University of California
Los Angeles

Multinomial Probit Model for Panel Data

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Statistics

by

Aparupa Das Gupta

2014

Abstract of the Thesis

Multinomial Probit Model for Panel Data

by

Aparupa Das Gupta

Master of Science in Statistics

University of California, Los Angeles, 2014

Professor Mark Stephen Handcock, Chair

In this thesis we applied the multinomial probit model to a panel dataset to study the brand preference for a consumer product that various households purchased over multiple purchase occasions by 100 households. There are four brands available for the consumer product. The dataset comprises information regarding the brand of item purchased by a household given their annual income, household size, and purchase quantity along with the price for each brand of the product. We analyzed the effect of the three individual specific covariates, namely, `annual income`, `household size`, and `quantity purchased`, and the choice specific covariate of `price` of each unit of the product for every brand. We used the **MNP** package in R by Imai and van Dyk for our analysis. The package does not have any function for model selection. Hence, we introduced a new approach to perform model selection for multinomial probit model by applying Kullback-Leibler divergence to evaluate the mean divergence of the average posterior predictive probabilities in the presence and absence of each of the three individual specific covariates. Finally, we obtain insights regarding how consumer preference across brands changes with the covariate values in terms of the posterior predictive probabilities of purchasing a brand.

The thesis of Aparupa Das Gupta is approved.

Yingnian Wu

Frederic R Paik Schoenberg

Mark Stephen Handcock, Committee Chair

University of California, Los Angeles

2014

To my family . . .

for their unconditional love and support

Table of Contents

List of Figures

List of Tables

Acknowledgments

This thesis would not have been possible without the guidance from my committee members. I am extremely grateful to my advisor, Professor Mark Handcock who has been an exceptional mentor and have facilitated my development as a scholar. I learnt from him how to formulate an idea into a model and how to present technical results effectively. I would also like to thank my thesis committee members, Professor Frederic R Paik Schoenberg and Professor Yingnian Wu for devoting their time to read my thesis and provide valuable feedback. I feel privileged to have them as my mentors. I would also like to thank Glenda Jones for her administrative support during the MS program.

Finally, I wish to acknowledge the unconditional love and support that I have received from my family throughout the years of my graduate studies.

CHAPTER 1

Introduction

For a long period of time, it was recognized that maximum likelihood analysis of nonlinear limited dependent variable (LDV) models with panel data is feasible only under restrictive assumptions [Butler and R.Moffitt, 1982]. Such models in general pose difficulty in the evaluation of a likelihood function containing multivariate integrals that are often analytically intractable. In the last few decades, advances in computation methodologies such as data augmentation [Tanner and Wong, 1987], has led to an increase of interest in Bayesian models that had previously been regarded as numerically infeasible. Under this framework, the latent variables within multivariate integrals are treated as model parameters and are sampled along with them. The Bayesian Gibbs sampling scheme is suitable for scenarios where the likelihood function comprises high-dimensional multivariate integrals which are factorized into sequences of low-dimensional conditional densities each of which is sampled individually. By embedding these low-dimensional subproblems within a Markov chain one can draw from the joint posterior distribution which are then used directly for inference.

Bayesian methods successfully compete against simulation-based frequentist techniques, such as Simulated Maximum Likelihood (SML) [Gourieroux and Monfort, 1996], due to their flexibility and conceptual simplicity. The advantages of Bayesian method is more pronounced when the dimension of the underlying problem is large. For instance, in our model setup SML approach would require a large set of latent variable draws in order to accurately approximate the integral likelihood function for each parameter value embedded within an optimization algorithm. In contrast, Gibbs sampling takes one latent variable draw for each parameter value until convergence which often leads to substantially faster parameter estimation than SML. Moreover, Bayesian hierarchical models can be readily extended to incorporate inference on latent classes of similar individuals Rossi et al. [2005]. Other advantages of Bayesian inference in latent variable models are discussed in Paap [2002].

In this thesis, we perform a Bayesian analysis of the multinomial probit model and apply it on scanner panel data of consumer purchase decision of yogurt from four different brands across multiple periods of time in response to product prices and other individual specific variables like household size, annual household income, and number of items that were purchased. The multinomial probit model is a generalization of the probit model where there are several possible categories that the dependent variable can fall into. The model is often used to describe how individuals choose among a number of different alternatives, for instance, how a voter chooses which candidate to vote for among four candidates running for a particular office. The multinomial probit model differs from the ordinal probit model in that the former does not assume any inherent ordering on the choices. Hence, although the individuals may have preferences among the available alternatives, these ordering are individual specific rather than being dependent on the nature of alternatives themselves.

This thesis is organized in the following way. In Chapter 2, we discuss the literature on probabilistic choice models and in particular the multinomial probit model. In Chapter 3, we discuss the model used in the **MNP** package by Imai and van Dyke. In this thesis we used the **MNP** package in R to analyze the data. The package does not provide any function to perform model selection. Hence, the primary contribution of this thesis is the introduction of a novel approach to perform model selection for multinomial probit models by applying Kullback-

Leibler divergence. This is further discussed in Chapter 4 along with complete data analysis for the yogurt panel dataset. In Chapter 5, we discuss the main insights obtained from the data using the multinomial probit model. Finally, we conclude with Chapter 6 by summarizing the model results, insights from the dataset, and the drawbacks of our analysis.

CHAPTER 2

Literature Review

For over a decade barcode readers at checkout have become ubiquitous in retailing. The opportunities offered by scanner data are particularly attractive in situations where customers typically purchase a number of items on each store visit and visit stores in a particular retail sector on a number of occasions over a period of time. The prime example is grocery retailing where the majority of households use one or more supermarkets frequently and regularly.

Two of the most widely used probabilistic choice models to estimate purchase probabilities from scanner data are the multinomial logit (MNL) and multinomial probit (MNP) models. These two models differ in the distribution of the error terms. MNL has errors which are independent and identically distributed according to the type-1 extreme value distribution, which is also sometimes called the log Weibull distribution (see Greene [2000], p.858 for a discussion on this distribution). MNP has errors which are not necessarily independent, and are distributed by a multivariate normal distribution (see Greene [2000], p.856).

The independent errors of MNL force an assumption called the independence of irrelevant alternatives (IIA) assumption. The IIA assumption requires that an individual's evaluation of an alternative relative to another alternative should not change if a third (irrelevant) alternative is added to or dropped from the analysis. So if a consumer is twice as likely to buy brand 1 over brand 2, she should remain twice as likely to do so even if a third brand becomes a viable option. When IIA is violated, MNL is an incorrectly specified model, and MNL coefficient estimates are biased and inconsistent. MNP does not assume IIA. In fact, an MNP model estimates the error correlations along with the coefficients. In the context of maximum likelihood estimation, a choice probability is a formula to predict the probability that an individual chooses a certain alternative and the likelihood function for such models is the product of the choice probabilities for each individual. Choice probabilities in an MNL model are relatively simple, and computers can maximize the resulting likelihood function almost instantaneously, even for a large number of choices. For MNP, choice probabilities involve multiple integrals: as many integrals as one fewer than the number of choices. Computers can typically maximize likelihood functions with double or triple integrals, and may take a while to do so. But when computers must deal with higher integrals, MNP will often fail to converge or provide useful estimation. MNL, therefore, is a much more stable model. But since MNP does not assume IIA it is often assumed to be more accurate than MNL.

Alvarez and Nagler [1998] strongly advocate the use of MNP as a less restrictive model. Chintagunta [1992] compared the multinomial logit model with a multinomial probit model, first by developing the probit model using the IRI saltine (Rome, Georgia) data and then applying it to Nielsen data on catsup from Springfield, Missouri. His argument is that the logit model's implicit IIA assumption is violated in these purchasing situations because of the similarity of directly competing brands leading to correlations in the data. Gonul and Srinivasan [1993] commented that the multinomial probit model is more difficult to compute and offer a variant of the multinomial logit model that is non-IIA at the aggregate but not at the household level.

In this thesis we adopted the multinomial probit model. There are two packages available in R that cover multinomial probit models, namely, `bayesm` by Rossi [2010] and **MNP** by Imai and van Dyk [2005]. We used the **MNP** package to estimate the purchase probabilities for four different brands of yogurt. Neither package provides a function for model selection. This thesis contributes to the literature of multinomial probit models by applying the Kullback-Leibler divergence measure to perform model selection.

CHAPTER 3

Methodology

We analyzed an unbalanced panel dataset that contains purchase history of N households over $T_i, i = 1, \ldots, N$ time periods. The dependent variable y_{it} takes the value { Brand 1, Brand 2, Brand 3, Brand 4} indicating the brand purchased by household i on purchase occasion t. The K-vector of control variables is denoted by $\underline{x}_{it}$ and the corresponding vector of parameters to be estimated by $\underline{\beta}$. We define matrix X_i corresponding to each household i, that has T_i rows and K columns. The independent variables refer to the household size, annual income, quantity of yogurt bought, and the price of yogurt for each brand.

3.1 The Pooled Panel Probit Model

The simplest probit estimator treats the entire sample as if it were a large cross-section. Specifically, it postulates the latent variable probit model specification

$$y^{\star}_{it} = \underline{\beta}_0^{'} \underline{x}_{it} + \epsilon_{it} \tag{3.1}$$

with the observation rule

$$y_{it} = \mathbf{1}(y^{\star}_{it} \geq 0), i : 1, \ldots, N; t = 1, \ldots, T_i \tag{3.2}$$

where $\mathbf{1}(.)$ denotes the indicator function. The error terms ϵ_{it} are normally distributed with zero mean and unit variance. This assumption rules out state persistence, or the presence of lagged dependent variables [Greene, 2004] in model (3.1).

3.2 Multinomial Probit Model

The multinomial probit model is a generalization of the probit model, and it is used when there are several possible categories that the dependent variable can fall into. Suppose that we have a series of observations $Y_i, i = 1, \ldots, n$ of the outcomes of multi-way choices from categorical distribution of size p (i.e., there are p possible choices). Each observation Y_i is associated with a set of K observed values $x_{i1}, \ldots, x_{iK}$. We introduce a set of latent variables $W_i = (W_{i1}, \ldots, W_{i,p-1})$, where

$$W_i = X_i\beta + e_i, e_i \sim N(0, \Sigma), i = 1, \ldots, n.$$

where X_i is a $(p-1) \times K$ matrix of covariates, β is $K \times 1$ vector of fixed coefficients, e_i is $(p-1) \times 1$ vector of disturbances, and Σ is a $(p-1) \times (p-1)$ positive definite matrix. We further use the restriction that the first diagonal element of Σ is constrained to be one, i.e., $\sigma_{11} = 1$. The response variable Y_i is the index of the choice of individual i among the alternatives in the choice set and is modeled as follows,

$$Y_i(W_i) = \begin{cases} 0 & \text{if } \max(W_i) < 0 \\ j & \text{if } \max(W_i) = W_{ij} > 0, \end{cases} \quad \text{for } i = 1, \ldots, n, \text{ and } j = 1, \ldots, p-1$$

where $Y_i = 0$ corresponds to a base category.

The matrix X_i can include both choice-specific and individual-specific variables. A choice-specific variable has a value for each of the p choices, and these p values may be different for each individual. Choice specific variables are recorded relative to the baseline choice and thus there are $p-1$ recorded values for each individual. In this way choice-specific variable is tabulated as a column in X_i. Individual specific variables, on the other hand, have a fixed value for each individual that is constant across choices, e.g., the age or gender of the individual. These variables are tabulated via their interaction with each of the choice indicator variables. Thus, an individual-specific variable corresponds to $p-1$ columns of X_i and $p-1$ components of β.

MNP implements the marginal data augmentation algorithms for posterior sampling in the multinomial probit model. The MCMC algorithm is described in Imai and Dyk [2005].

3.3 Prior Specification

The prior distribution for the multinomial probit model is $\beta \sim N(0, A^{-1})$ and $p(\Sigma) \propto |\Sigma|^{-(\nu+p)/2}[\text{trace}(S\Sigma^{-1})]^{-\nu(p-1)/2}$, where A is the prior precision matrix of β, ν is the prior degrees of freedom parameter for Σ, and the $(p-1)\times(p-1)$ positive definite matrix S is the prior scale for Σ. A scalar input can be used to set the scale matrix to a diagonal matrix with diagonal elements equal to the scalar input value. The default value is 1. We assume that the first diagonal element of S is one. The prior distribution on S is proper if $\nu \geq p-1$, the prior mean of Σ is approximately equal to S if $\nu > p-2$, and the prior variance of Σ increases as ν decreases as long as this variance exists. The model also allows for an improper prior on β, which is $p(\beta) \propto 1$. Other alternate prior specifications are available but this choice is preferred because it allows to directly specify the prior distribution on the identifiable model parameters, allows to specify an improper prior distribution on regression coefficient, and results in a Monte Carlo sampler that is relatively quick to converge.

3.4 Prediction under the Multinomial Probit Model

Predictions of individual preferences given particular values of the covariates can be useful in interpreting the fitted model. Consider a value of the $(p-1)\times k$ matrix of covariates, $X^{\star}$, that may or may not correspond to the values for one of the observed individuals. We are interested in the distribution of the preferences among the alternatives in the choice set given this value of the covariates. Let $Y^\star$ be the preferred choice among the available alternatives. As an example, one might be interested in $Pr(Y^\star = j|X^\star)$ for some j. By varying $X^\star$, one could explore how preferences are expected to change.

In Bayesian analysis, such predictive probabilities are computed via the posterior predictive distribution. This distribution conditions on the observed values of $Y = (Y_1, \ldots, Y_n)$, but averages over the uncertainty in the model parameters. For instance,

$$Pr(Y^\star = j|X^\star, Y) = \int Pr(Y^\star = j|X^\star, \beta, \Sigma, Y)p(\beta, \Sigma|Y)d(\beta, \Sigma).$$

Thus, the posterior predictive distribution accounts for both variability in the response variable given the model parameters and the uncertainty in the model parameters as quantified in the posterior distribution.

CHAPTER 4

Data Analysis

4.1 Data Description

In this section we analyze the yogurt data with R-package **MNP**, using Bayesian multinomial probit model by Markov Chain Monte Carlo (MCMC) techniques.

We look at customer purchase history of yogurt from retailer scanner panel data. The data set consists of 2430 purchases of four brands of yogurt by 100 households. The data comprises information regarding which brand of yogurt does a customer buy among four brands available at the retail store. The four brands were named as Brand 1, Brand 2, Brand 3, and Brand 4. At every purchase occasion, the `prices` of all the four brands, the `household size` and `income` of the customer, and `quantity` of yogurt purchased by her is generated by point-of-sales systems. We observed that `household size` varied from single person to five people. The `quantity` bought also varied from one to fourteen. Large size of `quantity` were purchased on fewer occasions. The household `income` was represented on a scale of one to fourteen, where one represented the lowest income level and fourteen represented the highest income level. There were 2430 purchase instances in the data set.

We took the logarithm of all the variables, both individual and choice specific, to normalize the data and ensure convergence of Gibbs sampling.

The description of variables is summarized in Tables 4.1 and 4.2.

Table 4.1: Description of Variables

Variables	Description
Choice	Brand purchased out of four brands
Price1	Log price of one unit of product of brand 1
Price2	Log price of one unit of product of brand 2
Price3	Log price of one unit of product of brand 3
Price4	Log price of one unit of product of brand 4
HHS	Log of total number of family members in a household
Income	Log of annual income of the household
Quantity	Log of number of units of the product purchased

Table 4.2: Summary of Variables

Choice	Statistic	Income	HHSize	Quantity	Price1	Price2	Price3	Price4
Brand1:831	Min	0.0000	0.0000	0.0000	-0.0120	0.00000	0.02500	0.00400
Brand2:975	1st Qu	0.7782	0.3010	0.0000	0.1030	0.08100	0.05000	0.07900
Brand3: 71	Median	0.9542	0.4771	0.3010	0.1080	0.08600	0.05400	0.07900
Brand4:553	Mean	0.8781	0.4027	0.3189	0.1062	0.08153	0.05362	0.07951
	3rd Qu	1.0792	0.6021	0.4771	0.1150	0.08600	0.06100	0.08600
	Max	1.1461	0.7782	1.1461	0.1930	0.11100	0.08600	0.10400

4.2 Fitting the Multinomial Probit Model using MNP

MNP returns the estimated probabilities for each brand in the model by MCMC. First, we ran pilot analysis with small iterations (10000) with 2000 burnin by calling "mnp" in MNP package. By calling the function

```
res=mnp(Choice   Income + HHSize + Quantity, choiceX=list(B1=Price1,
B2=Price2, B3=Price3, B4=Price4), cXnames=c("price"), data=yogurt,
n.draws=10000, burnin = 2000, thin = 3, verbose=TRUE)
```

we obtained the following result,

Call:

Coefficients:	mean	std.dev.	2.5%	97.5%
(Intercept):B2	0.26784	0.09514	0.08886	0.454
(Intercept):B3	-1.39954	0.24281	-1.86194	-0.932
(Intercept):B4	-1.27343	0.20894	-1.70977	-0.878
Income:B2	-0.47454	0.10708	-0.68816	-0.261
Income:B3	-0.87891	0.27848	-1.40878	-0.349
Income:B4	0.94758	0.19844	0.57243	1.355
HHSize:B2	-0.54429	0.13731	-0.82467	-0.287
HHSize:B3	-0.45546	0.31585	-1.06586	0.183
HHSize:B4	-1.78714	0.24706	-2.28164	-1.345
Quantity:B2	-0.02239	0.07912	-0.18237	0.135
Quantity:B3	0.53739	0.19145	0.16127	0.911
Quantity:B4	0.60966	0.13531	0.35734	0.889
price	-16.53935	1.68637	-20.21440	-13.524

```
Covariances:    mean    std.dev.    2.5%     97.5%
    B2:B2      0.44057   0.10647   0.28528   0.705
    B2:B3      0.23061   0.25879  -0.29389   0.619
    B2:B4      0.04471   0.18204  -0.36714   0.345
    B3:B3      1.38476   0.22072   1.01672   1.838
    B3:B4      0.99597   0.12700   0.71352   1.186
    B4:B4      1.11920   0.20544   0.74417   1.521

Base category:  B1
Number of alternatives:  4
Number of observations:  2430
Number of estimated parameters:  18
Number of stored MCMC draws:  2000
```

The dependent variable `choice` is defined in the model. We listed the choice-specific variable in `choiceX` and we also defined the name of these variables as "price" in `cXnames`. We named the data yogurt and MNP returned parameter estimates from 10000 replications with 2000 burnin. Here, we use non informative and improper prior, by default.

4.3 Running Convergence Diagnostic

We used `coda` package by Plummer et al. [2005] to perform convergence diagnostics. The `coda` package takes a matrix of posterior draws for relevant parameters to be saved as an mcmc object. We use the coda package to evaluate the Gelman Rubin convergence diagnostic statistic (Gelman and Rubin [1992]). This diagnostic is based on multiple independent Markov chains initiated at over-dispersed starting values. Here, we obtain these chains by independently running the `mnp()` command three times, specifying different starting values for each time. This can be accomplished by typing the following commands at the R prompt,

```
res1=mnp(Choice ~ Income+HHSize+Quantity, choiceX=list(B1=Price1,
B2=Price2, B3=Price3, B4=Price4), cXnames=c("price"), data=yogurt,
n.draws=120000, verbose=TRUE)
```

```
res2=mnp(Choice ~ Income+HHSize+Quantity, choiceX=list(B1=Price1,
B2=Price2, B3=Price3, B4=Price4), coef.start=c(1,-1,1,-1),
cov.start=matrix(0.5,ncol=3,nrow=3)+diag(0.5,3),cXnames=c("price"),
data=yogurt, n.draws=120000, verbose=TRUE)
```

```
res3=mnp(Choice ~ Income+HHSize+Quantity, choiceX=list(B1=Price1,
B2=Price2, B3=Price3, B4=Price4), coef.start=c(-1,1,-1,1),
cov.start=matrix(0.9,ncol=3,nrow=3) +diag(0.1,3),cXnames=c("price"),
data=yogurt, n.draws=120000, verbose=TRUE)
```

We ran three different chains to assess convergence by setting up the different initial values for each chain. For the first chain, the initial values for coefficients were zero for all coefficients and for covariance matrix initial values were set to be an identity matrix by default. The second chain is run starting from a vector of three 2's and three -2's and a matrix with all diagonal elements equal to 1 and all correlations equal to 0.5. Finally, the third chain is run starting from a vector of three 1's and three -1's, and a matrix with all diagonal elements equal to 1 and all correlations equal to 0.9. Each chain performed 120,000 draws and stored corresponding parameter estimates in `res1`, `res2`, and `res3`. The calculated estimates are summarized in Appendix A.

The R package `coda` gave the Gelman Rubin diagnostics. First, we combined the stored MCMC output into single list, `res.coda`. And then, the first element of diagonal in covariance matrix was eliminated since it is always equal to 1. Table 4.3 lists the value and a 97.5% upper limit of the Gelman-Rubin statistic for each parameter. The Gelman-Rubin statistics are all less than 1.1 for most cases,

Table 4.3: Potential scale reduction factors

	Point estimate	Upper C.I.		Point estimate	Upper C.I.
Intercept:B2	1.07	1.20	Quantity:B3	1.01	1.04
Intercept:B3	1.08	1.24	Quantity:B4	1.01	1.04
Intercept:B4	1.01	1.03	price	1.05	1.15
Income:B2	1.01	1.01	B2:B2	1.04	1.06
Income:B4	1.03	1.10	B2:B3	1.11	1.22
HHSize:B2	1.05	1.16	B2:B4	1.05	1.14
HHSize:B3	1.03	1.09	B3:B3	1.04	1.13
HHSize:B4	1.07	1.24	B3:B4	1.08	1.17
Quantity:B2	1.01	1.04	B4:B4	1.08	1.23
Multivariate perf	1.1				

suggesting satisfactory convergence has been achieved. We also plot the change in the value of the Gelman-Rubin statistic over the iterations in Figures 7.7-7.11 in Appendix B. The figure shows a cumulative evaluation of the Gelman-Rubin statistic over iterations for eighteen parameters.

We also use the coda package to produce univariate time-series plots of the three chains and univariate density estimate of the posterior distribution. Figures 7.1-7.6 in Appendix B present the resulting plots for selected parameters. The left panel shows the kernel-smoothed density estimate of the posterior distribution for each chain with a different color representing each chain. The right panel shows the density estimate plots for individual specific variables.

4.4 Model Comparison

In this section we evaluate the importance of each individual specific covariate for the predictive probabilities. In the complete model, we have three individual specific covariates, namely, `income`, `household size`, and `quantity`. In order to study the effect of each individual covariate we remove them one at a time or in pairs from the model and compare how much the posterior predictive probability diverges from the predictions obtained from the complete model on an average. We apply the Kullback-Leibler (KL) divergence criterion to measure the mean divergence.

The KL divergence is a measure of the difference between two probability distributions. For a discrete probability distribution **p** and **q**, the KL divergence of **q** from **p** is defined to be

$$D_{KL}(p||q) = \sum_i \ln\left(\frac{p(i)}{q(i)}\right) p(i). \tag{4.1}$$

In words, it is the expectation of the logarithmic difference between the probabilities **p** and **q**, where the expectation is taken using the probabilities **p**. The KL divergence is only defined if **p** and **q** both sum to 1 and if $\mathbf{q}(i) = 0$ implies $\mathbf{p}(i) = 0$ for all i (absolute continuity). The quantity $0 \ln 0$ in the formula is interpreted as zero because $\lim_{x\to 0} x\ln(x) = 0$.

Therefore, if **q** and **p** are the posterior predictive distribution from the full model and the model without an individual specific covariate respectively then the mean KL divergence is evaluated as

$$KL = \frac{1}{2430}\sum_{i=1}^{2430}\sum_{j=1}^{4} p_{ij}\log(\frac{p_{ij}}{q_{ij}}) \tag{4.2}$$

where $\mathbf{p_i}$ and $\mathbf{q_i}$ are the posterior predictive probabilities for the ith row of covariates from the yogurt data. We compared six models: model$_I$ that has only `income`, model$_H$ that has only `household size`, model$_Q$ that has only `quantity`, model$_{IH}$ has only `income` and `household size`, model$_{IQ}$ has only `income` and `quantity`, and model$_{HQ}$ has only `household size` and `quantity`. We compare each of these six models with the model with all covariates which we will call model$_{Full}$ henceforth in the paper.

We fit the model$_{Full}$ and use the function `predict()` as follows

```
Q=predict(res1,newdraw=rbind(res1$param[8909:10909,],
```

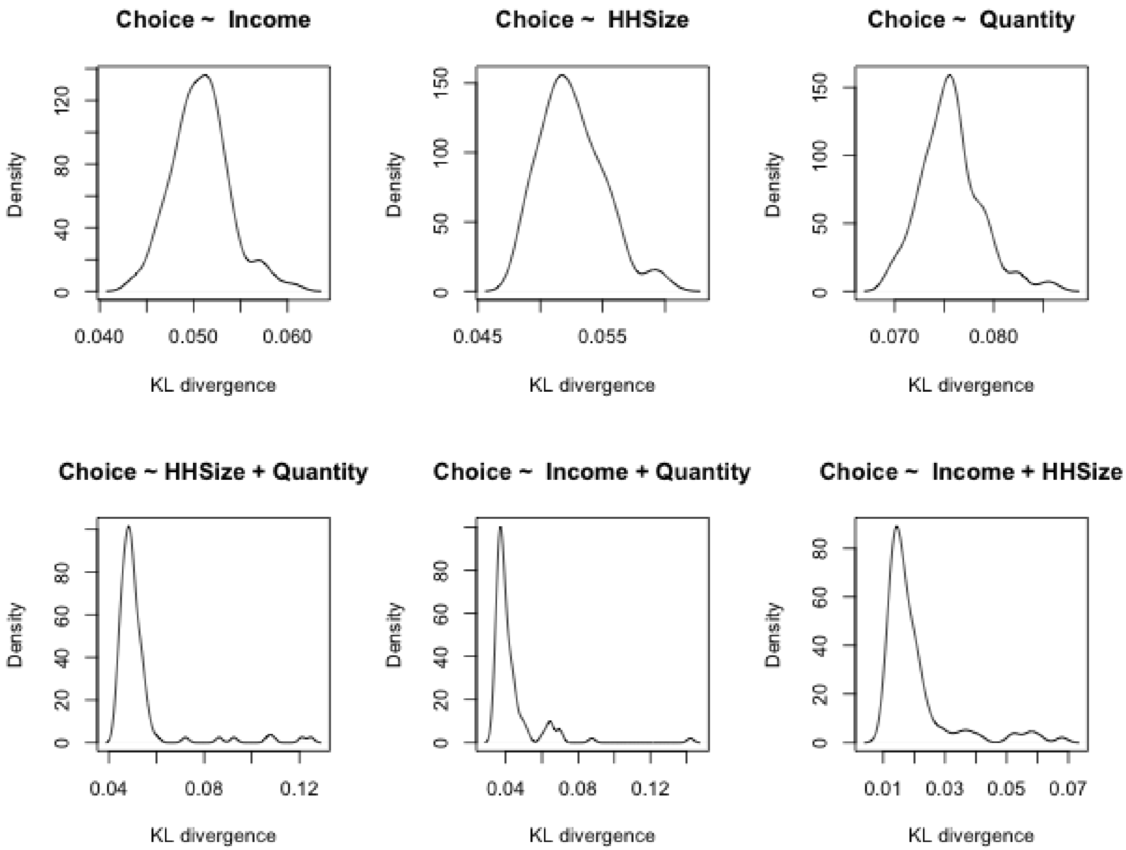

Figure 4.1: The mean Kullback-Leibler divergence from the full model.

`res2$param[8909:10909,], res3$param[8909:10909,]), type="prob")`
to obtain the posterior predictive probabilities for model$_{Full}$ using the last 20000 draws from the three chains. We further obtain the Monte Carlo sample of the preferred choice for each row of covariates from yogurt data by using the function
`Y=predict(res1,newdraw=rbind(res1$param[8909:10909,],`
`res2$param[8909:10909,], res3$param[8909:10909,]), type="choice")`.

We further use this Monte Carlo sample Y to fit all the six models and obtain the corresponding posterior predictive probabilities P_I, P_H, P_Q, P_{IH}, P_{IQ}, and P_{HQ}. These posterior predictive probabilities are then used to measure the mean divergence from Q using equation 4.2.

Table 4.4: Summary of Mean KL Divergence for Each Model

Statistic	model_{HQ}	model_{IQ}	model_{IH}	model_{I}	model_{H}	model_{Q}
Min	0.04314	0.03397	0.009359	0.04351	0.04827	0.06950
1st Qu	0.04667	0.03674	0.013720	0.04874	0.05082	0.07363
Median	0.04907	0.03856	0.016450	0.05077	0.05226	0.07547
Mean	0.05228	0.04296	0.020040	0.05076	0.05255	0.07565
3rd Qu	0.05222	0.04362	0.020710	0.05233	0.05419	0.07692
Max	0.12480	0.14210	0.067930	0.06074	0.06011	0.08604

The corresponding mean divergence for each sample of Y was plotted in Figure 4.1. Table 4.4 shows the descriptive statistic related to the mean divergence for each model. We see that average mean divergence is smallest for model_{IH} and largest for model_{Q}. Clearly, this indicates that income and household size are the most important covariates in the model and quantity is the least important covariate. This is so since the model with more covariates will have more accurate predictive probabilities, so we can use model_{Full} as a benchmark. Consequently, a model that diverges less from model_{Full} has more accurate predictive probabilities than a model that diverges greatly from model_{Full}.

Therefore, from Table 4.4 we conclude that model_{IH} is the preferred model among all other models since it has the smallest divergence from the full model and it requires fewer parameters for estimation.

CHAPTER 5

Predictions

The **MNP** model can be used to predict the likely outcome of an unobserved multi-way trial given the associated explanatory variables. In this thesis we apply the model to predict which brand of yogurt will be purchased by a consumer with given values for `income`, `household size`, and `quantity`. To analyze the data using the MNP model we combine the second half of each of the three chains. We use the command

```
res.coda=mcmc.list(chain1=mcmc(res1$param[5455:10909, -5],
start=5455), chain2=mcmc(res2$param[5455:10909, -5], start
=5455), chain3=mcmc(res3$param[5455:10909, -5], start=5455))
```

The summary of the output (see Appendix A) shows the mean, standard deviation, and various percentiles of the posterior distributions of the coefficients and the elements of the variance-covariance matrix. The base category is B1. Separate intercepts are estimated for each brand. The price coefficient is negative which is consistent with the fact that consumers are less likely to buy more expensive products.

We used the `predict()` function of **MNP** to calculate the posterior predictive probabilities of each alternative being the most preferred given a particular value of the covariates. We set the data used for prediction(newdata) to be the first 10 rows of the original yogurt dataset. To predict the probability for each brand we specified to use additional matrix of MCMC draws (newdraw) that are combined estimates from three chains. Option type = "prob" induces the function

Table 5.1: Posterior Predictive Probabilities for the Brands

	B1	B2	B3	B4
[1,]	0.3000000	0.3666667	0.00000000	0.3333333
[2,]	0.2666667	0.2666667	0.00000000	0.4666667
[3,]	0.4000000	0.3333333	0.03333333	0.2333333
[4,]	0.3333333	0.3000000	0.00000000	0.3666667
[5,]	0.3000000	0.3333333	0.03333333	0.3333333
[6,]	0.2666667	0.3666667	0.03333333	0.3333333
[7,]	0.3000000	0.4000000	0.00000000	0.3000000
[8,]	0.5000000	0.2333333	0.00000000	0.2666667
[9,]	0.5333333	0.2333333	0.06666667	0.1666667
[10,]	0.3666667	0.5333333	0.00000000	0.1000000

`predict()` to return the posterior predictive probabilities. The function called to obtain the predictive probabilities is as follows:

```
predict(res1, newdata=yogurt[1:10,] , newdraw = rbind(res1$
param[10900:10909,], res2$param[10900:10909,], res3$param[10900:10909,]),
type = "prob")
```

Table 5.1 summarizes the posterior predictive probabilities obtained for covariates corresponding to the first ten data points. These correspond to high income households with two family members. We observe that B3 is the cheapest brand during this period and these households have lowest probability of buying this brand. In a similar way, by changing the covariate values we can obtain the posterior predictive probabilities for choosing a brand.

Further analysis shows that the cheapest brand is not always most preferred by consumers. In fact if the household falls in the high income class and they have low household size then they are more likely to prefer the brand with higher price. If the household has lower income or higher household size then they are more sensitive to price and will prefer the cheaper brands. This indicates sensitivity towards price increases with decrease in income or increase in household size.

Whereas, when both income and household size are fixed such that income is low or household size is high then quantity purchased is inversely related to the sensitivity to price and consumer prefers cheaper brands.

CHAPTER 6

Conclusion

In this thesis we analyze a panel dataset of consumer purchase history for four different brands of yogurt. We apply the multinomial probit model to estimate the posterior predictive probability of any consumer purchasing a specific brand of yogurt given the price of all the brands and the individual specific information regarding the consumer's annual income, household size, and quantity purchased.

We perform model selection by applying Kullback-Leibler divergence and find that `income` and `household size` together fit the set of observations well and we can drop `quantity` from the model. This allows us to simplify the model and have fewer parameters to estimate.

The posterior predictive probabilities indicate that consumers with high income and low household size prefer more expensive brands whereas consumers with low income or high household size are more sensitive to price and have higher probability of purchasing cheaper brands.

The drawbacks of our analysis is that the estimated models may be misspecified to varying degrees and the insights may not extend to other datasets or for the same brands in future. In this thesis we only consider `price` as the choice specific variable. But we know that products can be differentiated by many attributes, like nutrition content. Therefore, inclusion of other choice specific variables that can be attributed to the products can provide more insights towards understanding the purchase preference of consumers.

CHAPTER 7

Appendices

7.1 Appendix A

This Appendix contains the summary of the results obtained by calling the mnp() function for model fitting. The summary of result from

```
res1=mnp(formula = Choice ~ Income + HHSize + Quantity, data = yogurt,
choiceX = list(B1 = Price1, B2 = Price2, B3 = Price3, B4 = Price4),
cXnames = c("price"), n.draws = 120000, thin = 10, verbose = TRUE)
```

Coefficients:	mean	std.dev.	2.5%	97.5%
(Intercept):B2	0.20814	0.13321	-0.05395	0.465
(Intercept):B3	-1.58266	0.37478	-2.29366	-0.918
(Intercept):B4	-1.49610	0.37693	-2.22416	-0.652
Income:B2	-0.48198	0.14856	-0.76452	-0.154
Income:B3	-1.12089	0.38779	-1.97337	-0.463
Income:B4	1.09943	0.31362	0.52146	1.769
HHSize:B2	-0.43763	0.16983	-0.77443	-0.106
HHSize:B3	-0.11401	0.42795	-0.90406	0.769
HHSize:B4	-1.75575	0.37565	-2.46026	-1.029
Quantity:B2	-0.06389	0.09247	-0.24971	0.104
Quantity:B3	0.51638	0.24652	0.05006	1.029
Quantity:B4	0.61064	0.15583	0.30892	0.922
price	-15.38517	2.97751	-20.19082	-7.710

Covariances:	mean	std.dev.	2.5%	97.5%
B2:B2	0.44997	0.16161	0.11855	0.775
B2:B3	-0.03556	0.51609	-0.91965	0.663
B2:B4	-0.12776	0.24691	-0.58258	0.280
B3:B3	1.42335	0.41348	0.68094	2.327
B3:B4	0.61948	0.57217	-1.12683	1.136
B4:B4	1.12668	0.35783	0.48482	1.877

The summary of result from

```
res2=mnp(formula = Choice   Income + HHSize + Quantity, data = yogurt,
choiceX = list(B1 = Price1, B2 = Price2, B3 = Price3, B4 = Price4),
cXnames = c("price"), n.draws = 120000, coef.start = c(1, -1, 1, -1,
1, -1, 1, -1, 1, -1, 1, -1, 1) * 2, cov.start = matrix(0.5, ncol = 3,
nrow = 3) + diag(0.5, 3), thin = 10, verbose = TRUE)
```

Coefficients:	mean	std.dev.	2.5%	97.5%
(Intercept):B2	0.261993	0.099336	0.073356	0.464
(Intercept):B3	-1.405677	0.286741	-2.040756	-0.903
(Intercept):B4	-1.521751	0.334661	-2.185126	-0.863
Income:B2	-0.517754	0.132125	-0.790069	-0.270
Income:B3	-1.061567	0.370136	-1.804490	-0.391
Income:B4	1.141650	0.290528	0.608201	1.740
HHSize:B2	-0.506294	0.145843	-0.796158	-0.219
HHSize:B3	-0.243944	0.394872	-0.945070	0.595
HHSize:B4	-1.898736	0.288693	-2.465401	-1.321
Quantity:B2	-0.046985	0.085327	-0.218483	0.114
Quantity:B3	0.455880	0.222366	0.008275	0.893
Quantity:B4	0.638921	0.143558	0.373204	0.934
price	-16.679083	1.966512	-20.404778	-12.685

Covariances:	mean	std.dev.	2.5%	97.5%
B2:B2	0.47678	0.13330	0.25232	0.770
B2:B3	0.22391	0.30783	-0.55474	0.667
B2:B4	-0.04936	0.17922	-0.41078	0.268
B3:B3	1.29550	0.29358	0.75235	1.914
B3:B4	0.65904	0.53074	-0.98818	1.190
B4:B4	1.22772	0.27474	0.70211	1.791

The summary of result from

```
res3=mnp(formula = Choice   Income + HHSize + Quantity, data = yogurt,
choiceX = list(B1 = Price1, B2 = Price2, B3 = Price3, B4 = Price4),
cXnames = c("price"), n.draws = 120000, coef.start = c(-1, 1, -1, 1,
-1, 1, -1, 1, -1, 1, -1, 1, -1) * 1, cov.start = matrix(0.9, ncol = 3,
nrow = 3) + diag(0.1, 3), thin = 10, verbose = TRUE)
```

Coefficients:	mean	std.dev.	2.5%	97.5%
(Intercept):B2	0.172041	0.199223	-0.458240	0.436
(Intercept):B3	-1.601575	0.363652	-2.350592	-0.972
(Intercept):B4	-1.549517	0.394158	-2.289440	-0.766
Income:B2	-0.505608	0.139611	-0.789343	-0.235
Income:B3	-1.173222	0.437004	-2.039085	-0.408
Income:B4	1.111678	0.366705	0.377940	1.809
HHSize:B2	-0.463692	0.183680	-0.847188	-0.116
HHSize:B3	-0.035211	0.455798	-0.848106	0.912
HHSize:B4	-1.751306	0.416066	-2.462519	-0.843
Quantity:B2	-0.060092	0.094495	-0.249844	0.124
Quantity:B3	0.488764	0.250608	-0.004726	0.984
Quantity:B4	0.619720	0.158270	0.305695	0.938
price	-16.053037	2.639978	-20.642921	-9.774

Covariances:	mean	std.dev.	2.5%	97.5%
B2:B2	0.49292	0.19210	0.16074	0.914
B2:B3	-0.05841	0.52645	-1.08738	0.655
B2:B4	-0.16768	0.25083	-0.72075	0.234
B3:B3	1.38573	0.39277	0.74984	2.335
B3:B4	0.47221	0.70381	-1.14868	1.149
B4:B4	1.12135	0.35591	0.36884	1.789

```
Base category:  B1
Number of alternatives:  4
Number of observations:  2430
Number of estimated parameters:  18
Number of stored MCMC draws:  10909
```

7.2 Appendix B

This Appendix contains all the figures related to the analysis of the estimates of the multinomial probit model.

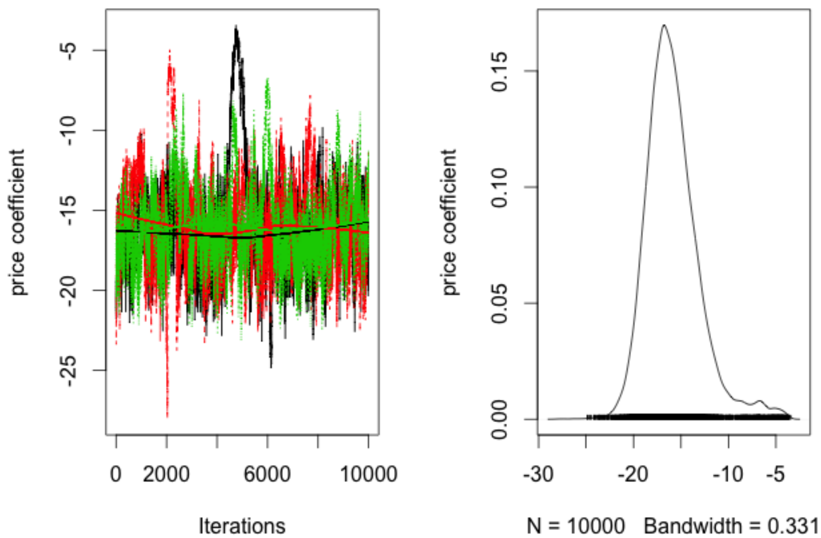

Figure 7.1: Time-series plot of three independent markov chains (Left Panel) and a density estimate of the posterior distribution of the price coefficient (Right Panel)

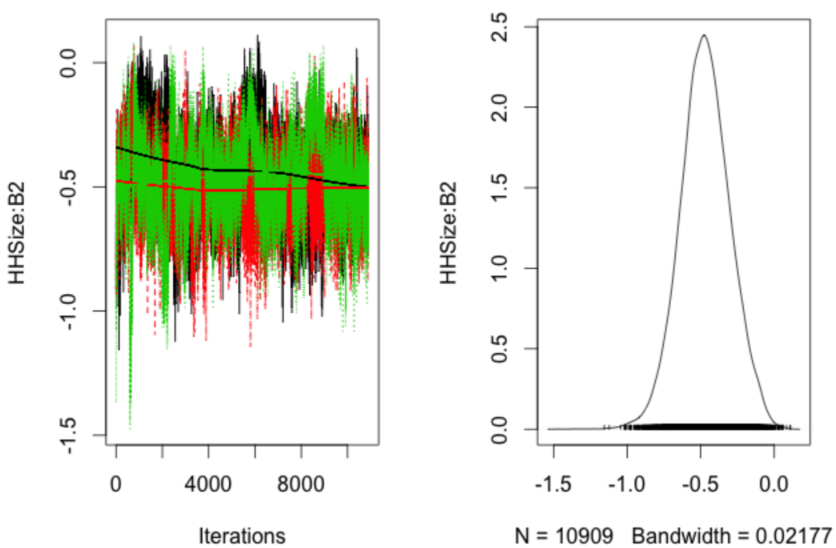

Figure 7.2: Time-series plot of three independent markov chains (Left Panel) and a density estimate of the posterior distribution of selected household size coefficient (Right Panel) for Brand 2.

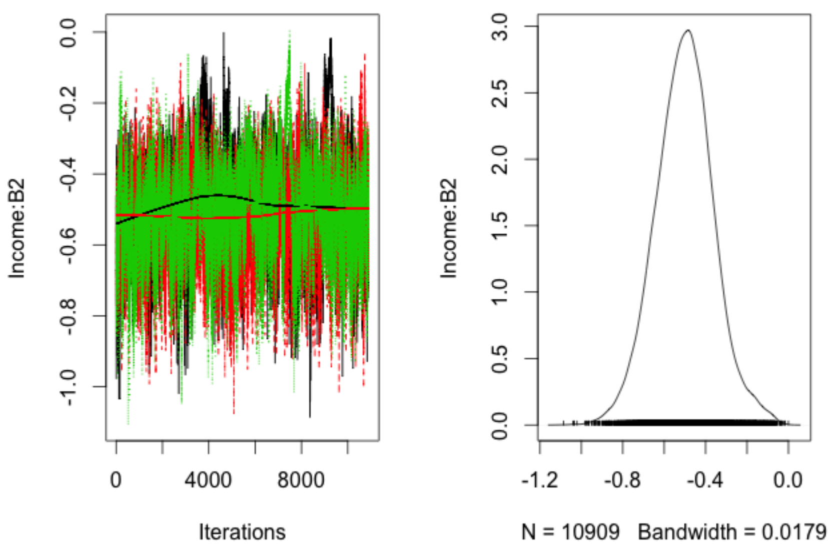

Figure 7.3: Time-series plot of three independent markov chains (Left Panel) and a density estimate of the posterior distribution of selected income coefficient (Right Panel) for Brand 2.

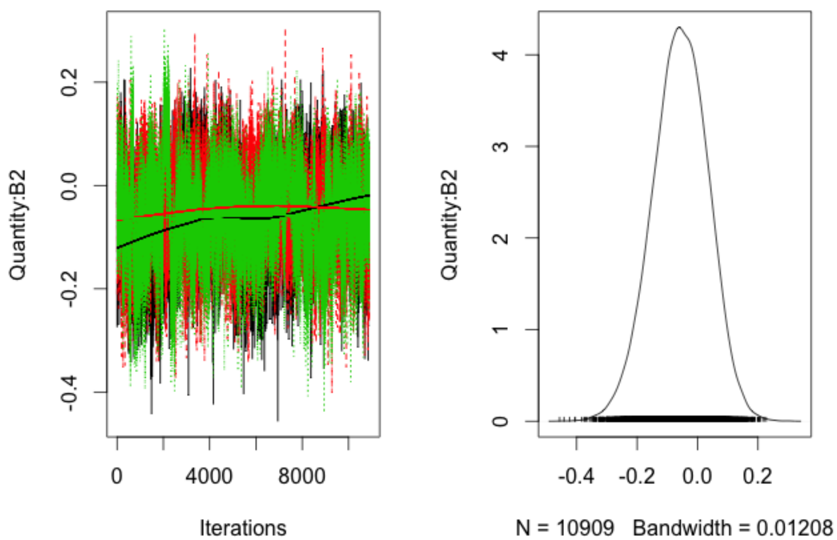

Figure 7.4: Time-series plot of three independent markov chains (Left Panel) and a density estimate of the posterior distribution of selected quantity coefficient (Right Panel) for Brand 2.

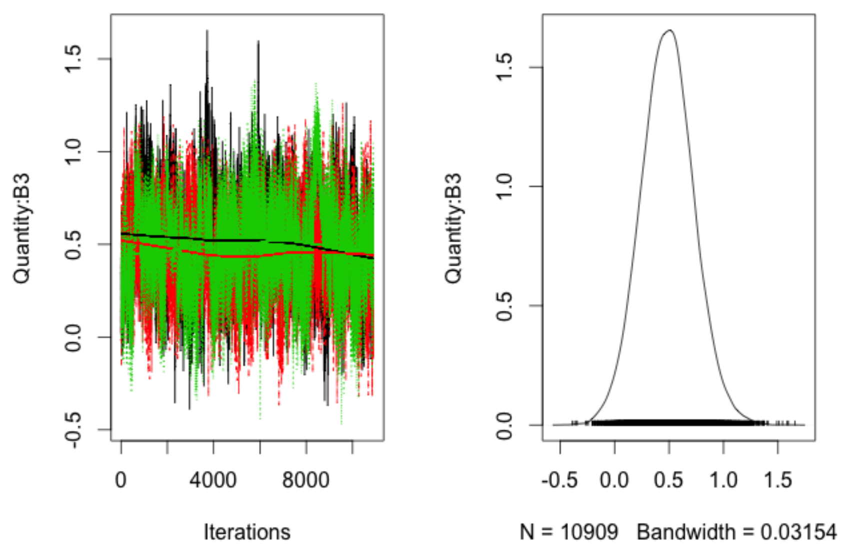

Figure 7.5: Time-series plot of three independent markov chains (Left Panel) and a density estimate of the posterior distribution of selected quantity coefficient (Right Panel) for Brand 3.

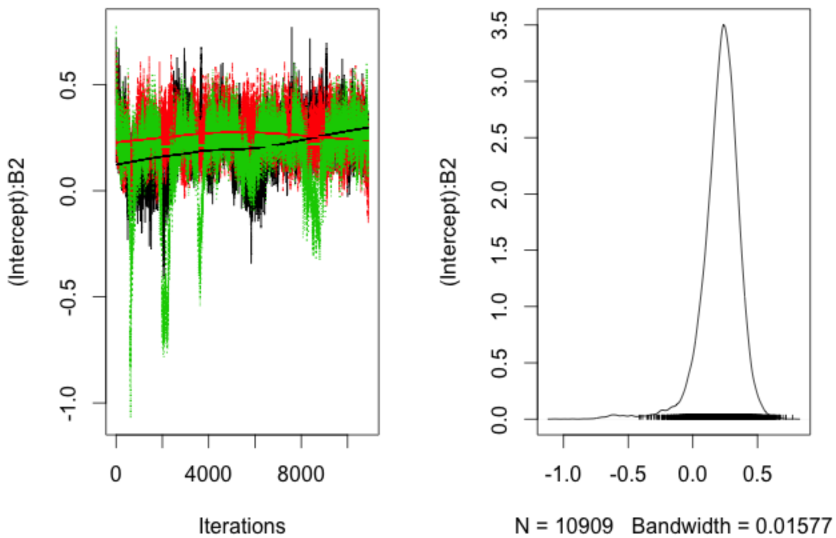

Figure 7.6: Time-series plot of three independent markov chains (Left Panel) and a density estimate of the posterior distribution of selected intercept (Right Panel) for Brand 2.

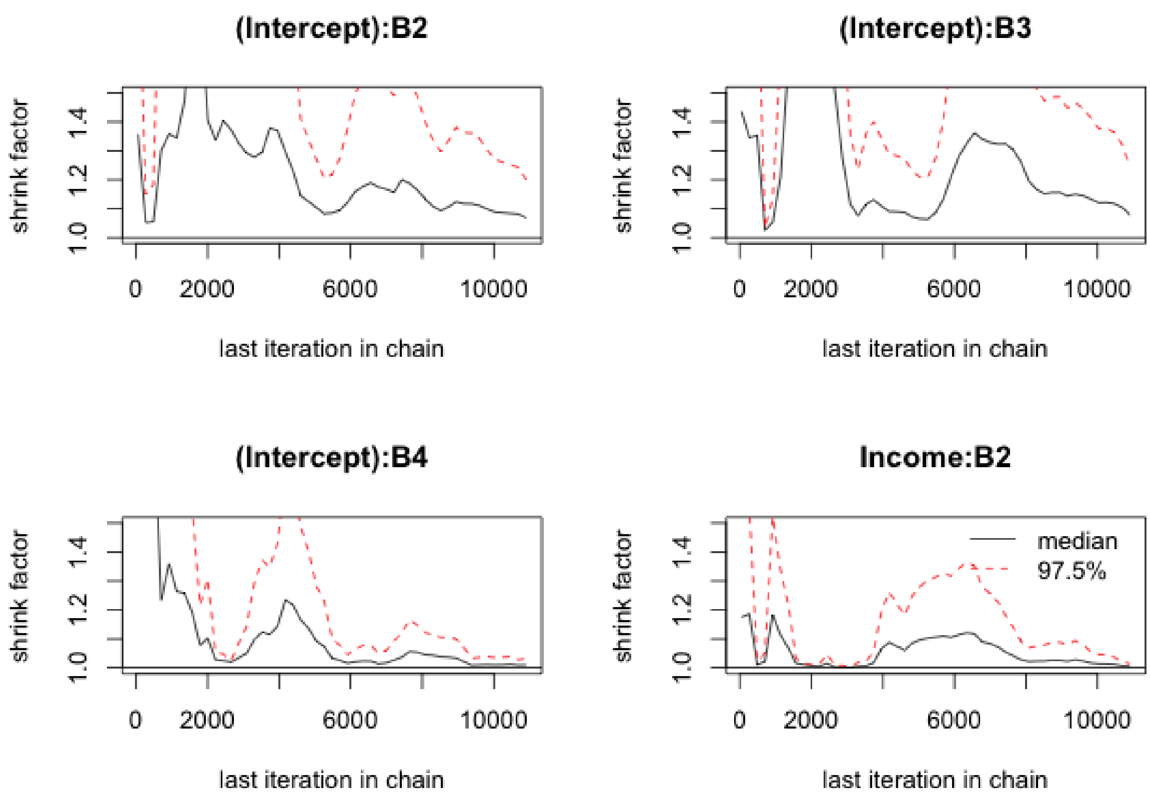

Figure 7.7: The Gelman-Rubin statistic computed with three independent Markov chains for selected parameters in the yogurt data.

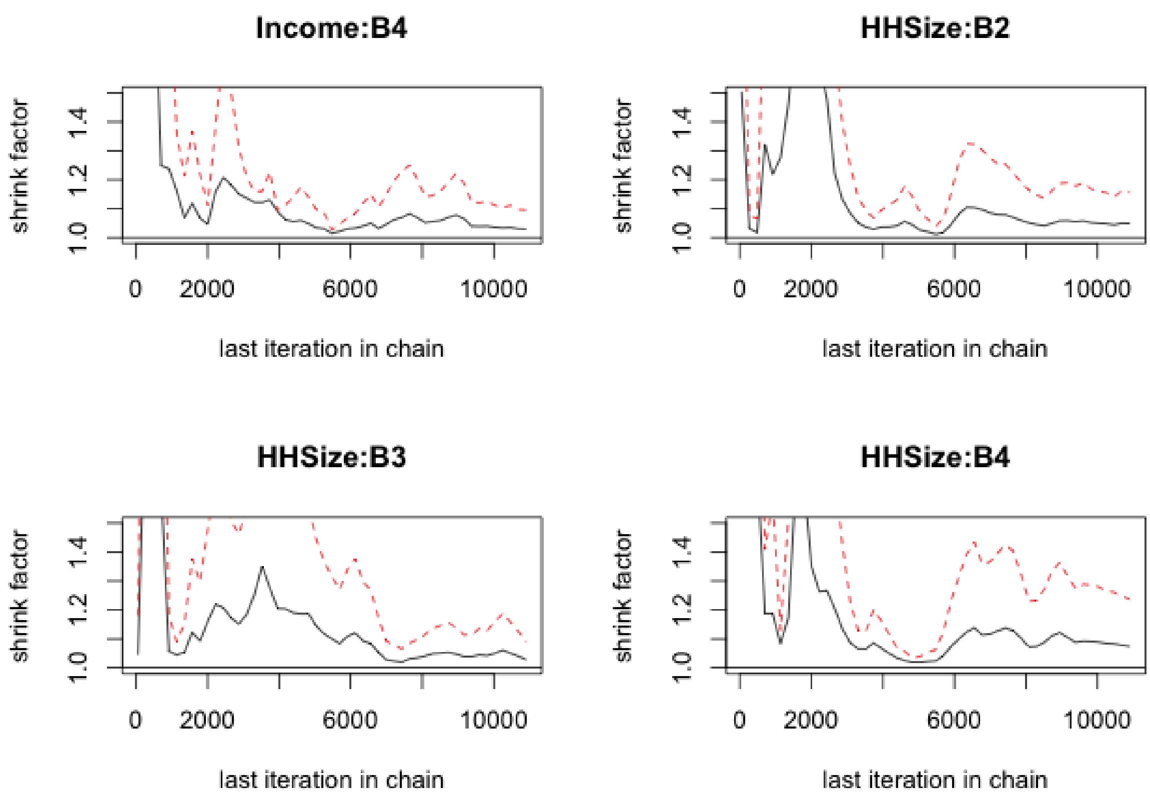

Figure 7.8: The Gelman-Rubin statistic computed with three independent Markov chains for selected parameters in the yogurt data.

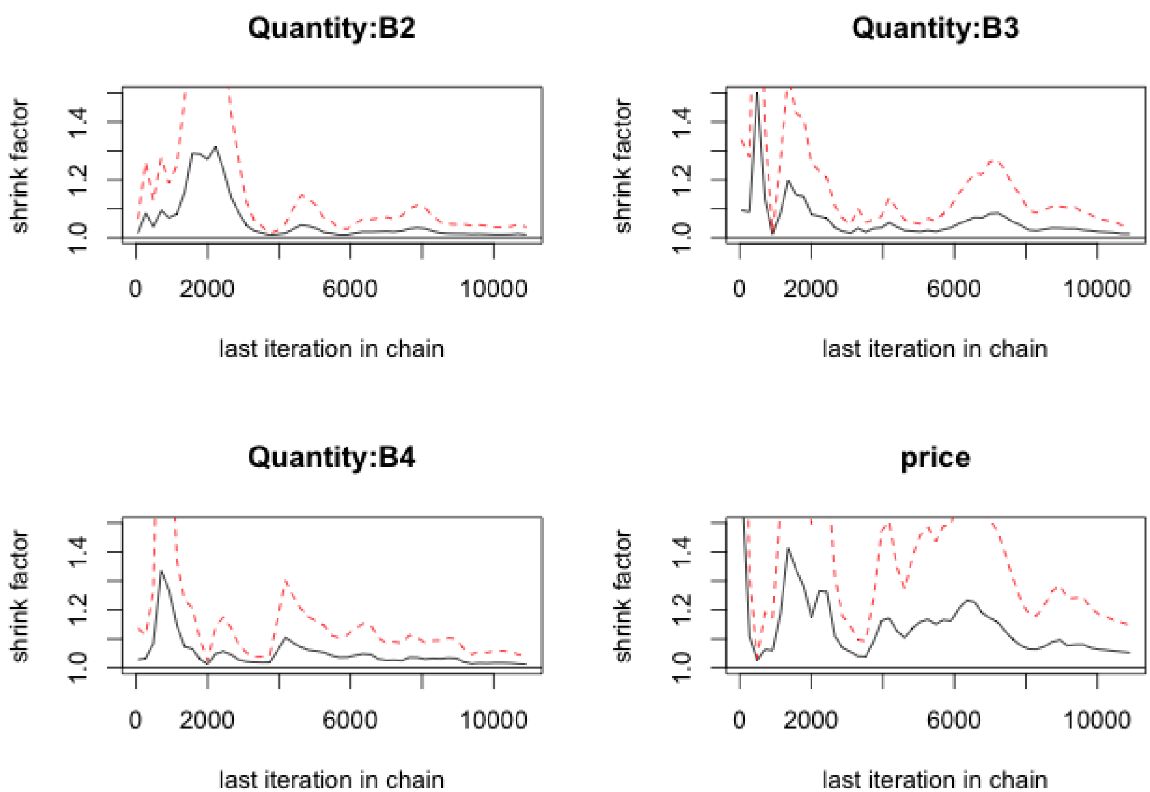

Figure 7.9: The Gelman-Rubin statistic computed with three independent Markov chains for selected parameters in the yogurt data.

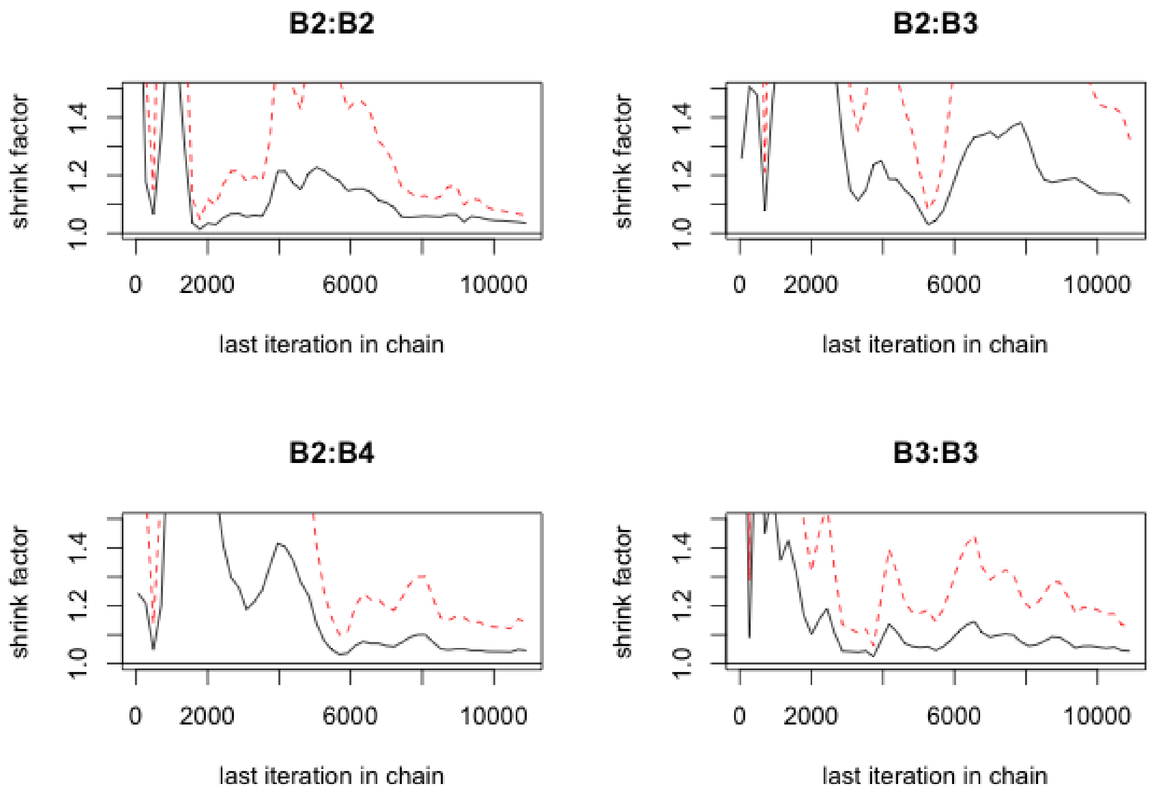

Figure 7.10: The Gelman-Rubin statistic computed with three independent Markov chains for selected parameters in the yogurt data.

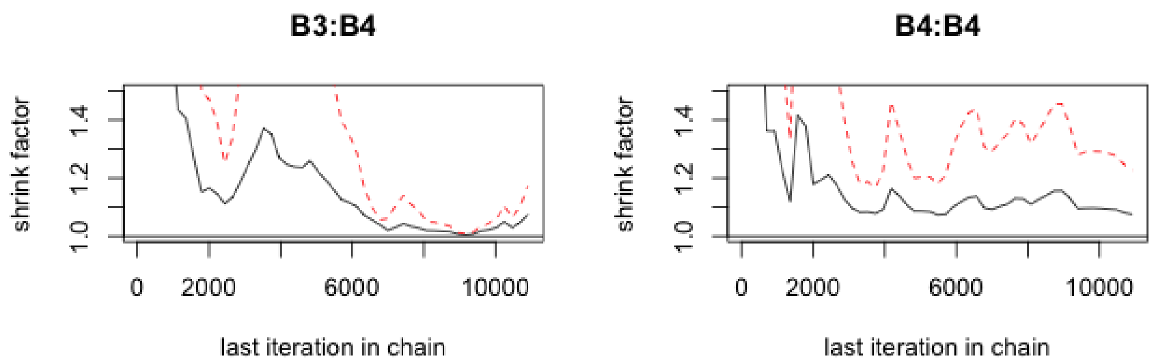

Figure 7.11: The Gelman-Rubin statistic computed with three independent Markov chains for selected parameters in the yogurt data.

Bibliography

R.M. Alvarez and J. Nagler. When politics and models collide: Estimating models of multiparty elections. *American Journal of Political Science*, 42:55–96, 1998.

J.S. Butler and R.Moffitt. A computationally efficient quadrature procedure for the one-factor multinomial probit model. *Econometrica*, 50(3):761–764, 1982.

P. Chintagunta. Estimating a multinomial probit model of brand choice using the method of simulated moments. *Marketing Science*, 11(4):386–407, 1992.

A. Gelman and D.B. Rubin. Inference from iterative simulations using multiple sequences (with discussion). *Statistical Science*, 7(4):457–472, 1992.

F. Gonul and K. Srinivasan. Modeling multiple sources of heterogeneity in multinomial logit models: methodological and managerial issues. *Marketing Science*, 12:213–229, 1993.

C. Gourieroux and A. Monfort. *Simulation-based econometric methods.* Oxford University Press. 1996.

W. Greene. Convenient estimators for the panel probit model: Further results. *Empirical Economics*, 29(1):21–47, 2004.

W.H. Greene. *Econometric Analysis.* Fourth edition ed. Upper Saddle River, NJ: Prentice Hall. 2000.

K. Imai and D.A. Dyk. A bayesian analysis of the multinomial probit model using marginal data augmentation. *Journal of Econometrics*, 124(2):311–334, 2005.

K. Imai and D.A. van Dyk. Mnp: R package for fitting the multinomial probit model. *Journal of Statistical Software*, 14:1–32, 2005.

R. Paap. What are the advantages of mcmc based inference in latent variable models? *Statistica Neerlandica*, 56(1):2–22, 2002.

M. Plummer, N. Best, K. Cowles, and K. vINES. *coda: Output analysis and diagnostics for MCMC.* R package version 0.9-2. 2005.

P. Rossi. bayesm: Bayesian inference for marketing/micro-econometrics. R package version 2.2-3. URL http://cran.r-project.org/package=bayesm. 2010.

P.E. Rossi, G.M. Allenby, and R. McCulloch. *Bayesian Statistics and Marketing.* Wiley series in Probability and Statistics. 2005.

M.A. Tanner and W.H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–550, 1987.