exp\left(\frac{1}{\sigma} IV^{mv}_{g_d}\right)} \right]^{S_g} \right] \\ &\quad \cdot I(0 < \sigma < 1)\end{aligned} \tag{4.9}$$

Note: same derivation of likelihood fn is applied as in $[\alpha_g \mid \cdot\,]$

$$\propto \frac{\exp\left[\sum_d \sum_{g \in C_d} A\right]}{\prod_d \left[1 + \exp\left(\frac{1}{\sigma} IV^{mv}_{g_d}\right)\right]^{\sum_{g \in C_d} S_g}} \tag{4.10}$$

4.2.5 Full Conditional Distribution of $[\eta^* \mid \cdot\,]$

$$\begin{aligned}[\eta^* \mid \cdot\,] &\propto [\alpha^*_g \mid \eta^*, \Sigma] \cdot [\eta^*] \\ &\sim \mathrm{MVN}_{15 \cdot 6}([Z'Z \otimes \Sigma^{-1} + I]^{-1}[(Z' \otimes \Sigma^{-1}) \cdot \alpha^*_g], [Z'Z \otimes \Sigma^{-1} + I]^{-1})\end{aligned} \tag{4.11}$$

4.2.6 Full Conditional Distribution of $[\Sigma \,|\, \cdot\,]$

$$
\begin{aligned}
[\Sigma \,|\, \cdot\,] &\propto [\alpha_g^* \,|\, \eta^*, \Sigma] \cdot [\Sigma] \\
&\sim \text{IW}\left(N + a\,, [X_g' X_g]^{-1} I + \sum_{g \in C_d} [\alpha_g - \eta Z_g][\alpha_g - \eta Z_g]' \right)
\end{aligned} \tag{4.12}
$$

Note that the results of the posterior distributions are left for Chapter 6 so as to make a comparison between models.

Chapter 5

Data

5.1 Description of Data

The attendance data are obtained from PATHE, one of the largest multiplex movie theater companies in Netherlands. The raw data set contains one-year of data from 2008 for a multiplex theater located in Amsterdam, including movie showing information such as (1) when the showing started, (2) how long ago the movies were first released, (3) whether it is played during holidays, (4) what day of week is the showing on and (5) the number of tickets sold. Furthermore, the characteristics of movies such as genre and age restriction are contained in the data set. A sample of data is shown in Table 5.1. In the data set, different language version of the same movies are treated as two different movies. Also, data with the same movies with same hour showing are aggregated to one observation. For example, when Harry Potter is playing at 9pm and 9:30pm, the ticket sales of the two showings are combined into one movie ticket sales with the starting time of showing set at 9pm. The reason for such aggregation of the same movie in the same hour is to compare the three models' predictions, which include a linear regression model. This is in line with how Elishberg et al. handle the data for the linear regression model.

Since Holiday, Day of Week, Age Restriction and Genre variables are dummy variables, a base case for each corresponding variable needs to be set. Therefore, the base case for Holidays is the normal days between the Easter holidays and the school May vacation, the base case for Day of Week is Saturday, the base case for Age Restriction is all ages and the base case for Genre is action movie.

Showdate	Slot	Movie Name	Language Version	Movie Sub	S_paid	Age Restriction	Holidays	Age	Day of Week	Genre
1/1/2008	13	Bee Movie	NL	1	59	R_6	New Year	2	Tues	Kids
1/1/2008	16	Bee Movie	OV	2	23	R_6	New Year	2	Tues	Kids
1/1/2008	12	Elizabeth	OV	2	4	R_12	New Year	1	Tues	Drama
1/1/2008	15	Elizabeth	OV	2	15	R_12	New Year	1	Tues	Drama

Table 5.1: Sample of Data

5.2 Implementation

At first, the hierarchical linear regression was implemented by the statistical programming language R with 1 year of data. There were over 2000 parameters to be estimated, but since Gibbs sampling was used in the linear model, R can handle it. However, due to the complicated likelihood function in the standard logit and nested logit models, R cannot handle such a massive implementation of MCMC as it would take a few years to run. R is notoriously bad with multiple for-loops and unfortunately, in the likelihood of standard logit and nested logit models, there are several for-loops. Therefore, for the standard logit and nested logit, I learned and used the programming language C. Surprisingly C is about 180 times faster than R. However, there are more than 2000 parameters needed to be estimated, and even C would take a few months of computation time for reasonable results. Therefore, in this project, only 2 weeks of data from January 10, 2008 to January 23, 2008 are used for the parameter estimations. Then for prediction purposes, one week of data January 24, 2008 to January 30, 2008 is used and predictions are compared to actual data. For the linear regression model, 60000 MCMC iterations are used and for the standard logit and nested logit model, 100000 iterations are used.

Chapter 6

Result

For the predictions of movie ticket sales, the posterior predictive distributions are used: 5000 iterations are drawn from posterior distributions after burn-in period and used toward the calculations of ticket sales. For the linear regression model, log ticket sales are directly calculated from parameter estimations. For the standard and nested logit model, since the likelihood functions are multinomial distributions, the probability of watching a specific movie is calculated and then multiplied by the population of Amsterdam to get the predicted ticket sales. However, since there are 5 new movies in the week from January 24 to January 30, 2008 and there are no parameter estimations for movie specific parameters, α_{jk}, the hyper parameters, η and Σ are used to calculate movie specific parameters, α_{jk}.

6.1 Actual Movie Ticket Sales vs. Predicted Median Movie Ticket Sales

In order to compare the three models' predictions, the actual movie ticket sales and predicted median of movie ticket sales are compared. If the predictions are good, then the actual and predicted median should align with a 45 degree line (Figure 6.1). Then the coefficients of determination (R^2) of three models are compared in Table 6.1 to find the best model out of linear, standard logit and nested logit model since R^2 measures the proportion of the total variation [2]. Since predictions for new movies are based on less information, the predictions in Figure 6.1 show new and previously viewed movies in different colours. A detailed discussion of the results is in Chapter 7.

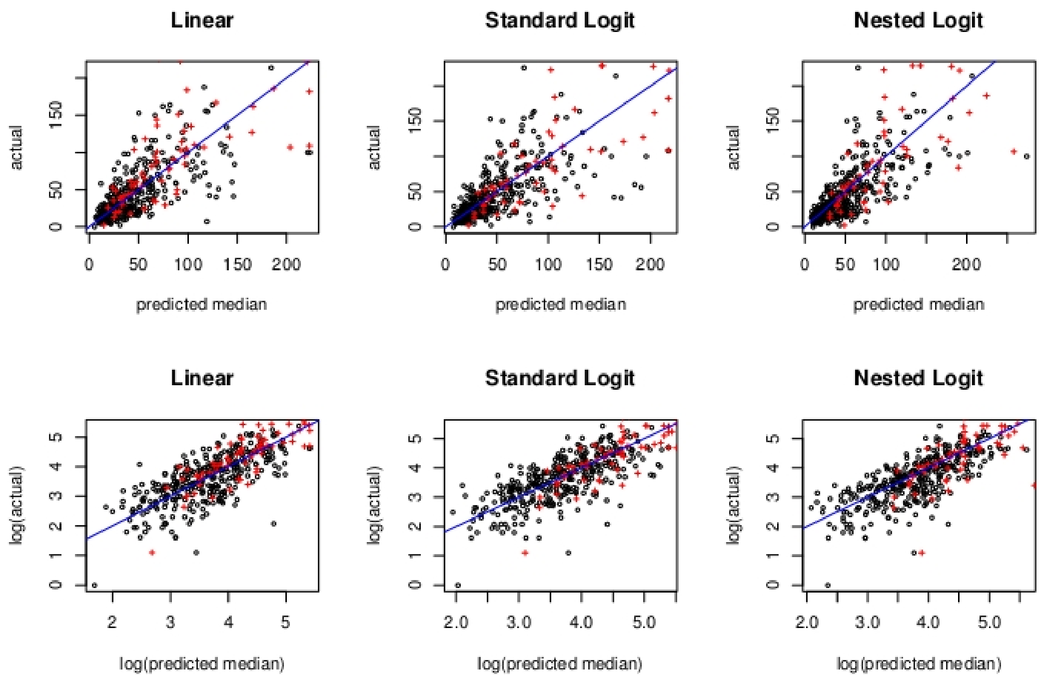

Figure 6.1: Scatter plot of actual movie ticket sales vs predicted median of movie ticket sales in the period of January 24, 2008 to Jan 30, 2008. Top graphs are actual scale of movie ticket sales and the bottom graphs are log scales of movie ticket sales. Black dots are predictions for existing movies and red plus signs are predictions for new movies. The blue line is the reference line for the perfect match of the actual and the predicted ticket sales.

Model	Linear	Standard Logit	Nested Logit
Existing Movies	0.67	0.63	0.70
New Movies	0.66	0.66	0.29
All Movies	0.71	0.72	0.59

Table 6.1: R^2 of three models on existing, new and all movies. By looking at R^2 for all movies, linear regression model and standard logit model are much better than nested logit model. However, when the movies are broken down into new and existing movies, the nested logit model is the clear winner. A more detailed discussion of the results is in Chapter 7.

6.2 Posterior Predictive Distribution for Existing Movies

Since there are 409 combinations of movie showings in the one week period of January 24, 2008 to January 30, 2008, three existing movies (i.e, Bee Movie, The Nanny Diaries and Moordwijven) are selected as an example. The posterior predictive distributions of those three selected movies with different days of week and hours of showings are in Figure 6.2 to Figure 6.4.

1. Bee Movie: kid's movie with age restriction R_6 and predictions for week 6 (Figure 6.2).
   (a) Saturday at 10am
   (b) Saturday at 2pm
   (c) Wednesday at 1pm
2. The Nanny Diaries: commedy movie with no age restriction R_{all} and predictions for week 5 (Figure 6.3).
   (a) Friday at 5pm
   (b) Sunday at 5pm
   (c) Monday at 12pm
3. Moordwijven: miscellaneous movie with age restriction R_{12} and predictions for week 2 (Figure 6.4).
   (a) Thursday at 10pm
   (b) Saturday at 9pm
   (c) Sunday at 10am

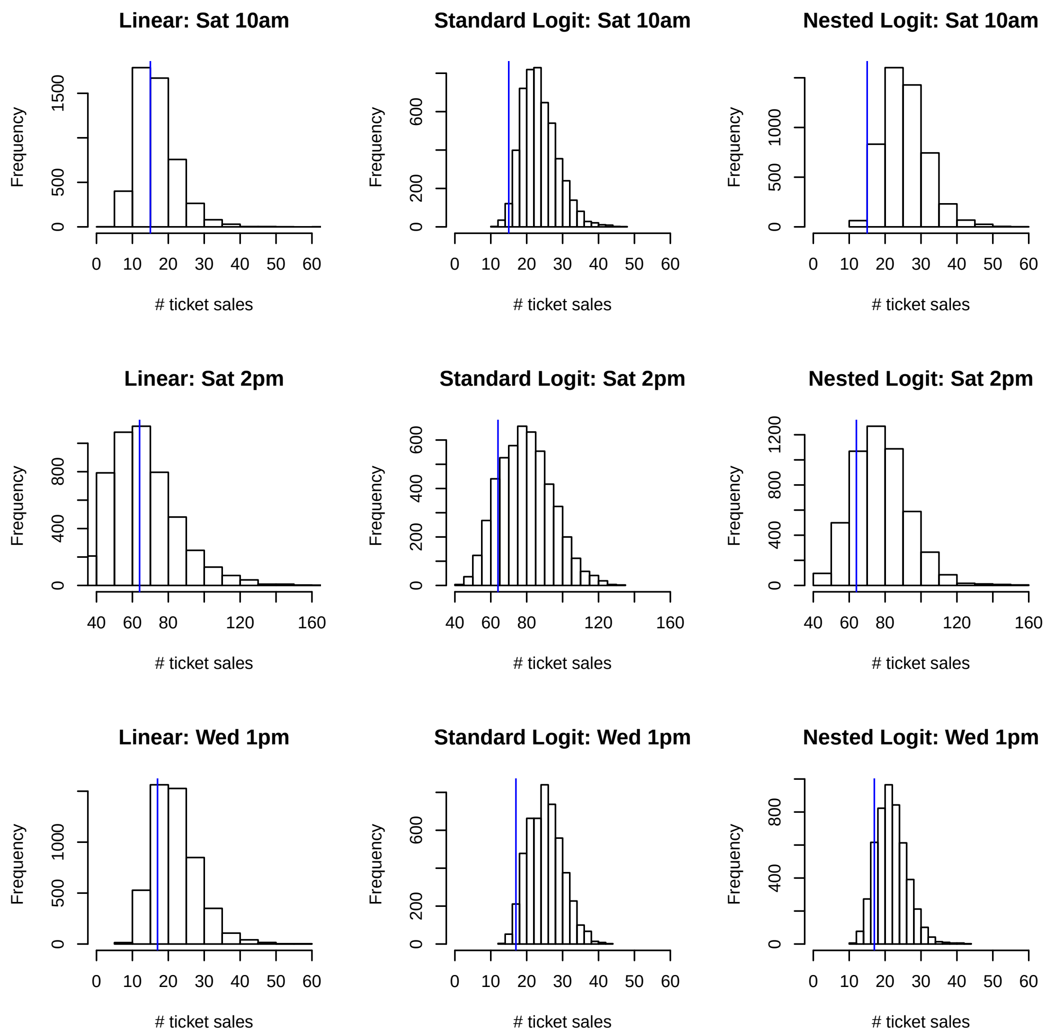

Figure 6.2: Posterior predictive distributions of Bee Movie on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales.

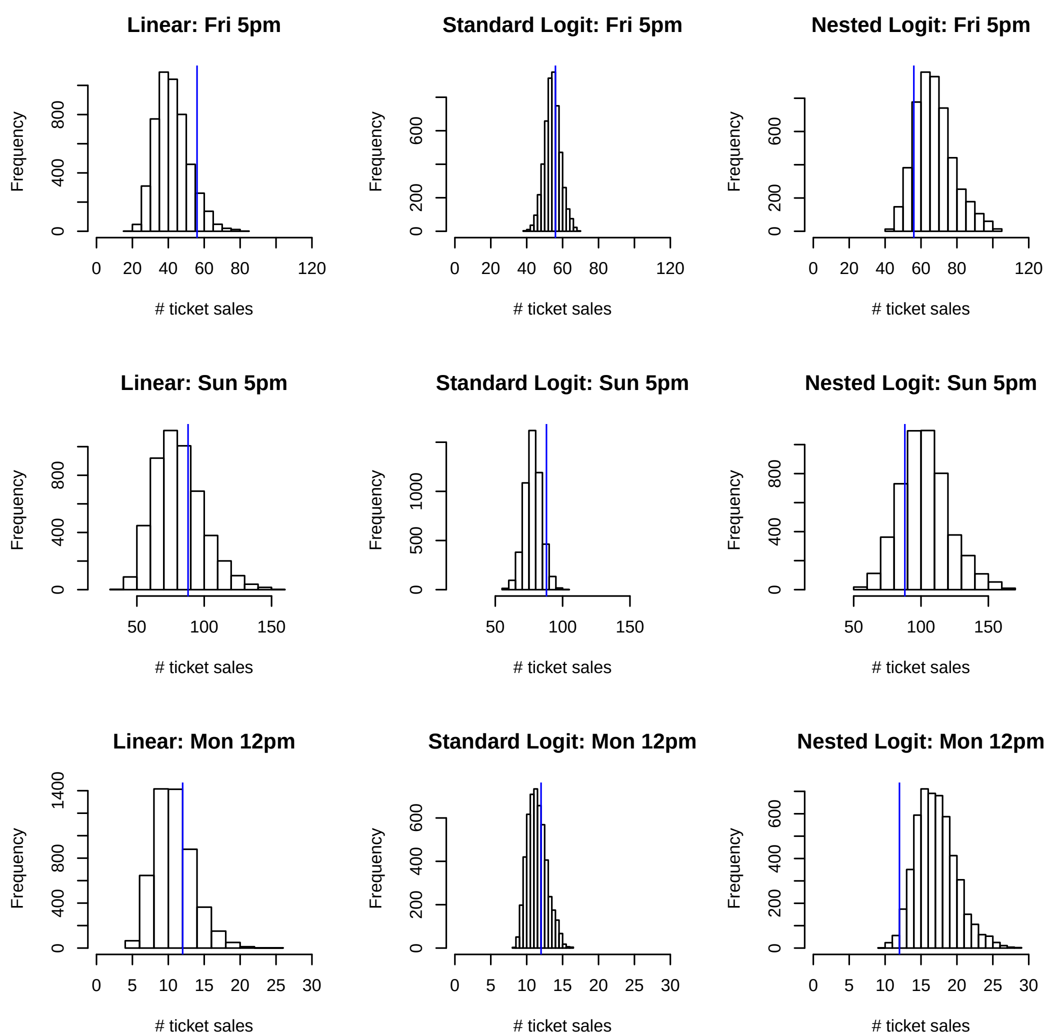

Figure 6.3: Posterior predictive distributions of The Nanny Diaries on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales.

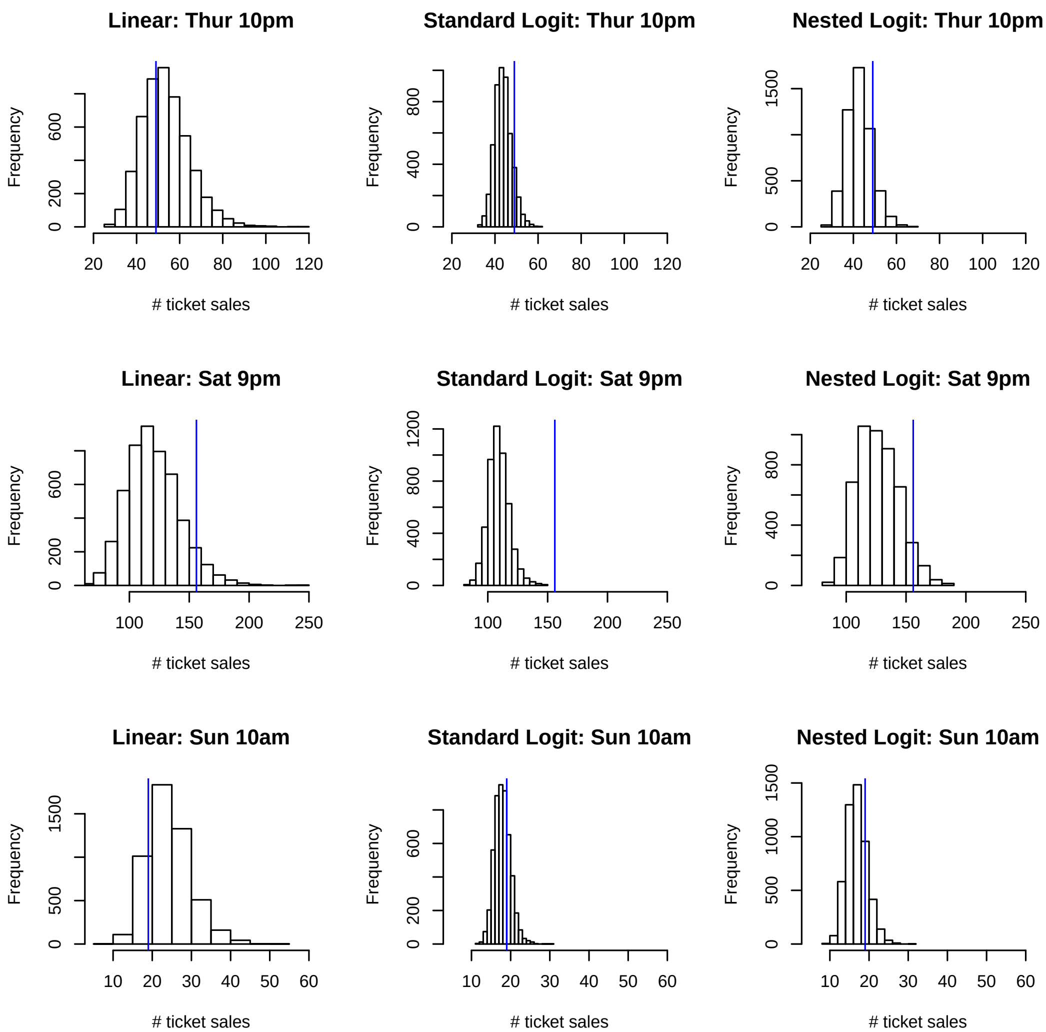

Figure 6.4: Posterior predictive distributions of Moordwijven on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales.

6.3 Posterior Predictive Distributions for New Movies

Two new movies, Cloverfield and We Own the Night, with different day of the week and hour of showings are selected as an example. Note that in the one week period of January 23, 2008 to January 30, 2008, 5 new movies are released.

1. Cloverfield: action movie with age restriction R_{16} and predictions for week 1 (Figure 6.5).
   (a) Friday at 12pm
   (b) Sunday at 9pm
   (c) Monday at 7pm
2. We Own the Night: drama movie with age restriction R_{16} and predictions for week 1 (Figure 6.6).
   (a) Thursday at 9pm
   (b) Friday at 9pm
   (c) Wednesday at 9pm

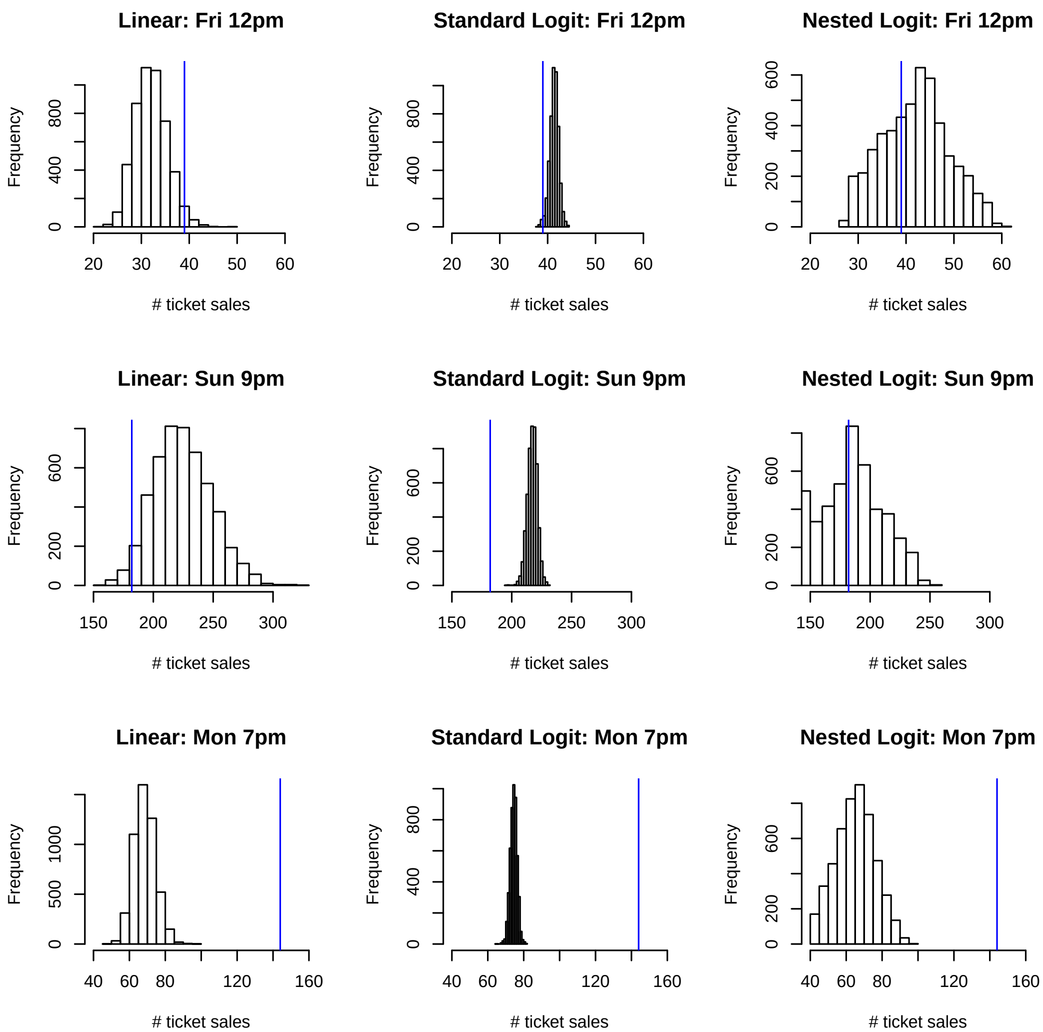

Figure 6.5: Posterior predictive distributions of Cloverfield on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales.

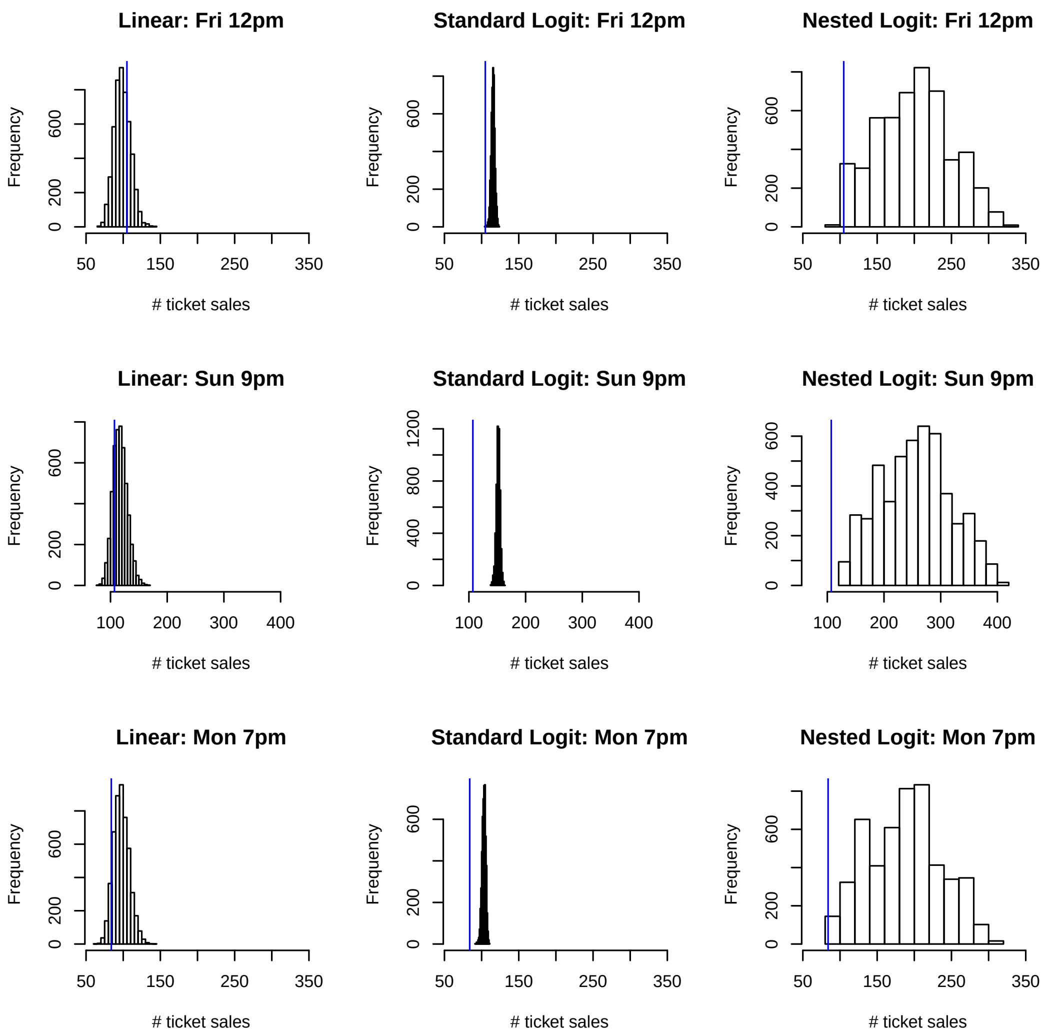

Figure 6.6: Posterior predictive distributions of We Own the Night on selected day of week and hour of showings from linear, standard logit and nested logit model. The blue line represents the actual ticket sales.

Chapter 7

Conclusion and Recommendation

As shown in the previous section, the predictions from three different models (linear regression, standard logit and nested logit models) are not as good as expected. However, the movie predictions for existing movies are good compared to the new movies. The reason is that only two weeks of data are used to predict the next weeks ticket sales. Two weeks are too short period of time to capture the trend of age decay and hour of movie showing effects. Among the three models discussed in this project, there is no superior model for predicting the movie ticket sales as seen in scatter plot of actual vs. predicted ticket sales (Figure 6.1), all three models' predictions are behaving very similarly and on par predict decently. It is obvious that the predictions of existing movies are better than the new movies' predictions (Figure 6.1). The reason is that the new movies' predictions are based on the movies' genre and age restrictions and two weeks of data is too short to capture the effects of genre and age restrictions for all combinations.

By looking at the three models R^2, linear regression model and standard logit model predict better than nested logit model for all movies (Table 6.1). However, when looking at only existing movies, the nested logit model has better predictions because the model captures better market expansion as well as cannibalization effects. The reason why the nested logit model's R^2 for all movies is smallest is because it predicts new movies poorly. That is, the nested logit model captures both market expansion and cannibalization effects well for existing movies but not new movies. The major reason why nested the logit model fails to give a good prediction for the new movies is the uninformative prior imposed on the parameter capturing the substitution of alternatives of the same movie, σ_{jk}. In the nested logit model, instead of using an uninformative prior on σ_{jk}, beta distributions with hyperparameters related to genre and age restriction covariates can be used to improve model predictions. Currently predicted values of σ_{jk} for new movies are based on the prior of uniform distribution between 0 and 1 meaning that any substitution behaviour is possible. The derivations of posterior distribution of σ_{jk} is in Appendix A.

Also, predictions for both existing and new movies can be improved if one year of data is used. It will potentially improve the hierarchical effects of genre and age restrictions since there are more movies in the data set with more variety of combinations for genre and age restrictions. Also, the opening week attractiveness, age decay, hour of showing effects, holidays and day of weeks will have better parameter estimations. The issues in here though is the time efficiency due to complicated likelihood functions in MCMC and the massive dataset.

This project can be further extended with another model: nested logit by hour. It may be interesting to compare the nested logit by hour model with the nested logit by movie model. Also, a Poisson model is be another candidate for predicting the counts of ticket sales. The derivation of posterior parameter distributions for the Poisson model is in Appendix B.

Appendix A

Posterior Distribution of σ_{jk}

A.1 σ_{jk} where genre and age restriction are used to construct a hierarchical layer

Prior:

$\sigma_g \sim \text{Beta}(\gamma_g, \delta_g)$

Hyper Prior:

$\gamma_g \sim \text{Log Normal}(\beta'_{\gamma_g} Z_g, 1)$
$\delta_g \sim \text{Log Normal}(\beta'_{\delta_g} Z_g, 1)$

Hyper Hyper Prior:

$\beta_{\gamma_g} \sim \text{MVN}(0, I)$
$\beta_{\delta_g} \sim \text{MVN}(0, I)$

Distributions:

$$
\begin{aligned}
[\sigma_g \,|\, \gamma_g, \delta_g] &= \frac{\Gamma(\gamma_g+\delta_g)}{\Gamma(\gamma_g)\cdot\Gamma(\delta_g)}\sigma_g^{(\gamma_g-1)}(1-\sigma_g)^{(\delta_g-1)} \\
\left[\gamma_g \,|\, \beta_{\gamma_g}, Z\right] &= \frac{1}{\gamma_g\sqrt{2\pi}}\exp(-\frac{1}{2}(\ln\gamma_g - \beta'_{\gamma_g}Z)^2) \\
\left[\delta_g \,|\, \beta_{\delta_g}, Z\right] &= \frac{1}{\delta_g\sqrt{2\pi}}\exp(-\frac{1}{2}(\ln\delta_g - \beta'_{\delta_g}Z)^2) \\
\left[\beta_{\gamma_g}\right] &\propto \exp(-\frac{1}{2}\beta'_{\gamma_g}I^{-1}\beta_{\gamma_g}) \\
\left[\beta_{\delta_g}\right] &\propto \exp(-\frac{1}{2}\beta'_{\delta_g}I^{-1}\beta_{\delta_g})
\end{aligned}
$$

A.2 Posterior for σ_g

$[\sigma_g \,|\, \cdot] \propto [S_g \,|\, \alpha_g, \omega, \sigma_g, \sigma] \cdot [\sigma_g \,|\, \gamma_g, \delta_g]$

$$
\begin{aligned}
[\sigma_g \,|\, \cdot] &\propto [S_g \,|\, \alpha_g, \omega, \sigma_g, \sigma] \cdot [\sigma_g \,|\, \gamma_g, \delta_g] \\
&= \prod_d \left[\frac{M!}{S_g!}\prod_{g\in C_d}\left[\frac{\exp(\sigma_g V_{g_{jkhd}})}{\exp(IV_{g_{jkd}})}\cdot\frac{\exp\left(\frac{\sigma}{\sigma_g}IV_{g_{jkd}}\right)}{\exp\left(IV_{g_d}^{mv}\right)}\cdot\frac{\exp\left(\frac{1}{\sigma}IV_{g_d}^{mv}\right)}{1+\exp\left(\frac{1}{\sigma}IV_{g_d}^{mv}\right)}\right]^{S_g}\right] \\
&\quad \cdot\frac{\Gamma(\gamma_g+\delta_g)}{\Gamma(\gamma_g)\cdot\Gamma(\delta_g)}\sigma_g^{(\gamma_g-1)}(1-\sigma_g)^{(\delta_g-1)} \\
&\propto \frac{\exp\left[\sum_d\sum_{g\in C_d}A\right]}{\prod_d\left[1+\exp\left(\frac{1}{\sigma}IV_{g_d}^{mv}\right)\right]^{\sum_{g\in C_d}S_g}}\cdot\frac{\Gamma(\gamma_g+\delta_g)}{\Gamma(\gamma_g)\cdot\Gamma(\delta_g)}\sigma_g^{(\gamma_g-1)}(1-\sigma_g)^{(\delta_g-1)}
\end{aligned}
$$

A.3 Posterior for γ_g

$$\begin{aligned}
[\gamma_g \,|\, \cdot] &\propto [\sigma_g \,|\, \gamma_g, \delta_g] \cdot [\gamma_g \,|\, \beta_{\gamma_g}, Z] \\
&= \frac{\Gamma(\gamma_g + \delta_g)}{\Gamma(\gamma_g) \cdot \Gamma(\delta_g)}\, \sigma_g^{(\gamma_g - 1)}\, (1 - \sigma_g)^{(\delta_g - 1)}\, \frac{1}{\gamma_g} \exp\left(-\frac{1}{2}[\ln(\gamma_g) - \beta_{\gamma_g} Z]^2\right) \\
&\propto \frac{\Gamma(\gamma_g + \delta_g)}{\Gamma(\gamma_g)}\, \sigma_g^{(\gamma_g - 1)}\, \frac{1}{\gamma_g} \exp\left(-\frac{1}{2}[\ln(\gamma_g) - \beta_{\gamma_g} Z]^2\right)
\end{aligned}$$

A.4 Posterior for δ_g

$$\begin{aligned}
[\delta_g \,|\, \cdot] &\propto [\sigma_g \,|\, \gamma_g, \delta_g] \cdot [\delta_g \,|\, \beta_{\delta_g}, Z] \\
&= \frac{\Gamma(\gamma_g + \delta_g)}{\Gamma(\gamma_g) \cdot \Gamma(\delta_g)}\, \sigma_g^{(\gamma_g - 1)}\, (1 - \sigma_g)^{(\delta_g - 1)}\, \frac{1}{\delta_g} \exp\left(-\frac{1}{2}[\ln(\delta_g) - \beta_{\delta_g} Z]^2\right) \\
&\propto \frac{\Gamma(\gamma_g + \delta_g)}{\Gamma(\delta_g)}\, (1 - \sigma_g)^{(\delta_g - 1)}\, \frac{1}{\delta_g} \exp\left(-\frac{1}{2}[\ln(\delta_g) - \beta_{\delta_g} Z]^2\right)
\end{aligned}$$

A.5 Posterior for β_{γ_g}

$$\begin{aligned}
[\beta_{\gamma_g} \,|\, \cdot] &\propto [\gamma_g \,|\, \beta_{\gamma_g}, Z] \cdot [\beta_{\gamma_g}] \\
&= \frac{1}{\gamma_g} \exp\left(-\frac{1}{2}[\ln(\gamma_g) - \beta_{\gamma_g} Z]^2\right) \cdot \exp\left(\beta'_{\gamma_g} I^{-1} \beta_{\gamma_g}\right) \\
&= \frac{1}{\gamma_g} \exp\left(-\frac{1}{2}[\ln(\gamma_g) - \beta_{\gamma_g} Z]^2 + \beta'_{\gamma_g} I^{-1} \beta_{\gamma_g}\right)
\end{aligned}$$

A.6 Posterior for β_{δ_g}

$$
\begin{aligned}
[\beta_{\delta_g} \mid \cdot] &\propto [\delta_g \mid \beta_{\delta_g}, Z] \cdot [\beta_{\delta_g}] \\
&= \frac{1}{\delta_g} \exp\left(-\frac{1}{2}[\ln(\delta_g) - \beta_{\delta_g} Z]^2\right) \cdot \exp\left(\beta'_{\delta_g} I^{-1} \beta_{\delta_g}\right) \\
&= \frac{1}{\delta_g} \exp\left(-\frac{1}{2}[\ln(\delta_g) - \beta_{\delta_g} Z]^2 + \beta'_{\delta_g} I^{-1} \beta_{\delta_g}\right)
\end{aligned}
$$

Appendix B

Poisson Model For Counts of Ticket Sales

Likelihood: $p(S \,|\, \mu) \sim \text{Poisson}(\mu)$
Prior: $p(\mu) \sim \text{Gamma}(a, b)$ where $\mu = \exp(X\beta)$

Note: $X\beta = \alpha_{jk} x_{jkhd} + \omega y$

B.1 Likelihood $p(S \,|\, \mu)$

$$\begin{aligned}
p(S \,|\, \mu) &= \prod_{i=1}^{n} \frac{\mu^{S_i} \exp(-\mu)}{S_i!} = \frac{\mu^{\sum_{i=1}^{n} S_i} \exp(-n\mu)}{\prod_{i=1}^{n} S_i!} \\
\text{Trasform } \mu &= \exp(X\beta) \\
p(S \,|\, \beta) &= p(S \,|\, \beta) \left\| \frac{d\mu}{d\beta} \right\| \\
&= \frac{\exp(X\beta)^{\sum_{i=1}^{n} S_i} \exp(-n \exp(X\beta))}{\prod_{i=1}^{n} S_i!} X \exp(X\beta) \\
&= \frac{X \exp\left(X\beta \sum_{i=1}^{n} S_i - n \exp(X\beta)\right)}{\prod_{i=1}^{n} S_i!}
\end{aligned}$$

B.2 Prior $p(\mu)$

$$\begin{aligned} p(\mu) &\sim \text{Gamma}(a,b) \\ &= \frac{b^a \mu^{a-1}\exp(-b\mu)}{\Gamma(a)} \end{aligned}$$

$$\text{Trasform } \mu = \exp(X\beta)$$

$$\begin{aligned} p(\beta) &= p(\mu)\left|\left|\frac{d\mu}{d\beta}\right|\right| \\ &= \frac{b^a \exp(X\beta(a-1))\exp(-b\exp(X\beta))}{\Gamma(a)} X\exp(X\beta) \\ &= \frac{Xb^a \exp(X\beta)\exp(-b\exp(X\beta))}{\Gamma(a)} \end{aligned}$$

B.3 Full Conditional Distribution

$$\begin{aligned} p(\beta|S) &\propto p(S|\beta)\times p(\beta) \\ &\propto \frac{X\exp\left(X\beta\sum_{i=1}^{n} S_i - n\exp(X\beta) + X\beta\right)}{\prod_{i=1}^{n} S_i!} \times \frac{Xb^a\exp(X\beta)\exp(-b\exp(X\beta))}{\Gamma(a)} \\ &\propto \exp\left(X\beta\sum_{i=1}^{n} S_i - n\exp(X\beta) + \beta + aX\beta - b\exp(X\beta)\right) \\ &= \exp\left(X\beta\sum_{i=1}^{n} S_i - n\exp(X\beta) + (a+1)X\beta - b\exp(X\beta)\right) \\ &= \exp\left(X\beta\left(\sum_{i=1}^{n} S_i + a + 1\right) - (n+b)\exp(X\beta)\right) \end{aligned}$$

Bibliography

[1] M. Ben-Akiva and Lerman S. R. *Discrete Choice Analysis.* The MIT Press, London, 1985.

[2] G. Casella and R. L. Berger. *Statistical Inference.* Duxbury, 2002.

[3] C. Eaton, D. Deroos, T. Deutsch, G. Lapis, and P. Zikopoulos. *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data.* McGraw Hill, 2012.

[4] J. Eliashberg, Q. Hegie, J. Ho, D. Huisman, S. J. Miller, S. Swami, C. B. Weinberg, and B. Wierenga. Demand-driven scheduling of movies in multiplex. *Intern. J. of Research in Marketing*, 26:75–88, 2009.

[5] A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis.* Chapman and Hall/CRC, Boca Raton, 2003.

[6] J. Ho. *Marketing Models of Entertainment Products.* PhD thesis, University of British Columbia, 2005.