KU LEUVEN

FACULTY OF SCIENCE

# Comparing different logit models to describe heterogeneous choice behaviour

**Cheng CHENG**

Supervisor: Prof. M. Vandebroek

Mentor: Deniz Akinc

Thesis presented in fulfillment of the requirements for the degree of Master of Science in Statistics

Academic year 2013-2014

# Preface

In marketing, it is becoming more and more important for companies to understand customers' heterogeneous choice behaviour during their decision making process.

This thesis presents six different logit models including the multinomial logit (MNL), the mixed logit (MIXL), the latent class (LC), the scale-Heterogeneity logit (S-MNL), the generalized multinomial logit (G-MNL) and the mixed-mixed multinomial logit (MM-MNL) and compares their modelling abilities. Two estimation methods, the maximum simulated likelihood (MSL) and the hierarchical Bayesian estimation (HB), are described and compared for their performance.

The first objective of this thesis is to replicate the estimation results of two datasets (PIZZA A and PIZZA B) from Keane and Wasi (2013) and then to compare model estimations of both MSL and HB procedures. Finally, two simulations are developed to compare estimation and prediction accuracy of different models and different estimation procedures.

The thesis would not have been finished without the support of many people. First I would like to thank my thesis promoter Prof. M. Vandebroek for giving me constant support during the whole project. I would also like to thank thesis advisor D. Akinc for her inspiration with insightful suggestions. Second I also want to thank Prof. W. Greene for his thoughtful suggestions and comments.

Finally I want to give particular thanks to my parents and all the friends for their support during my lifelong journey.

# Summary

The aim of marketing is to understand heterogeneous behaviour of consumers. Being able to discover the consumers' diversity is crucial for new product development and product differentiation. In this thesis, six different models including the multinomial logit (MNL), the mixed logit (MIXL), the latent class (LC), the scale-heterogeneity logit (S-MNL), the generalized multinomial logit (G-MNL) and the mixed-mixed multinomial logit (MM-MNL) are presented and compared for their modelling abilities. Two estimation methods, the maximum simulated likelihood (MSL) and the hierarchical Bayesian estimation (HB), are described and compared for their performance.

The thesis starts with an introduction of the basic concepts, such as discrete choice model, the features of six different logit models, the estimation process of both classical and Bayesian procedures. After giving an overview of the theoretical background, the next step is to replicate the six models' results of two experiments from a previous study of Keane and Wasi (2013), and to compare two estimation approaches using the software Nlogit 5 and the *bayesm* package in R. Afterwards, two simulations are developed to assess the estimation and prediction accuracy of different models and different estimation procedures.

By applying the MSL procedure, this analysis starts from two PIZZA delivery experiments and finds that all model selection criteria are in favour of the MIXL model. However, Keane and Wasi (2013) found that MIXL was never preferred compared with G-MNL and MM-MNL when capturing the complex structure of heterogeneity in data.

The estimates from simpler models, such as MNL, MIXL, LC and S-MNL are very close to the reported results from the paper of Keane and Wasi (2013). In terms of the relatively complex models, such as G-MNL and MM-MNL, some discrepancies in estimates are discovered. One possible reason is because G-MNL and MM-MNL are complex models, so that the estimation is very sensitive to starting values and requires educated guesses. Due to the technical issue of setting different starting values in Nlogit 5, the model might fail to converge or only report the local maximum.

Concluded from simulation studies, both estimation procedures, MSL and HB, provide the asymptotically equivalent information for the parameter estimates. The computation of MSL is very heavy and takes longer time to estimate the same model than using HB procedure.

# Table of Contents

# List of Figures



# List of Tables

## List of Symbols

| | |
|---|---|
| $U_{jsn}$ | Utility function for decision maker $n$ at choice occasion $s$ choose alternative choice $j$ |
| $V_{jsn}$ | Representative utility for decision maker $n$ at choice occasion $s$ choose alternative $j$ |
| $P_{jsn}$ | Choice probability for decision maker $n$ at choice occasion $s$ choose alternative $j$ |
| $\boldsymbol{\Omega}$ | Covariance matrix of unobserved factors $\varepsilon_n$ |
| $\boldsymbol{\theta}$ | General symbol for model parameters |
| $\boldsymbol{\beta}$ | Vector of fixed coefficients |
| $\boldsymbol{\beta_n}$ | Vector of random coefficients |
| $\boldsymbol{b}$ | Mean vector of random coefficient $\boldsymbol{\beta_n}$ |
| $\boldsymbol{\Sigma_\beta}$ | Covariance matrix of random coefficient $\boldsymbol{\beta_n}$ |
| $\boldsymbol{\eta_n}$ | Matrix of MVN (0, $\boldsymbol{\Sigma_\beta}$) |
| $\boldsymbol{Y} = \{y_1, \ldots, y_N\}$ | Matrix of choices from $N$ decision makers |
| $MAE_\beta$ | Mean absolute error |
| $RMSE_\beta$ | Root mean squared estimation error |
| $RMSE_p$ | Root mean squared prediction error |

## List of Abbreviation

| | |
|---|---|
| MNL | The multinomial logit model |
| MIXL | The mixed logit model |
| LC | The latent class model |
| S-MNL | The scale heterogeneity logit model |
| G-MNL | The generalized multinomial logit model |
| MM-MNL | The mixed-mixed multinomial logit model |
| MSL | Maximum simulated likelihood method |
| HB | Hierarchical Bayesian estimation method |

# 1. Introduction

"The aim of marketing is to know and understand the consumer so well that the product or service fits him and sells itself." This can be regarded as the guideline of Peter Drucker for marketing activities.

In practice, marketers are eager to understand consumers' preferences to determine the features of the product, the best price, the communication & promotional strategies and the distribution channel in order to reach their target consumers. Companies, which can understand the heterogeneity of consumers, is offered the opportunity to explore new market needs, wants or demands that are not discovered by the competitors.

Differentiated products or services are becoming the new competitive advantage of a company, as consumers are internally diverse. Therefore, "the modelling of consumer heterogeneity is the central focus of many statistical marketing applications (Allenby & Rossi, 1999)." One useful approach is to use the discrete choice models family. The cornerstone of it is the multinomial logit model (MNL); based on it, the mixed logit (MIXL) and the latent class (LC) are the two most frequently used models for capturing heterogeneous behaviour of consumers and evaluating market preference segmentation. By claiming different sources of heterogeneity, the scale heterogeneity logit (S-MNL) and the generalized multinomial logit (G-MNL) models have been developed. The final mixed-mixed multinomial logit (MM-MNL) model is a newly developed, generalized model based on the combination of MIXL and LC models.

One popular methodology for estimating discrete choice models in the econometrics literature is called classical procedure (Revelt & Train, 1998), which makes use of the maximum likelihood approach and the other one is from Bayesian tradition, which is to utilize the hierarchical Bayesian estimation approach (Train , 2001). Asymptotically, the two procedures provide the same information. Train (2001) and Huber & Train (2001) found that these two methods produce very similar results by using their samples. In this thesis, the software used for classical estimation is Nlogit 5; for Bayesian tradition, it is estimated by the *bayesm* package in R.

The first purpose of this thesis is to replicate the estimation results of two datasets (PIZZA A and PIZZA B) from the study of Keane and Wasi (2013) and then to compare model estimations by both classical and Bayesian approaches. Finally, two simulation studies are developed in order to compare estimation and prediction accuracy of different models and different estimation procedures.

The rest of this paper is organized as follows: Chapter 2 is providing an overview of discrete choice models. Chapter 3 provides the theoretical explanation of the maximum simulated estimation and hierarchical Bayes estimation. Chapter 4 introduces the two software/ package used for analysis. Chapter 5 replicates and compares the results of two datasets (PIZZA A and B) by using Nlogit 5 and the *bayesm* package. In chapter 6, the estimation and prediction accuracy based on two simulated datasets are compared. Finally, chapter 7 gives conclusions and remarks.

# 2. Discrete Choice Models

A discrete choice model is to describe or explain the decision makers' choices among a set of alternatives (Train, 2009). It is becoming a popular technique in marketing and has become an essential tool in modelling individual behaviour (Greene & Hensher, 2003). This chapter provides an overview of six different discrete choice models.

The concept of 'Utility Maximization' is first proposed by Marschak (1960). Based on this fundamental idea, several derivations have been developed later on.

McFadden (1974) first proposed random utility models (RUMs). In a discrete choice experiment, the decision maker is faced with a number of choice occasions and each choice occasion contains several choice alternatives. For example, for decision maker $n$ in the choice occasion $s$, the choice of $i$ among $J$ alternatives means that the utility that he obtains from choice $i$ would provide the greatest utility compared with other alternatives, namely, $U_{isn} > U_{jsn}, \forall\ i \neq j$.

The utility function for each decision process is decomposed into a representative utility function ( $V_{jsn}, \forall\ j$ ) and a random function ( $\varepsilon_{jsn}, \forall\ j$ ), as expressed in equation (2.1) :

$$U_{jsn} = V_{jsn} + \varepsilon_{jsn} \tag{2.1}$$

During the decision making process, the representative utility function $\left(V_{jsn}\right)$ includes the part that is observable by researchers, such as the attributes of the alternatives ($x_{jsn}$, which is a p-dimensional vector) and is defined as:

$$U_{jsn} = \boldsymbol{\beta} x_{jsn} + \varepsilon_{jsn}$$

where $\boldsymbol{\beta}$ is the vector of fixed coefficients. Among the observable part, characteristics of decision makers $s_n$ (which is a q-dimensional vector) can also be added. This representative utility function is assumed to be linear in the parameters for each alternative. For the part of random function $(\varepsilon_{jsn})$, since it is not possible to be observed by researcher, so is treated as random and defined as

$$\varepsilon_{jsn} = U_{jsn} - V_{jsn}.$$

Based on the above notation, the probability that a decision maker $n$ would choose $i$ in choice occasion $s$ among all $J$ alternatives is expressed in equation (2.2):

$$P_{isn} = Prob\left(U_{isn} > U_{jsn}, \forall i \neq j\right)$$

$$= Prob\left(V_{isn} + \varepsilon_{isn} > V_{jsn} + \varepsilon_{jsn}, \forall i \neq j\right) \quad (2.2)$$

$$= Prob\left(\varepsilon_{jsn} - \varepsilon_{isn} < V_{isn} - V_{jsn}, \forall i \neq j\right)$$

Equation (2.2) indicates that the choice probability is the cumulative density of the error differences $(\varepsilon_{jsn} - \varepsilon_{isn})$ that is smaller than the observed quantity of $(V_{isn} - V_{jsn})$. It could also be rewritten as follows in equation (2.3):

$$P_{is} = \int_{\varepsilon} I\left(\varepsilon_{jsn} - \varepsilon_{isn} < V_{isn} - V_{jsn}, \forall i \neq j\right) f(\varepsilon_n) d\varepsilon_n \quad (2.3)$$

The indicator function $I(\cdot)$ equals $1$ when the statement in bracket is true and $0$ otherwise.

On the basis of equation (2.3), different choice models have been developed by making different assumptions on the representative function ( $V_{jsn}$ ) and error term $(\varepsilon_n)$. For example, under the assumption that the error term is identically independently distributed extreme value, the multinomial logit model is obtained.

## 2.1 Multinomial logit model (MNL)

The multinomial logit model is first proposed by Luce (1959). It is assumed that the unobserved error factor $\varepsilon_{nj}$ is identical independently distributed (iid) extreme value. In order to identify this model, the variance is normalized to the value of $\pi^2/6$ , which is also implicitly normalized the scale of utility. If $\varepsilon_{isn}\ and\ \varepsilon_{jsn}$ are iid extreme values, equation (2.4) shows that the error difference $\varepsilon^*_{ns(jn)} = \varepsilon_{jsn} - \varepsilon_{isn}$ follows logistic distribution:

$$f\left(\varepsilon_{jsn}\right) = e^{-isn} e^{-e^{-\varepsilon_{jsn}}} \quad (2.4)$$

The choice probability of MNL is derived in equation (2.5), the observable representative utility function $V_{isn}$ and $V_{jsn}$ are known, but the error component of $\varepsilon_{isn}$ on the right hand side is unknown and has to integral over all possible values of $\varepsilon_{isn}$ . Equation (2.6) expresses the derived probability of MNL, where $\boldsymbol{\beta}$ is the vector of fixed coefficients:

$$P_{isn} = Prob\left(\varepsilon_{jsn} < \varepsilon_{isn} + V_{isn} - V_{jsn}, \forall i \neq j\right) \quad (2.5)$$

$$= \frac{e^{\beta x_{isn}}}{\sum_J e^{\beta x_{jsn}}} \quad (2.6)$$

### 2.1.1 Limitations of the multinomial logit model

The MNL model has a lot of nice properties, for instance, the integral has a closed analytical form; the choice probabilities for all alternatives sum to one. However, on account of its assumptions, there are certain limitations existed in the literature (Train, 2009).

Firstly, regarding taste variation, since the coefficient $\boldsymbol{\beta}$ here is assumed homogeneous across different decision makers, MNL would only capture limited taste variations in population, such as when taste is varied systematically with reference to the observed variables. It cannot incorporate the taste variation which differs in unobserved or purely random variables (Train, 2009).

Secondly, as for the substitution patterns, the MNL model exhibits the IIA property (independence from irrelevant alternatives), which means that the probability ratio for any 2 alternatives $(P_{isn}, P_{jsn})$ will only depend on the choice $i$ & $j$, not on any other alternatives. In reality, the property is very unrealistic.

Thirdly, when dealing with repeated choices, the MNL model cannot handle the situation where unobserved factors are correlated over time (Train, 2009).

Although the MNL model is the cornerstone of different discrete choice models, it is insufficient to capture the heterogeneity of consumers' choice behaviour. Based on it, several alternative models have been proposed.

## 2.2 Mixed logit model (MIXL)

The mixed logit model might be the most frequently used model to describe respondents' heterogeneity among a number of recent developments (Greene & Hensher, 2003). It could deal with all three limitations of the MNL model and is a highly flexible model that can approximate almost any random utility functions (McFadden & Train, 2000).

The probability of MIXL model is the integral of the standard logit probabilities over a density of parameters. In other words, It is a mixture of the standard logit function evaluated at different $\beta_n$ with $f(\beta_n)$ as the mixing distribution (Train, 2009). The intuitive expression is shown in equation (2.7).

$$\pi_{isn} = \int_{\beta_n} P_{isn}(\beta_n)\, f(\beta_n) d\beta_n \tag{2.7}$$

where $P_{isn}$ is the same as equation (2.6)

It can be seen that the MNL model is nested in the MIXL model. When MIXL is degenerated at a fixed parameter $b$, with $f(b) = 1$ for $\beta_n = b$ and $f(b) = 0$ otherwise, the MIXL model reduces to the MNL model.

The essential idea of MIXL is to have random coefficients as individual-specific parameters to capture heterogeneity in respondents. When the decision maker $n$ chooses $i$ from $J$ alternatives in choice occasion $s$, the utility function is specified in function (2.8), which is in consistent with MNL and the only different is the occurrence of random effect ($\boldsymbol{\beta_n}$) for decision maker $n$. The researcher can specify any distribution for $\boldsymbol{\beta_n}$ and then estimate the parameters.

$$U_{isn} = \boldsymbol{\beta_n} x_{isn} + \varepsilon_{isn} \tag{2.8}$$
$$= (\boldsymbol{b} + \boldsymbol{\eta_n}) x_{isn} + \varepsilon_{isn}$$

in which $x_{isn}$ are the observed attributes of alternatives and characteristics of decision makers. $\varepsilon_{isn}$ still follows *iid* extreme value distribution. $\boldsymbol{b}$ is the mean vector of random coefficient $\boldsymbol{\beta_n}$. In this thesis, we restrict to the most commonly used multivariate normal distribution, namely $\boldsymbol{\beta_n} \sim MVN(\boldsymbol{b}, \boldsymbol{\Sigma_\beta})$. $\boldsymbol{\eta_n}$ is the matrix of multivariate normal (0,$\boldsymbol{\Sigma_\beta}$), with 0 mean and covariance matrix $\boldsymbol{\Sigma_\beta}$. If $\boldsymbol{\Sigma_\beta}$ is a full covariance matrix, then it is the MIXL model with correlated errors; otherwise If $\boldsymbol{\Sigma_\beta}$ is a diagonal matrix, then the estimation is MIXL with uncorrelated errors.

In summary, the MIXL model utilizes the individual – specific random parameters $\boldsymbol{\beta_n}$ to capture the respondents' heterogeneity; MIXL is also called random parameter logit model.

## 2.3 Latent class model (LC)

The latent class model has been used for analysing decision makers' heterogeneity for a long time. It is similar to MIXL, but instead of making specific assumptions of the parameter distributions, it assumes that respondents implicitly come from a set of classes (Greene & Hensher, 2003).

In equation (2.7), $f(\beta_n)$ is continuous distribution for MIXL. If $f(\beta_n)$ is specified as a discrete function with a finite set of distinct numbers, then the MIXL model is converted to LC, which is a very useful tool to explore consumers' segmentations in the data (Keane & Wasi, 2013).

This LC model assumes that consumers belong to one of the classes. Each class is a latent factor and the number of class is not known as priori (Keane & Wasi, 2013). For example, $f(\beta)$ takes $M$ possible values and the probability for each value is $\{C_1, C_2 \ldots, C_M\}$; in other words, $\boldsymbol{\beta_n}$ is different across classes but is identical within each class. The choice probability for latent class is expressed in equation (2.9).

$$P_{isn} = \sum_{m=1}^{M} C_m \left\{ \frac{e^{\beta_m x_{isn}}}{\sum_j e^{\beta_m x_{jsn}}} \right\} \tag{2.9}$$

In a typical application, the number of class is not known, so the researcher estimates models by specifying different number of classes and chooses the one which is favoured by information criteria, such as AIC, BIC or CAIC (Keane & Wasi, 2013).

## 2.4 Scale heterogeneity logit model (S-MNL)

Although MIXL has long been popular in practice, Louviere et al. (2008) argues that the source of heterogeneity in a population is coming from the error term, rather than the random coefficient as assumed by the MIXL model.

In MNL and MIXL, the error term has been implicitly normalized for model identification. If we write it explicitly, it is shown in equation (2.10)

$$U_{isn} = \beta x_{isn} + \frac{\varepsilon_{isn}}{\sigma} \tag{2.10}$$

As suggested by Louviere et al. (2008), the error term $\varepsilon_{isn}$ in equation (2.10) is heterogeneous and has a value of $\sigma_n$ for each decision maker. Thus $\sigma_n$ will replace $\sigma$ in equation (2.10) and is called the scaling factor (Fiebig, Keane, & Jordan, 2009). It is the scale of the individual specific error and the scaling factor $\sigma_n$ captures the main source of heterogeneity.

By multiplying $\sigma_n$ to both sides of equation (2.10), equation (2.11) is derived and called scale heterogeneity logit model (S-MNL model). It implies that $\boldsymbol{\beta}$ is "scaled up or down proportionately across consumers by the scaling factor $\sigma_n$ (Fiebig, Keane, & Jordan , 2009)", rather than having a random coefficient in the MIXL model ($\boldsymbol{\beta_n}$ in equation (2.8)).

$$U_{isn} = (\boldsymbol{\beta}\sigma_n) x_{isn} + \varepsilon_{isn} \tag{2.11}$$

Since $\sigma_n$ is the scaling factor, it has to be defined positively and follows a lognormal distribution: $ln(\sigma_n) \sim N(\bar{\sigma}, \tau^2)$, with the mean vector $\bar{\sigma}$ and variance-covariance matrix $\tau^2$. If S-MNL is adequate to describe the data, it is a much more parsimonious model compared with MIXL. In S-MNL, the source of heterogeneous is coming only from the error term $\varepsilon_n$ and it has one parameter, $\tau$ (Tau), which is far less parameters than the corresponding random coefficient matrix in MIXL.

## 2.5 Generalized multinomial logit Model (G-MNL)

The generalized multinomial logit model is first proposed by Keane (2010) and the estimation of G-MNL model should give insight on "whether heterogeneity is better described by scale heterogeneity, normal mixing, or some combination of the two" (Fiebig, Keane, & Jordan , 2009). The G-MNL model is a generalized model based on the combination of MIXL and S-MNL.

The utility function of choice occasion $s$ for decision maker $n$ choose $i$ is

$$U_{isn} = [\sigma_n b + \gamma\eta_n + (1-\gamma)\sigma_n\eta_n]x_{isn} + \varepsilon_{isn} \quad (2.12)$$

Equation (2.12) incorporates the MIXL and S-MNL models by adding a special weighting parameter $\gamma$ between 0 and 1. By applying different restrictions on this parameter, Table 1 indicates that the G-MNL model can degenerate to simpler models.

| **Table 1 the G-MNL Model and Its Special Cases** | |
|---|---|
| When $\gamma = 1$ and $var(\eta_n) = 0$, | G-MNL $\longrightarrow$ S-MNL |
| When $\gamma = 0$ and $\sigma_n = \sigma = 1$, | G-MNL $\longrightarrow$ MIXL |

Similarly as in the S-MNL model, the scale parameter is defined positively and follows a lognormal distribution, $LN(\sigma_n) \sim N(\bar{\sigma}, \tau^2)$.

## 2.6 Mixed-mixed multinomial logit model (MM-MNL)

In the literature of discrete choice models, mixture-of-multivariate normals are used as an "alternative flexible distribution" to incorporate heterogeneity (Keane and Wasi, 2013). The mixed-mixed multinomial logit model is developed based on the MIXL model by specifying the mixture-of-multivariate normals as the mixing distribution rather than having a unimodal distribution in the MIXL model.

The MM-MNL model can also be regarded as extension of the LC model, which allows the preference heterogeneity within each class by the random parameters (In the LC model as discussed in section 2.3, homogenous preference is assumed within each class). Therefore, MM-MNL is also called the "random parameter latent class model" or the "latent class mixed multinomial logit model" (Greene & Hensher, 2013).

Using the same notation as before, the utility function of MM-MNL is

$$U_{isn} = \boldsymbol{\beta_n} x_{isn} + \varepsilon_{isn} \tag{2.13}$$

where $\boldsymbol{\beta_n} \sim MVN(\beta_m,\ \Psi_m)$ with probability $C_m,\ \sum_m C_m = 1$ and $C_m > 0,\ \forall m$.

Table 2 shows that MIXL and LC are just two special cases of MM-MNL:

**Table 2 MM-MNL Model and its Special Cases**

| | |
|---|---|
| When $C_m \to 0$ for all but one class; It means there is only one class in the model. | MM-MNL $\longrightarrow$ MIXL |
| When $\Psi_m \to 0$; it means that there is no heterogeneity in each latent class. | MM-MNL $\longrightarrow$ LC |

The choice probability of MM-MNL is

$$P_{isn} = \sum_{m=1}^{M} C_m \left\{ \int \left[ \prod_{S} \prod_{i} \left( \frac{e^{\beta_m x_{isn}}}{\sum_j e^{\beta_m x_{jsn}}} \right)^{y_{isn}} \right] f(\beta_m) d\beta_m \right\} \tag{2.14}$$

# 3. Estimation Methods

An important part of this thesis is to compare the results from both maximum simulated likelihood estimation and hierarchical Bayesian estimation methods. An overview of these two procedures is given in this chapter.

## 3.1 Maximum Simulated Likelihood Method (MSL)

### 3.1.1 Choice Probabilities and integration

As discussed in the MIXL model, the probability of a certain choice is a multidimensional integral over all possible values of unobserved factors. If we want to calculate the probabilities of choice alternatives, the integrals must be evaluated.

For the simple MNL model, the integral of the probability expression has a closed-form and the choice probability can be calculated directly. In the MIXL model, the integral does not have a closed-form and cannot be calculated analytically; thus, it has to be evaluated numerically, in most cases via simulation.

"Simulation relies on the fact that integration over a density is a form of averaging (Train, 2009)." This is the fundamental for all simulation methods. Although getting exact analytical integrals are rarely possible, by using large number of draws, approximation is often plausible. In general, the more draws are used in the simulation; the more accurate the results.

For the MSL method, simulation gives researchers freedom to approximate the choice probability by specifying a model without too many assumptions as the resulting probability expression does not have to be a closed-form.

### 3.1.2 Maximum Simulated Likelihood for the discrete choice models

The likelihood function is a function of parameters for a statistical model. Given $\theta$ as the symbol of model parameters, it "expresses the plausibility of the observed data as a function of the parameters (Lesaffre & Lawson, 2012)".

The maximum likelihood method (MLE) makes use of the product of the probability densities as the measure of the consistency of the parameters with the sample data. The product is called the likelihood value of the parameter $\theta$ and is denoted as $L(\theta)$ (Kutner, Nachtsheim, Neter, & Li , 2004). If $\theta$ is consistent with the sample, the probability will be relatively large and also the likelihood values; if $\theta$ is not consistent with the data, the probabilities and the likelihood values will be relatively small.

The MLE of $\theta$ is the one which maximizes the likelihood value. The procedure of MSL is the same as the MLE and the only difference is that simulated probabilities are used in the maximum process rather than exact probabilities (Train, 2009).

In the estimation of maximum likelihood, the log-likelihood function is:

$$LL(\theta) = \ln\left\{\prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{j=1}^{J}\left(P_{jsn}(\theta)\right)^{y_{jsn}}\right\} \tag{3.1}$$

in which the $\theta$ is a vector of unknown parameters, $P_{jsn}(\theta)$ is the probability of the observed choice for decision maker $n$. And $y_{isn} = 1$ if decision maker $n$ chooses $j$ in choice occasion $s$ and zero otherwise. The function is to multiply over $N$ independent decision makers and $S$ choice occasions. MLE is the value of $\theta$ that maximizes $LL(\theta)$. Given the fact that the gradient of $LL(\theta)$ is zero at the maximum, the MLE is calculated as:

$$\frac{\partial LL(\theta)}{\partial\theta} = \frac{\partial \ln}{\partial\theta}\left\{\prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{j=1}^{J}\left(P_{jsn}\right)^{y_{jsn}}\right\} = 0 \tag{3.2}$$

Using the expression of $p_{isn}(\hat{\theta})$ as the simulated probabilities of $P_{isn}(\theta)$, the MSL is defined similarly in equation (3.3):

$$\frac{\partial LL(\theta)}{\partial\theta} = \frac{\partial \ln}{\partial\theta}\left\{ln\left[\prod_{n=1}^{N}\prod_{s=1}^{S}\prod_{j=1}^{J}\left(p_{isn}(\hat{\theta})\ \right)^{y_{jsn}}\right]\right\} = 0 \tag{3.3}$$

The concern of MSL lies in the bias of $p_{isn}(\hat{\theta})$ ; however, the bias diminishes as more draws are utilized in the simulation. If the number of draws ($R$) increases faster than $\sqrt{N}$, the MSL estimator is consistent, efficient, asymptotically equivalent to MLE (Train, 2009).

## 3.2 Hierarchical Bayesian Estimation Method (HB)

The hierarchical Bayesian estimation method is one of the alternative estimation methods for estimating discrete choice models. In 1999, Allenby and Rossi showed how the Bayesian procedure could be used to assess individual – level parameters within a model with random parameters and since then Bayesian procedure is becoming an alternative to the classical procedure (Rossi, Allenby, & McCulloch, 2006).

The HB estimation includes two parts: first, a model that is written in a set of sub-models; second, the model is estimated by Bayesian methods. The combination of hierarchical Bayesian model and the MCMC method works well and has made it flexible for making inferences for both fixed effect coefficients and the estimates at consumer level (Allenby & Rossi, 1999). In contrast to classical methodologies, HB is a more powerful procedure to make inference on heterogeneous consumers and more flexible to explore consumers' diversities.

In this chapter, first an introduction of Bayesian concepts and its inference is given; then it is followed by a brief overview of MCMC estimation methods. Finally, the HB method is discussed in detail with respect to the estimations of MIXL and MM-MNL models.

### 3.2.1 General Concepts of Bayesian Inference

The essential concept in Bayesian analysis is to combine prior knowledge or past experiences (prior probability) to the current study (the likelihood of the data) in order to update the final conclusion (posterior probability).

The prior distribution is some initial knowledge about the unknown parameter $\theta$ and denoted as $k(\theta)$. It can be the historical data or the experts' knowledge. If there is no prior knowledge available for $\theta$, a non-informative prior can also be constructed, in that case, the posterior distribution is based almost entirely on the current data (Agresti, 2013).

The information of data collected in the current study is represented by the likelihood function, $L(\boldsymbol{Y} \mid \theta)$ and researchers could update or change their prior knowledge about $\theta$ based on the likelihood function. The updated information is summarized by a new distribution of $\theta$ (denoted as $K(\theta \mid \boldsymbol{Y})$ ) and called posterior distribution. By applying Bayes' rule, the posterior distribution is defined as (Train, 2009):

$$K(\theta \mid Y) = \frac{L(Y \mid \theta)\, k(\theta)}{L(Y)} \tag{3.4}$$

Where $L(Y)$ is the marginal probability of Y, integrated over $\theta$

$$L(Y) = \int L(Y \mid \theta)\, k(\theta) d\theta$$

Since $L(Y)$ is a normalizing constant with respect to $\theta$, equation (3.4) could be expressed in a way that the posterior distribution is proportional to the product of prior distribution and the likelihood function:

$$K(\theta \mid Y) \propto L(Y \mid \theta)\, k(\theta) \tag{3.5}$$

For model comparison with the MSL, the posterior mean can be calculated by:

$$\bar{\theta} = \int \theta \, K(\theta \mid Y) \, d\theta$$

During the integration process, the method of simulation is usually applied. In general, the simulated mean from the posterior is computed by taking draws from the posterior distribution and then average the total $R$ draws (Train, 2009).

$$\check{\theta} = \frac{1}{R} \sum_{r=1}^{R} \theta^r$$

where $\theta^r$ is the $r$th draw from posterior $K(\theta \mid Y)$

By this way, it allows the estimator to be implemented even if the integral that defines the estimator does not have a closed form (Rossi, Allenby, & McCulloch, 2006).

### 3.2.2 MCMC Methods

The MCMC method includes two important sampling algorithms, Metropolis – Hasting and Gibbs-sampler. For both algorithms, some initial draws have to be taken before reaching the equilibrium; this part is called the burn-in part. The values in the burn-in part are discarded and only the remaining values are used to make inferences on the posterior distribution.

#### 3.2.2.1 Metropolis – Hasting (MH) Algorithm

The Metropolis algorithm was first proposed by Metropolis and his colleagues (1953). Later on, Hastings (1970) developed this algorithm and applied in the statistical world. Now the algorithm is called the Metropolis-Hasting algorithm in general.

The idea of MH algorithm is to generate a series of random samples based on a proposal distribution and decide whether to accept or reject the proposed moves. It is similar to the Accept - Reject algorithm; however, the MH algorithm is more efficient and applicable for high-dimensional parameters and generates dependent samples.

Using the same notation as before, if we want to sample from posterior distribution $p(\theta|\boldsymbol{Y})$, the procedure to obtain the Metropolis – Hasting sampled value for $\theta^{t+1}$ is shown in Table 3.

**Table 3: Metropolis – Hasting Algorithm (Rossi, Allenby, & McCulloch, 2006)**

(1) . Choose the starting values of $\theta^0$
(2) . At iteration $(t+1)$, sample a candidate $\theta^*$ from proposal distribution $q_{t+1}(\theta^*|\theta^t)$
(3) . Compute an acceptance ratio (probability): $r = \frac{p(\theta^*|Y) / q_{t+1}(\theta^*|\theta^t)}{p(\theta^{t-1}|Y) / q_{t+1}(\theta^t|\theta^*)}$
(4) . Draw $U \sim \text{Uniform}(0,1)$
(5) . The next value of $\theta^{t+1}$ equals:
   a. Accept $\theta^*$ with probability $\alpha = \min(r, 1)$
   b. Reject $\theta^*$, then $\theta^{t+1} = \theta^t$

Repeat step (2)~step(4), as necessary

The proposal distribution $q_{t+1}(\theta^*|\theta^t)$ decides where the next iteration of Markov chain and the MH algorithm will be more efficient if the proposal distribution is more close to the posterior distribution.

If we have a symmetric proposal distribution that is dependent on $\theta^t$, such as $\theta^* = \theta^t + \epsilon$, with $\epsilon \sim N(0, \Omega)$, then $q_{t+1}(\theta^*|\theta^t) = q_{t+1}(\theta^t|\theta^*)$ and the corresponding $r = \frac{p(\theta^*|\boldsymbol{Y})}{p(\theta^{t-1}|Y)}$, in this case, this algorithm is called “random walk Metropolis-Hastings sampling”. If the proposal distribution does not depend on $\theta^t$, with $q_{t+1}(\theta^*|\theta^t) = q_{t+1}(\theta^*)$, it is called “independent Metropolis –Hastings sampling”.

### 3.2.2.2 Gibbs sampler algorithm

The term Gibbs sampler is first introduced by Geman and Geman (1984) in the image processing context. Later in 1990, Gelfand and Smith developed the concept of Gibbs sampler and applied it for model estimation in the Bayesian context.

Mathematically, the Gibbs sampler is a special case of the MH algorithm. Both are MCMC algorithms for generating a sequence of samples which are approximation of the target distribution under mild conditions (Lesaffre & Lawson, 2012).

Using the same notation as before, if we want to sample from posterior distribution $p(\theta|\boldsymbol{Y})$ **,** suppose the chain is at the current value of $(\theta_1^t, \theta_2^t, \dots, \theta_d^t)$ ( $d$ dimensional unknown parameter $\boldsymbol{\theta}$ at the $t$ th iteration), the procedure for general Gibbs sampler at iteration $(t+1)$ is summarized in Table 4:

| Table 4: The general Gibbs sampler(Lesaffre & Lawson, 2012) |
| --- |
| (1) Choose the starting values of $\theta^0$<br>(2) Update $\theta_1^{(t+1)}$ from $p(\theta_1 \mid \theta_2^t, \ldots, \theta_{(d-1)}^t,\ \theta_d^t, \boldsymbol{y})$<br>(3) Update $\theta_2^{(t+1)}$ from $p(\theta_2 \mid \theta_1^{(t+1)}, \theta_3^t \ldots,\ \theta_d^t, \boldsymbol{y})$<br><br>(d) Update $\theta_d^{(t+1)}$ from $p(\theta_d \mid \theta_1^{(t+1)}, \ldots, \theta_{(d-1)}^{(t+1)},\ \boldsymbol{y})$<br>Repeat step (2)~step(d) as necessary |

In this way, under mild regularity conditions, the sampled observations $\theta^t, \theta^{t+1}, \ldots$ are observations from the target posterior distribution (Lesaffre & Lawson, 2012).

### 3.2.3 The hierarchical Bayesian estimation for the MIXL model

If using the utility function of decision maker $n$ in choice occasion $s$ for the $i$th choice, the hierarchical form of MIXL model is expressed in Table 5.

| Table 5: Hierarchical model expression for MIXL (Rossi, Allenby, & McCulloch, 2006): | | |
| --- | --- | --- |
| Utility function of Mixed Logit model: | Hierarchical model expression: | |
| $U_{isn} = \boldsymbol{\beta_n} x_{isn} + \ \varepsilon_{isn}$<br>(Where $\varepsilon_{isn}$ is iid extreme value) | $\boldsymbol{\ell}(\boldsymbol{\beta_n} \mid y_n, x_{isn})$ | (3.6) |
| $\boldsymbol{\beta_n} \sim iid\ N(\boldsymbol{b}, \boldsymbol{\Sigma_\beta})$ | $\boldsymbol{\beta_n} \mid \boldsymbol{b}, \boldsymbol{\Sigma_\beta}$ | (3.7) |

in which $\boldsymbol{\beta_n}$ represents the individual - level parameters for each decision makers and it is useful to describe the heterogeneous tastes across respondents. b & $\Sigma_\beta$ are hyper-parameters, for which there is a hierarchy of priors expressed in Table 6.

| Table 6: Priors for MIXL model (Rossi, Allenby, & McCulloch, 2006) |
| --- |
| $b \sim N\ (b_0, A^{-1})$<br>with very diffuse prior $A^{-1} = 100I$ or larger |
| $\Sigma_\beta \sim Inverse\ Wishart(\nu, V)$<br>$\nu$ is the degree of freedom and $V$ is $\nu$- dimensional identity matrix |

By applying Bayes' Rule, the hierarchical model for the posterior distribution is specified as a set of conditional distributions (Rossi, Allenby, & McCulloch, 2006):

$$P\left(b, \Sigma_\beta \,\middle|\, \boldsymbol{Y}\right) \propto \prod_n L\,(y_n \mid b, \Sigma_\beta)\; p(\mathrm{b}, \Sigma_\beta)$$

in which $p(\mathrm{b}, \Sigma_\beta)$ is the specified prior and $P\left(b, \Sigma_\beta \,\middle|\, Y\right)$ is the posterior distribution based on a given prior and the current observations.

It is possible to make draws directly from the posterior distribution by MH algorithm; however, it is computationally very slow (Train, 2009). By considering $b$ and $\Sigma_\beta$ as the parameters for $\beta_n$, the posterior distribution can also be expressed as below:

$$P\left(b, \Sigma_\beta, \beta_n, \forall n \,\middle|\, \boldsymbol{Y}\right) \propto \prod_n L\,(y_n \mid \beta_n)\; \phi(\beta_n \mid \mathrm{b}, \Sigma_\beta)\, p(\mathrm{b}, \Sigma_\beta)$$

Figure 1 gives the diagram of drawing order within each simulation (Rossi, Allenby, & McCulloch, 2006):

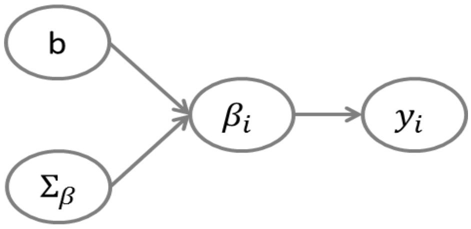


**Figure 1: Diagram of the HB estimation for MIXL Model**

Train (2009) suggests that one possible way of making draws for the MIXL model is to use the Gibbs sampler. The procedure is summarized in Table 7:

**Table 7: HB estimation for MIXL model (Train, 2009):**

Suppose the starting values $b^0$, $\Sigma_\beta^0$ and $\beta_n^0$, $\forall n$, the $t$ th iteration for each parameters:

1) Draw $b^t$ from $N\,(\overline{\beta^{t-1}}, \frac{\Sigma_\beta^{t-1}}{n})$, where $\overline{\beta^{t-1}}$ is the mean of all draws of $\beta_n^{t-1}$
2) Draw $\Sigma_\beta^t$ from $IW\,(V+N, \frac{VI+N\,S^{t-1}}{V+N})$,

   where $s^{t-1} = \sum_n (\beta_n^{t-1} - b^t)(\beta_n^{t-1} - b^t)'/N$
3) For each n, draw $\beta_n^t$ by one iteration of MH algorithm based on the normal distribution $\phi\,(\beta_n \mid b^t, \Sigma_\beta^t)$

The iteration process has to be repeated many times until convergence. The draws after the burn-in part will converge to the real values from the joint posterior distribution. Then the mean and standard deviation can be calculated in order to obtain the estimates and standard errors (Train, 2009).

A relatively diffuse prior is used in this analysis. More complex priors can also be used as introduced by McCulloch, Polson and Rossi (2000).

### 3.2.4 The hierarchical Bayesian estimation method for MM-MNL model

The mixture components of MM-MNL provide a great deal of added flexibility (McCulloch, Polson, & Rossi, 2000). The estimation of MIXL provides the basis for estimation of the MM-MNL model. In this section, an overview of the HB estimation for the MM-MNL model is presented.

By using the same notation, the utility function and the corresponding hierarchical model expressions for $M$ mixture components are shown in Table 8. Table 9 summaries the corresponding priors.

**Table 8: Hierarchical Model expression for MM-MNL model (Rossi, Allenby, & McCulloch, 2006):**

| Utility function for MM - MNL model | Hierarchical model expression |
|---|---|
| $U_{isn} = \boldsymbol{\beta_n} x_{isn} + \varepsilon_{isn}$ | $\boldsymbol{\ell}(\boldsymbol{\beta_n} \mid y_n, x_{ns})$ |
| $\boldsymbol{\beta_n} \sim MVN(\beta_m, \Psi_m)$<br><br>$\boldsymbol{\beta_n}$ with probability $C_m$, $\sum_m C_m = 1$ and $C_m > 0, \ \forall m$. | $\boldsymbol{\beta_n} \sim N(b_{ind_i}, \Sigma_{ind_i})$<br><br>$ind_i \sim Multinomial_M(pvec)$<br>*Where $ind_i$ is the indicator of latent variable for which component observation $i$ is from and it takes on value $1, \ldots, M$.*<br>*$pvec$ is a vector of mixture probabilities of length $M$.* |

**Table 9: Priors for MM-MNL model (Rossi, Allenby, & McCulloch, 2006)**

| |
|---|
| $pvec \sim Dirichlet\ (\alpha)$<br><br>$b_m \sim N\left(\bar{b}, \Sigma_{\mathrm{m}} \otimes A^{-1}\right)$ with very diffuse prior $A^{-1} = 100I$ or larger |
| $\Sigma_\beta \sim IW(\nu, V)$<br><br>$\nu$ is the degree of freedom and $V$ is $\nu -$ dimensional identity matrix |

It is similar to the MIXL model that the estimation procedure can be written as a set of conditional models and the ordering is shown in Figure 2 :

$$p(pvec, ind, b_m, \Sigma_m, \beta_n, \forall n \mid \boldsymbol{Y}) \propto \prod_n L\,(y_n \mid \beta_n)\, \phi(\beta_n \mid b_n, \Sigma_m)\ p(b_n, \Sigma_m \,|ind, pvec)\ f(pvec\,|ind)\ q(ind)$$

in which $ind_i \sim multinomial\ (pvec_i)$ and given $ind_i, pvec \sim Dirichlet\ (\alpha)$.

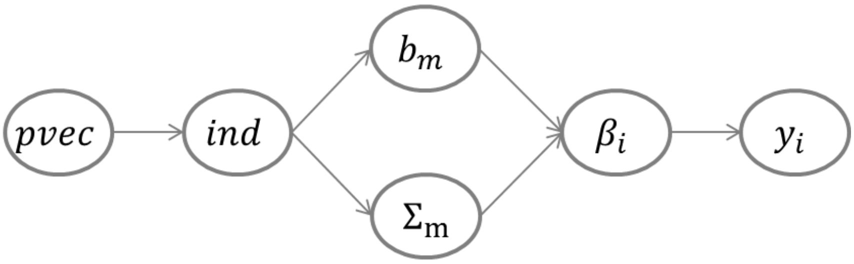


**Figure 2: Diagram of the HB estimation for MM-MNL Model**

It is possible to make draws directly from the joint posterior distribution ( $p(pvec, ind, b_m, \Sigma_m, \beta_n, \forall n \mid \boldsymbol{Y})$ ) with a Metropolis step; given the mixture of multivariate normals, empirically, Metropolis algorithm is not only slow but it is also very difficult to reach a good performance for high dimensions. Rossi, Allenby and McCulloch (2006) define an MCMC chain of Gibbs style for the estimation which is outlined in Table 10.

| **Table 10: HB estimation for MM-MNL model** |
|---|
| The $t$th iteration for each parameters:<br>1) To make draws from $q(ind)$, the draw of the indicators is a multinomial draw: $ind_i \sim multinomial\ (pvec_i)$ where $pvec' = (pvec_{i,1},,,pvec_{i,M})$<br>And $pvec_{(i,m)} = \frac{pvec_m\ \phi\ (\theta_i^{t-1}\ \vert b_m, \Sigma_m)}{\sum_m pvec_m\ \phi\ (\theta_i^{t-1}\ \vert b_m, \Sigma_m)}$<br>2) To make draws from $f(pvec\ \vert ind)$, the draws of $pvec$, given the indicators $ind$, is a Dirichlet draw: $pvec \sim Dirichlet\ (\bar{\alpha})$ where $\bar{\alpha} = n_m + \alpha_m$<br>And $n_m = \sum_{i=1}^{N} i\ (ind_i = m)$<br>3) To make draws for $b_m^t, \Sigma_m^t$ and $\beta_n^t$ , which is similar to the algorithm introduced in Table 7. |

For the estimation of the discrete choice models, it is very common to embed the MH algorithm within steps of the Gibbs sampler and it is called the Hybrid Gibbs sampler. For example, when the full distribution is known, but not trivial to compute or the full conditions for one or some of the parameters are not known in the closed form. In that case, it is possible to use different MH variants within steps of the Gibbs sampler, such as random walk Metropolis-Hasting or Metropolis algorithms (Guindani, 2008).

# 4. Software

Chapter 4 introduces the software/package which is used for estimating the discrete choice models in this thesis. Section 4.1 introduces the software Nlogit 5.0, which is used for estimation by the maximum simulated likelihood (MSL) (Greene, William H.; Econometric Software, Inc, 2012). Section 4.2 is about the *bayesm* package in R which applies the hierarchical Bayesian estimation method (HB) (Rossi, Allenby, & McCulloch, 2006).

## 4.1 Nlogit Version 5.0

Nlogit is a major suite of software for the estimation of discrete choice models (Greene, William H.; Econometric Software, Inc, 2012). It is an extension of the econometric and statistical software LIMDEP Version 6, which provides more features for newly developed discrete choice models. Each discrete choice model introduced in Chapter 2 associates to a specific function for estimation and is summarized in Table 11 :

**Table 11: Relative Functions of *Nlogit* (Version 5)**

| Function | Discrete Choice Model |
|---|---|
| *NLOGIT OR CLOGIT* | *Multinomial logit model (MNL)* |
| *RPLOGIT* | *Mixed logit model (MIXL) or random parameter logit model* |
| *LCLOGIT* | *Latent Class model (LC)* |
| *SMNLOGIT* | *Scale-heterogeneity logit model (S-MNL)* |
| *GMXLOGIT* | *Generalized multinomial logit (G-MNL)* |
| *LCRPLOGIT* | *Mixed-mixed multinomial logit model (MM-MNL)* |

Nlogit 5.0 utilizes the maximum likelihood estimator for discrete choice data and panel data from repeated observations and choice situations (Greene, William H.; Econometric Software, Inc, 2012).

## 4.2 The *bayesm* Package in R

The *bayesm* package is developed by Rossi (2005) as accompany of the book titled "Bayesian Statistics and Marketing" (Rossi, Allenby, & McCulloch, 2005). It introduces the Bayesian methods in the context of marketing problems and also illustrates the practical use of the MCMC method. However, the *bayesm* package could only handle three out of six models introduced in Chapter 2. The corresponding functions are summarized in Table 12.

| **Table 12: Relative Functions of the *bayesm* package (Version 2.2-5)** | |
|---|---|
| Function | Discrete Choice Model |
| *rmnlIndepMetrop* (Independence Metropolis for the MNL) | *Multinomial logit model (MNL)* |
| *rhierMnlRwMixture* *(*A hybrid Gibbs Sampler with a RW Metropolis step for the MNL coefficients for each panel unit.*)* | *Mixed logit model (MIXL)* |
| | *Mixed-mixed multinomial logit model (MM-MNL)* |

"*rhierMnlRwMixture*" function is a MCMC algorithm for the hierarchical multinomial logit model with a mixture of normals heterogeneity distribution (Rossi, 2012). When only one normal distribution is to capture the heterogeneity, it estimates the MIXL model. If a mixture of normals is specified for the prior distribution, this function is used to estimate the MM-MNL model. In this thesis, the *bayesm* package version 2.2-5 is used for model estimation.

# 5. Empirical results of two experiments from Keane and Wasi's (2013) study

Although there is a fast development of discrete choice models in recent years, only few papers have compared different approaches for modelling consumer heterogeneity (Kragt, 2013) and most of them only compare 2~4 models. Similarly, although Bayesian estimation becomes popular recently, so far, few papers have compared model performances the different estimation procedures (e.g., (Train, 2001)). This thesis is based on the study of Keane and Wasi (2013), first to replicate two experiment results of six discrete choice models by MSL procedure and then to re-estimate the MNL, MIXL and MM-MNL models by HB procedure, finally the estimates from the two procedures are compared.

In the study of Keane and Wasi (2013), ten empirical data sets are estimated by MATLAB by utilizing the MSL estimation procedure. Of ten datasets, two experiments about PIZZA delivery services are investigated in detail. PIZZA A is a relatively small datasets with 8 attributes and 2848 observations. PIZZA B is a fairly large datasets with doubled attributes (#16) and 10496 observations. The structured details are given in Table 19 in the appendix. The datasets were collected by the Centre for the Study of Choice (CenSoC), University of Technology Sydney (Fiebig, Keane, & Jordan, 2009).

## 5.1 Some notes on the model estimation

Same choices are presented for each decision makers in both experiments and they are labelled "A" and "B" for each decision maker. Thus, no ASC (alternative specific constant) is needed in the model specification (Keane & Wasi, 2013).

There is no attribute about the characteristic of decision makers ($s_n$) in both datasets, so that we will neglect the influence of choice - invariant variables in both estimation procedures.

During the model estimation process, the number of classes for the LC model and the number of mixture components for the MM-MNL model are the same with the ones used by Keane and Wasi (2013).

For the models with scale heterogeneity factors, such as S-MNL and G-MNL, due to some technical reasons given by the support team of Nlogit 5, the reported values of tau (variance parameter) and gamma (weighting parameter) are wrong due to one bug in Nlogit5. The estimates are still included in the summary table for completeness.

For model comparison between the MSL and HB procedures, “the mean of the posterior distribution is interpreted as being asymptotically equivalent to the MLE from a classical perspective (Train, 2009).” Therefore, the posterior mean is reported as the estimate from HB procedure.

During the integration and simulation estimation processes, different number of draws (#200, #500 and #1000) has been tried for all models. The reported results are the best ones that are favoured by BIC.

## 5.2 Empirical Results from Maximum Simulated Likelihood estimation

Table 20 and Table 21are the summary tables for the results from Keane and Wasi (2013) and the model estimates by using Nlogit 5.

Fiebig, Keane and Jordan (2009) recommended to use different information criteria in conjunction for model selection: BIC and/or CAIC are reliable measures for testing whether scale heterogeneity exists; AIC is useful for evaluating the existence of error correlation. In this thesis, all three information criteria are reported. Table 13 is the summary table for AIC, BIC and CAIC.

**Table 13: Summary of Information Criteria**

| Information Criteria | Formula<br>$K$: Number of parameters<br>$N$: The number of observations<br>$\ell$ : The likelihood function of the specified choice model | Comments |
|---|---|---|
| AIC | $AIC = -2\ell + 2K$ | AIC prefers complex models, might lead to overfitting. |
| BIC | $BIC = -2\ell + Kln(N)$ | BIC prefers parsimonious models, might lead to underfitting. |
| CAIC | $\mathrm{CAIC} = -2\ell + K(1 + \ln(N))$ | CAIC prefers parsimonious models, might lead to underfitting. |

Table 20 in the appendix presents the results for the PIZZA delivery experiment A (PIZZA A). Only attributes of the PIZZA and the delivery service are included in the model, such as gourmets, price and delivery time; there is no characteristic about the decision makers, such as income, family size and etc.

In the study of Fiebig, Keane and Jordan (2009) and Keane and Wasi (2013), they found that G-MNL with uncorrelated error outperformed all others with the smallest BIC and CAIC values (the smallest AIC value favours MM-MNL model with the largest number of parameters). In Nlogit 5, we find that MIXL is the best performing model with the smallest values for all information criteria, AIC, BIC and CAIC.

The results for the MNL and the LC models are almost the same compared with the reported ones from Keane and Wasi (2013), with slight differences in the second-decimal-point of attributes coefficients in the LC model.

The third and fourth columns are the estimates from S-MNL and G-MNL with scale heterogeneity factors. Due to one bug existed in Nlogit 5, the estimates of Tau and Gamma are wrong. This study identifies the same significant attributes with the expected sign compared with original study (refer to Keane and Wasi (2013)) at 5% the significant level; however, the estimates are not very close in magnitude as for other models.

The model estimate of MIXL (with uncorrelated error covariance matrix) gives an unexpected improvement in the likelihood function from $-1403$ to $-1382$ (the reported result from Keane and Wasi (2013)). Given the same number of parameters (#16), this leads to improvements in all information criteria accordingly. Note that Nlogit 5 also identify the same sets of significant attributes with more or less same estimates (e.g., with only one estimate deviates from 0.96 to 1.13; all others only have small differences at the second – decimal point.)

The final model is MM-MNL. By using the same number of parameters (#33), the performance of Nlogit 5 is poor, with a log-likelihood 198 points less than the one from Keane and Wasi (2013). One possible reason is because Nlogit 5 fails to reach the global maximum by the defaulted starting values. If other starting values could be tested, the model might be improved.

Table 21 summaries the results for the PIZZA delivery experiment B (PIZZA B). Compared with PIZZA A, PIZZA B is a much larger dataset, including doubled attributes (from 8 to 16), almost twice the number of respondents (from 178 to 328) and around 3.7 times more number of observations (from 2848 to 10496).

In the study of Keane and Wasi (2013), all three information criteria are in favour of the MM-MNL model due to the increased complexity in data. Based on the BIC criteria, G-MNL is following MM-MNL as the second best one (BIC is 11527 for MM-MNL and 11693 for G-MNL) and is way ahead the 3$^{rd}$ best, the MIXL model (BIC: 12080). However, in this analysis using Nlogit 5, MIXL is found as the best performing model with the support of all three information criteria.

The first column is the simple MNL model. Not surprisingly, both studies produce the same estimates. The next one is S-MNL, although the reported estimation of Tau is not interpretable, all the other coefficient estimates are quite similar (with only up to $0.04$ difference in magnitude). The estimates of G-MNL for both studies are similar, but not very close.

The fourth one is MIXL with uncorrelated errors. With the same number of parameters (#32), there is a substantial improvement in log-likelihood, which is $-5872$ compared with $-5892$ for Keane and Wasi (2013). Both studies recognize the same statistical significant attributes at 1% level and most estimates are quite close (within 1 standard deviation, with only 1 estimate (for “crust”) deviating around 3 standard deviations).

The fifth one is LC with 6 latent classes. There is a 23 points improvement in the log-likelihood with the cost of 6 additional parameters compared with results from Keane and Wasi (2013). By specifying the same model, the estimated variance matrix is singular in Nlogit 5, thus an extra constant is added in each class for avoiding singularity. The results are very close with the original study, including the class probabilities, significant attributes and the estimates of attribute coefficients.

For the final MM-MNL model, similar to the PIZZA A experiment, it again fails to reach global maximum with the default starting values and reports poor fitting, with a log-likelihood value of $-6279$ compared with $-5310$ from Keane and Wasi (2013).

Using Nlogit 5, the performance of MIXL is enhanced, such as the estimates always give better fit function (e.g., 21 and 20 points improvement for the likelihood function for two experiments with the same number of parameters) compared with the reported MIXL model by Keane and Wasi (2013).

A summary of all six models is presented in Table 14.

**Table 14: Summary for estimates from Nlogit 5.0 and Keane and Wasi (2013)**

| | |
|---|---|
| *MNL* | Both studies have the same results. |
| *S-MNL* | Both studies produce very similar estimates, with same LL function and similar attribute coefficients. |
| *G-MNL* | The fit of Nlogit 5 is a little worse than the reported value from Keane and Wasi (2013). The estimates of coefficients have the expected sign but are not very close. |
| *MIXL* | Nlogit 5 reports better fit results; similar estimated coefficients. |
| *LC* | Both studies have very similar estimates. |
| *MM-MNL* | Nlogit 5 fails to reach the global maximum. Keane and Wasi (2013) report a better fitting result. |

## 5.3 Empirical Results from Hierarchical Bayes Estimation

As discussed in Chapter 3.2, the hierarchical model is specified by a set of conditional distributions and proper priors are needed to construct the posterior distribution. By applying the settings in the *bayesm* package, the priors for both MIXL and MM-MNL can be rewritten as below (Rossi, 2012):

$$\boldsymbol{\beta_n} \sim iid\ N(\boldsymbol{b}, \boldsymbol{aV_\beta})$$

$$\boldsymbol{V_\beta} \sim Inverse\ Wishart(\boldsymbol{\nu}, \boldsymbol{V_0})$$

where a vector of Dirichlet prior parameters and the length equals the number of components. $\nu$ and $\boldsymbol{V_o}$ are the degree of freedom and location parameters for the Inverse Wishart prior. The Bayesian procedure reports the estimates and full covariance matrix of $\boldsymbol{\beta_n}$.

A non-informative prior is employed for a fair comparison, more specifically:

$$\boldsymbol{b} = \boldsymbol{0}\ ;$$

$$\boldsymbol{a} = \boldsymbol{5}\ (\boldsymbol{default});$$

$$\boldsymbol{\nu} = \boldsymbol{K}\ (\boldsymbol{no.of\ \ parameters});$$

$$\boldsymbol{Vo} = \boldsymbol{KI}$$

In the analysis, different values of $a$ and $\nu$ have been tried and the results are not sensitive to this change, as predicted by Train (2001), it indicates that the prior is non-informative compared to the likelihood when there is a large number of observations.

For the hierarchical Bayesian analysis in this section, the number of iterations/draws is ranging from $30{,}000$ to $120{,}000$ iterations depending on the examination of the trace plot. The first half of the chain is discarded as burn-in part.

The output of the *bayesm* package is quite different from Nlogit 5. Since HB does not involve a maximization process, no log-likelihood value is reported in R. The *bayesm* package focuses on the overall mean of parameters instead of the mixture components; no class probability and attribute coefficient is reported for each component. Therefore, it is difficult to compare the performance of the MM-MNL model by different procedures.

Table 22 and Table 23 are the summary results for PIZZA A and PIZZA B experiments. For the simple MNL model, both Nlogit 5 and the *bayesm* package report the same estimates as the original study. Given the full covariance matrix setting in the *bayesm* package, we compare the MIXL model with correlated errors from Keane and Wasi (2013). In the PIZZA A experiment, the fit (*LL*) from Nlogit 5 is almost the same as the reported value, but the estimates are not as close as the ones from the *bayesm* package. The same is also found in the PIZZA B experiment, the mean of the posterior distribution from the *bayesm* is closer to the estimates of original study.

The number of parameters estimated in MIXL (correlated error) is different from that is in Nlogit5 (#152) and in the original study (#48). Instead of using a full variance-covariance matrix in Nlogit 5, Keane and Wasi (2013) made a compromise and set it for all classes to be proportional, namely, $\Sigma_m = k_m\Sigma$, where $\Sigma$ is a full variance-covariance matrix. In this way, many parameters are saved.

For the HB estimation of MM-MNL model, the class (or cluster) probabilities are calculated based on the characteristics of decision makers, since there is no such attribute in the PIZZA delivery experiments, it is not estimable in HB. In addition, there is no separate parameter estimates for each component. It is not convenient to make a direct comparison between two estimation procedures. The empirical results are summarized in Table 15:

**Table 15: Summary of *bayesm* and Nlogit 5.0 estimation with Keane and Wasi (2013)**

| | |
|---|---|
| *MNL* | All procedures report the same estimates. |
| *MIXL (correlated error)* | The mean of the posterior distribution of *bayesm* package is closer to the estimates from Keane and Wasi (2013). In general, there is no substantial difference in the estimates between MSL and HB procedures. |
| *MM-MNL* | There is no direct comparison between both procedures. |

## 5.4 Estimation Comparison

Keane and Wasi (2013) based on their results and found that the MIXL model is never preferred compared to G-MNL and MM-MNL. By using Nlogit 5, the two PIZZA delivery experiments (A and B) indicate that the MIXL model is the best one with respect to the AIC, BIC and CAIC information criteria.

One important fact about the classical procedure is worth mentioning, since only default starting values can be applied in Nlogit 5 (the bug problem), this might be the main reason why the performance of G-MNL and MM-MNL is less satisfactory in this analysis.

If setting the estimates from Keane and Wasi (2013) as benchmark and only comparing MNL and MIXL models, the classical procedure reach the similar conclusion as the Bayesian procedure, which is in accordance with the finding by Huber and Train (2001). In addition, the HB estimates are even closer to the benchmark.

# 6. Simulation results

In this section, for comparing the estimated and prediction accuracy of different models by procedures, two simulations are developed. Following a previous study at the Research Centre ORSTAT (Zhang , 2013), two simulated datasets are discussed in detail. The first one (*simu1*) is simulated from the MIXL model (Zhang , 2013) and second dataset (*simu2)* is generated by a given example code in the *bayesm* package. **Table 16** is the summary of the two simulated datasets.

**Table 16: Simulated dataset setup**

| | Simulated MIXL dataset (*simu1*) | Simulated MM-MNL dataset (*simu2*) |
|---|---|---|
| Number of respondents | 50 | 100 |
| Number of Choice Sets | 20 | 20 |
| Number of Alternatives | 3 | 3 |
| Number of Coefficients | 6 | 3 |

In order to assess the accuracy and precision, three criteria based on the *compare* function (Zhang , 2013) are calculated and defined in Table 17.

**Table 17: Three Precision Criteria in the *compare* function**

* Where $\widehat{\beta_n}$ is the estimated parameter coefficient and $\beta_n^*$ is the true value for each decision maker.

| | |
|---|---|
| **Mean Absolute Error ($\boldsymbol{MAE_\beta}$)** | $MAE_\beta = \frac{1}{N}\sum_{n=1}^{N} \lvert \widehat{\beta_n} - \beta_n^* \rvert$ |
| $MAE_\beta$ is an average of absolute errors measuring how close the estimates are to the true values. | |
| **Root Mean Squared Estimation Error ($\boldsymbol{RMSE_\beta}$)** | $RMSE_\beta = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\widehat{\beta_n} - \beta_n^*)'(\widehat{\beta_n} - \beta_n^*)}$ |
| $RMSE_\beta$ represents the root-mean-square error of estimation. | |
| **Root Mean Squared Prediction Error ($\boldsymbol{RMSE_P}$)** | $RMSE_p = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(p(\widehat{\beta_n}) - p(\beta_n^*))'(p(\widehat{\beta_n}) - p(\beta_n^*))}$ |
| $RMSE_P$ reports the root-mean-square error of prediction. | |

All three criteria are frequently used measures for comparing estimation and prediction accuracy. However, since it is scale-dependent, it is not suitable for comparing between variables (Hyndman & Koehler, 2006).

## 6.1 Some notes on the simulation / estimation procedures

In the MSL procedure, a series of different models are tested, such as different number of classes, different number of draws, correlated / uncorrelated error structures; the reported models are the best ones favoured by BIC.

For each HB estimation procedure, It has been updated at least 60,000 times (keeping every $15^{th}$ draws) and the first half is discarded as burn-in part. The convergence is also examined by the trace plot.

## 6.2 Summary for simulated data from the MIXL model

Table 24 in the appendix is the summary results for the data generated from MIXL model (*simu1*). The first column indicates the true value of each parameter.

In the context of the MSL, the MIXL model with uncorrelated errors has the best fit indicated by BIC, CAIC and all of the three precision criteria. The more complex model G-MNL is favoured by the lowest AIC criteria.

Not able to assess the random parameters in the simulated dataset, the fit of the first two models, MNL and S-MNL are quite poor and there are large discrepancies between the estimates and the true ones. The similar reason can also explain the poor fit of LC model (log-likelihood: $-1041$ vs $-1064$ for MNL).

The variance estimates for both G-MNL and MIXL are very significant, because they capture the heterogeneity across respondents. The estimates of MIXL are closer to the true values compared with G-MNL since *simu1* is simulated from the MIXL model. Although the correlated errors of MIXL achieve lower log-likelihood value ($-983$ compared with -996 for uncorrelated errors) by using 15 extra parameters, MIXL with uncorrelated error is superior based on BIC by 94 points.

For MM-MNL, several models have been estimated by using different number of latent classes. Nlogit 5 again fails to reach a global maximum and the reported estimates are less satisfactory.

The MIXL model with correlated errors is estimated by both procedures and both report similar estimates. All three precision criteria are in favour of the HB procedure and this is in consistent with Allenby & Rossi (1999) that HB is more flexible to assess the heterogeneity at the individual level.

The last column reports the MM-MNL model by HB procedure. Since HB avoids the maximization process that troubles MSL, it produces improved precision measures.

## 6.3 Summary for simulated data from the MM-MNL model

Table 25 in the appendix gives the summarized results for the simulated data from the MM-MNL model. The dataset (*simu2*) is generated from a mixture of 3 normal components and the table on top presents the true values for each component.

By the MSL procedure using Nlogit 5, G-MNL is the best-fit model supported by all model selection criteria and MIXL is superior on the individual-level precision criteria. For both G-MNL and MIXL, the estimates of the mean preference and variance are similar, having the expected sign and magnitude.

Following G-MNL and MIXL, the third best model is LC. It approximately identifies one of the three components by a slightly larger predicted class probability (LC estimates at 0.49 compared with 0.4 of true value). However, for the other two latent classes, LC model only grasp the expected sign and there is discrepancies between parameter estimates compared with true values.

The estimates of MM-MNL by Nlogit 5 again fail to reach the global maximum; however, by utilizing the flexible HB estimation procedure, the estimate with three mixtures of normal components is close to the true ones and it has the best precision among all possible models.

Given the relative small number of respondents in *simu2* (#100), another simulated datasets with 500 respondents is generated from MM-MNL model and estimated by the MM-MNL model using Nlogit 5. The output again suggests that the estimation procedure is unable to converge.

A summary of these results is presented in Table 18. Given that the Bayesian procedure is not computed by maximization of the log-likelihood, it is only compared by the precision criteria.

**Table 18: Summary of simulation results**

| | | **simu 1** (from **MIXL** model) | **simu 2** (from **MM-MNL** model) |
|---|---|---|---|
| Classical (Nlogit 5) | Best by BIC | MIXL (UNCORR) | G-MNL* |
| | Best by precision | MIXL (UNCORR) | MIXL (CORR) |
| Bayesian (bayesm) | Best by precision | MIXL (CORR) | MM-MNL (3 mixture) |
| All Procedures | Best by BIC | MIXL (UNCORR) | G-MNL* |
| | Best by precision | MIXL (UNCORR) | MM-MNL (3 mixture) |

*for G-MNL model, it is estimates from uncorrelated coefficient specification.

## 6.4 Simulation Comparison

Using Nlogit 5, MIXL and G-MNL are the top two preferred models for both simulated datasets. By comparing both MSL and HB procedures, MIXL is the best model to fit dataset *simu1 (which* is generated from MIXL model) either for the model selection criteria or for the precision criteria.  For the second simulated data (which is simulated from MM-MNL model), since MM-MNL model fails to reach global maximum, the best fit is obtained from the G-MNL model, with MIXL as the second best one.

In the HB context, the results are as expected and MIXL and MM-MNL produce the best precision measures for modelling datasets which are generated from MIXL and MM-MNL respectively.

In the classical procedure, many reasons might influence the model performance, such as the starting values, random number generator and the number of draws, especially for complex model, such as G-MNL and MM-MNL. Because the specific setting of original study from Keane and Wasi (2013) is not known and the technical difficulties of testing different starting values in Nlogit 5, these might be the main reasons why the results in this analysis are different.

When comparing the procedures between the MSL and HB, there are three important differences. Firstly, the maximization process for MSL can be very difficult (Train, 2009); once it fails to converge, the right stating values are crucial. In general, a series of starting values have to be tested. On the other hand, the Hierarchical Bayes avoids the maximization process (Rossi, Allenby, & McCulloch, 2006).

Secondly, the comparison of classical and Bayesian procedures in literature shows that under certain conditions, both estimates are asymptotically equivalent (Train, 2001). In other words, the researcher who is not familiar with the Bayesian concept could also use the estimates and interpret them in the classical way (Train, 2009). More details about the parallel interpretation can be found in Geweke (1989).

Thirdly, the Bayesian estimates process is much faster than the classical process. It is in consistent with the findings by Allenby & Rossi (1999) that the classical approach is more computationally demanding for estimating attributes coefficients than the Bayesian procedure. Furthermore, the classical procedure could only produce approximate results; while the Bayesian procedure provides precise information about the posterior distribution (Train, 2010).

# 7. Conclusion

This thesis compares the model performances of six logit models including MNL, MIXL, S-MNL, G-MNL, LC and MM-MNL estimated by maximum simulated likelihood (MSL) and hierarchical Bayesian estimation method (HB).

By applying MSL in Matlab, Keane and Wasi (2013) analysed ten empirical datasets and found that G-MNL and MM-MNL are always superior to the MIXL model. This analysis makes use of two datasets (PIZZA A and PIZZA B) from their study and identifies MIXL as the best model with respect to all model selection criteria, AIC, BIC and CAIC.  One possible reason to explain the different results is because G-MNL and MM-MNL are very complex and the estimation process is sensitive to starting values. The software Nlogit 5 failed to reach global maximum.

In the literature, the MSL and HB procedures asymptotically provide similar model estimates; in this analysis, the estimates for MNL and MIXL models are very close; however, some discrepancies are noticeable in MM-MNL model.

Allenby & Rossi (1999) claim that the HB procedure has relative advantages to assess the parameters of individual consumers; however, in the first simulated dataset (which was generated from MIXL), we find that the lowest precision criteria is achieved by MSL; only in the second simulated dataset (which is simulated from MM-MNL), the Bayesian procedure achieves better accuracy.

The estimation time for the MSL in Nlogit 5 is very long, especially for the complex model and large datasets; for example, the computation time of the MM-MNL model in PIZZA B is 1.5 times that of the MIXL model. Also the technical issue of setting different starting values prevents model converge or reaching the global maximum. However, Nlogit 5 is still powerful software for estimating discrete choice models since it is user-friendly and can cope with the majority of discrete choice models.

Compared with classical procedure, the computation time for the HB estimation is shorter and the *bayesm* package is more flexible as users can specify different numbers of mixture normals. However, the *bayesm* package is only capable of estimating MNL, MIXL and MM-MNL models.

# Appendix A – Tables

Table 19 lists the structured details of two PIZZA Delivery experiments from the study of Keane and Wasi (2013).

**Table 19: Structured details of two PIZZA Delivery experiments (Keane & Wasi, 2013)**

| | No. of choices | No. of choice occasions | No. of respondents | No. of observations | No. of attributes |
|---|---|---|---|---|---|
| PIZZA A | 2 | 16 | 178 | 2848 | 8 |
| PIZZA B | 2 | 32 | 328 | 10496 | 16 |

PIZZA A: attributes 1-8; PIZZA B: attributes 1_16

| Attributes | Levels |
|---|---|
| 1. Gourmet | -1 (Traditional),1(Gourmet) |
| 2. Price | -1 ($13),1 ($17) |
| 3. Ingredient freshness | -1 (some canned),1(all fresh ingredients) |
| 4. Delivery time | -1(45 mins),1(30 min) |
| 5. Crust | -1(thin),1(thick) |
| 6. Sizes | -1(single size),1(3 sizes) |
| 7. Steaming hot | -1(warm),1(steaming hot) |
| 8. Late open hours | -1(till 10 pm.), 1 (till 1 am.) |
| 9. Free delivery charge | -1($2),1 (free) |
| 10. Local store | -1(chain),1(local) |
| 11. Baking methods | -1(traditional),1(wood fire) |
| 12. Manners | -1(friendly),1(polite & friendly) |
| 13. Vegetarian available | -1(no),1(yes) |
| 14. Delivery time Guaranteed | -1(no),1(yes) |
| 15. Distance to the outlet | -1(in other suburb),1(in own suburb) |
| 16. Variety availability | -1(restricted menu),1(large menu) |

## Table 20: Empirical analysis results for PIZZA A (by Nlogit 5)

**Analysis for the fourth data set - PIZZA A**

* the original results by Keane & Wasi (2013) is highlighted in colour. (Bold estimates are statisticcally significant at **5%**)

| | MNL | | S-MNL | | | | G-MNL (UNCORR) | | | | MIXL (CORR) | | | | MIXL (UNCORR) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EST | SE | EST | SE | Nlogit | SE | EST | SE | Nlogit | SE | EST | SE | EST(N) | SE | EST | SE | Nlogit | SE |
| gourmet | 0.02 | 0.02 | 0.03 | 0.04 | 0.04 | 0.05 | **0.45** | 0.22 | **0.28** | 0.13 | 0.01 | 0.05 | -0.02 | 0.05 | 0.03 | 0.05 | 0.03 | 0.06 |
| price | **-0.16** | 0.02 | **-0.19** | 0.05 | **-0.23** | 0.08 | **-1.67** | 0.65 | **-1.33** | 0.21 | **-0.38** | 0.06 | **-0.38** | 0.06 | **-0.35** | 0.06 | **-0.34** | 0.06 |
| ingredient | **0.48** | 0.03 | **1.45** | 0.29 | **1.74** | 0.70 | **4.65** | 1.69 | **3.06** | 0.33 | **1.06** | 0.08 | **1.39** | 0.13 | **0.96** | 0.08 | **1.13** | 0.11 |
| delivery | **0.09** | 0.03 | **0.16** | 0.08 | **0.20** | 0.09 | **0.74** | 0.35 | **1.02** | 0.16 | **0.17** | 0.05 | **0.19** | 0.07 | **0.16** | 0.05 | **0.17** | 0.05 |
| crust | 0.02 | 0.03 | 0.01 | 0.04 | 0.01 | 0.06 | 0.42 | 0.26 | -0.25 | 0.20 | 0.08 | 0.06 | 0.02 | 0.10 | 0.02 | 0.06 | 0.10 | 0.07 |
| size | **0.09** | 0.03 | **0.12** | 0.06 | **0.15** | 0.08 | **0.81** | 0.37 | **0.82** | 0.16 | **0.17** | 0.05 | **0.24** | 0.07 | **0.20** | 0.05 | **0.18** | 0.06 |
| steaming | **0.38** | 0.03 | **1.02** | 0.24 | **1.25** | 0.63 | **4.46** | 1.64 | **3.03** | 0.30 | **0.86** | 0.08 | **0.93** | 0.13 | **0.87** | 0.08 | **0.86** | 0.08 |
| late_hours | **0.04** | 0.02 | 0.08 | 0.06 | 0.09 | 0.06 | 0.29 | 0.17 | 0.17 | 0.12 | 0.07 | 0.05 | 0.03 | 0.06 | 0.07 | 0.05 | 0.07 | 0.05 |
| Tau | N.A. | | 1.69 | 0.18 | 0 | _ | **1.79** | 0.24 | 1.8 | 5.44 | N.A. | | | | N.A. | | | |
| Gamma | | | N.A. | | N.A. | | 0.01 | 0.01 | 0.048 | 0.14 | | | | | | | | |
| Class Probability | | | | | | | N.A. | | N.A. | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| No. of Parameters | 8 | | 9 | | 9 | | 18 | | 18 | | 44 | | 44 | | 16 | | 16 | |
| LL | -1657 | | -1581 | | -1579 | | -1373 | | -1377 | | -1379 | | -1378 | | -1403 | | -1382 | |
| AIC | 3330 | | 3179 | | 3176 | | 2782 | | 2790 | | 2847 | | 2844 | | 2838 | | **2795** | |
| BIC | 3378 | | 3234 | | 3230 | | **2889** | | 2897 | | 3108 | | 3106 | | 2933 | | **2891** | |
| CAIC | 3386 | | 3243 | | 3239 | | **2907** | | 2915 | | 3152 | | | | 2949 | | **2907** | |

| | Latent Class (estimates with 4 classes) | | | | | | | | | | | | MM-MNL (estimate with 2 independent normals) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cla 1 | SE | Cla 1 | SE | Cla 2 | SE | Cla 2 | SE | Cla 3 | SE | Cla 3 | SE | Cla 1 | SE | Cla 1 | SE | Cla 2 | SE | Cla 2 | SE |
| gourmet | -0.01 | 0.05 | -0.02 | 0.06 | 0.02 | 0.02 | 0.02 | 0.04 | 0.08 | 0.10 | 0.07 | 0.08 | 0.02 | 0.07 | 0.04 | 0.03 | 0.14 | 0.47 | -0.04 | 0.07 |
| price | **-0.20** | 0.06 | **-0.20** | 0.06 | **-0.16** | 0.03 | **-0.16** | 0.04 | **-0.39** | 0.09 | **-0.38** | 0.10 | **-0.18** | 0.06 | **-0.18** | 0.03 | -4.63 | 2.71 | **-0.20** | 0.06 |
| ingredient | **1.57** | 0.09 | **1.57** | 0.10 | **0.12** | 0.06 | **0.13** | 0.05 | 0.30 | 0.16 | **0.31** | 0.09 | **0.59** | 0.08 | **0.16** | 0.04 | 13.47 | 7.73 | **1.63** | 0.13 |
| delivery | 0.10 | 0.09 | 0.10 | 0.09 | **0.10** | 0.04 | **0.10** | 0.04 | **0.32** | 0.09 | **0.30** | 0.10 | 0.06 | 0.05 | **0.10** | 0.03 | 3.95 | 2.36 | 0.12 | 0.11 |
| crust | **-0.12** | 0.06 | -0.11 | 0.08 | 0.01 | 0.05 | 0.01 | 0.05 | **-0.30** | 0.09 | **-0.27** | 0.10 | -0.06 | 0.08 | 0.03 | 0.03 | 1.18 | 1.05 | -0.03 | 0.08 |
| size | **0.15** | 0.07 | **0.15** | 0.07 | 0.06 | 0.04 | 0.06 | 0.05 | **0.23** | 0.11 | **0.23** | 0.08 | **0.23** | 0.07 | **0.09** | 0.03 | 0.92 | 0.81 | 0.15 | 0.07 |
| steaming | **0.50** | 0.08 | **0.50** | 0.10 | **0.12** | 0.06 | **0.12** | 0.06 | **1.60** | 0.18 | **1.61** | 0.19 | **0.50** | 0.08 | **0.44** | 0.03 | 9.85 | 5.76 | **0.47** | 0.10 |
| late_hours | 0.09 | 0.08 | 0.09 | 0.07 | **0.06** | 0.03 | 0.06 | 0.04 | 0.02 | 0.07 | 0.02 | 0.08 | **0.12** | 0.06 | 0.04 | 0.03 | -0.97 | 0.72 | **0.08** | 0.08 |
| Tau | N.A. | | | | | | | | | | | | N.A. | | | | | | | |
| Gamma | | | | | | | | | | | | | | | | | | | | |
| Class Probability | 0.36 | 0.04 | 0.36 | 0.04 | 0.32 | 0.04 | 0.32 | 0.05 | 0.23 | 0.04 | 0.23 | 0.04 | 0.57 | 0.04 | 0.65 | 0.04 | 0.43 | 0.04 | 0.35 | 0.04 |
| | | | | | | | | | | | | | | | | | | | | |
| No. of Parameters | 35 | | 35 | | | | | | | | | | 33 | | 33 | | | | | |
| LL | -1418 | | -1418 | | | | | | | | | | -1328 | | -1526 | | | | | |
| AIC | 2907 | | 2907 | | | | | | | | | | **2722** | | 3118 | | | | | |
| BIC | 3114 | | 3114 | | | | | | | | | | 2918 | | 3314 | | | | | |
| CAIC | 3149 | | 3149 | | | | | | | | | | 2951 | | 3347 | | | | | |

## Table 21: Empirical analysis results for PIZZA B (by Nlogit 5)

**Analysis for the seventh data set - PIZZA B**

* the original results by Keane & Wasi (2013) is highlighted in colour (Bold estimates are statisticcally significant at **1%**)

| | MNL | | S-MNL | | | | G-MNL (UNCORR) | | | | MIXL (CORR) | | | | MIXL (UNCORR) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EST | SE | EST | SET | Nlogit | SE | EST | SE | Nlogit | SE | EST | SE | EST (TH) | SE | EST | SE | Nlogit | SE |
| gourmet | 0.01 | 0.01 | **0.05** | 0.01 | **0.05** | 0.02 | 0.03 | 0.03 | **0.12** | 0.05 | 0.01 | 0.02 | 0.02 | 0.02 | 0.01 | 0.02 | 0.01 | 0.02 |
| price | **-0.17** | 0.01 | **-0.25** | 0.02 | **-0.28** | 0.04 | **-0.79** | 0.07 | **-0.99** | 0.13 | -0.32 | 0.03 | -0.21 | 0.02 | **-0.3** | 0.03 | **-0.36** | 0.04 |
| ingredient | **0.21** | 0.01 | **0.36** | 0.03 | **0.40** | 0.04 | **1.05** | 0.08 | **1.11** | 0.13 | 0.39 | 0.03 | 0.40 | 0.02 | **0.34** | 0.03 | **0.28** | 0.03 |
| delivery | **0.03** | 0.01 | 0.04 | 0.02 | 0.04 | 0.02 | **0.15** | 0.04 | **0.22** | 0.07 | 0.05 | 0.02 | 0.06 | 0.02 | 0.05 | 0.02 | 0.04 | 0.02 |
| crust | **0.08** | 0.01 | **0.09** | 0.01 | **0.10** | 0.02 | **0.59** | 0.06 | **0.23** | 0.06 | 0.16 | 0.03 | 0.11 | 0.03 | **0.08** | 0.03 | **0.16** | 0.03 |
| size | **0.07** | 0.01 | **0.08** | 0.02 | **0.09** | 0.02 | **0.23** | 0.03 | **0.28** | 0.05 | 0.1 | 0.02 | 0.13 | 0.02 | **0.11** | 0.02 | **0.09** | 0.02 |
| steaming | **0.20** | 0.01 | **0.35** | 0.03 | **0.38** | 0.04 | **1.15** | 0.09 | **1.23** | 0.14 | 0.34 | 0.03 | 0.36 | 0.02 | **0.34** | 0.02 | **0.36** | 0.03 |
| late_hours | **0.04** | 0.01 | 0.02 | 0.02 | 0.02 | 0.01 | 0.08 | 0.04 | 0.09 | 0.05 | 0.07 | 0.02 | 0.07 | 0.02 | **0.08** | 0.02 | **0.07** | 0.02 |
| free_delivery | **0.12** | 0.01 | **0.15** | 0.02 | **0.16** | 0.02 | **0.56** | 0.06 | **0.78** | 0.09 | 0.21 | 0.02 | 0.20 | 0.02 | **0.20** | 0.02 | **0.20** | 0.02 |
| local | **0.08** | 0.01 | **0.06** | 0.02 | **0.07** | 0.02 | **0.42** | 0.05 | **0.68** | 0.08 | 0.13 | 0.02 | 0.14 | 0.02 | **0.15** | 0.02 | **0.13** | 0.02 |
| baking | **0.07** | 0.01 | **0.07** | 0.02 | **0.07** | 0.02 | **0.25** | 0.04 | **0.38** | 0.06 | 0.1 | 0.02 | 0.10 | 0.02 | **0.11** | 0.02 | **0.09** | 0.02 |
| manners | 0.01 | 0.01 | -0.004 | 0.02 | 0.003 | 0.02 | 0.01 | 0.04 | -0.09 | 0.05 | 0.02 | 0.02 | 0.04 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| vegetarian | **0.09** | 0.01 | **0.06** | 0.01 | **0.07** | 0.02 | **0.34** | 0.06 | **0.56** | 0.09 | 0.11 | 0.03 | 0.14 | 0.02 | **0.13** | 0.03 | **0.15** | 0.02 |
| time_guarant | **0.07** | 0.01 | **0.07** | 0.02 | **0.07** | 0.02 | **0.15** | 0.04 | **0.21** | 0.05 | 0.11 | 0.02 | 0.11 | 0.02 | **0.11** | 0.02 | **0.12** | 0.02 |
| distance | **0.06** | 0.01 | 0.04 | 0.02 | 0.04 | 0.02 | 0.1 | 0.04 | **0.17** | 0.05 | 0.09 | 0.02 | 0.09 | 0.02 | **0.09** | 0.02 | **0.09** | 0.02 |
| variety | **0.06** | 0.02 | 0.04 | 0.02 | **0.05** | 0.02 | **0.14** | 0.05 | 0.12 | 0.07 | 0.09 | 0.02 | 0.08 | 0.02 | **0.09** | 0.02 | **0.09** | 0.02 |
| **Tau** | N.A. | | 1.22 | 0.08 | 0 | _ | 1.26 | 0.06 | 0.001 | 6 | N.A. | | | | N.A. | | | |
| Gamma | | | N.A. | | | | 0.01 | 0.01 | – | – | | | | | | | | |
| Class Probability | | | | | | | N.A. | | N.A. | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | |
| No. of Parameters | 16 | | 17 | | 17 | | 34 | | 34 | | 48 | | 152 | | 32 | | 32 | |
| LL | -6747 | | -6607 | | -6606 | | -5689 | | -5903 | | -5857 | | -5867 | | -5892 | | -5872 | |
| AIC | 13525 | | 13249 | | 13246 | | 11446 | | 11875 | | 11810 | | 12038 | | 11849 | | **11807** | |
| BIC | 13642 | | 13371 | | 13369 | | 11693 | | 12121 | | 12158 | | 13141 | | 12080 | | **12040** | |
| CAIC | 13658 | | 13388 | | 13386 | | 11727 | | 12155 | | 12206 | | 13293 | | 12112 | | **12072** | |

| | Latent class (estimate with 6 classes) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Class **1** | SE | Cla 1 | SE | Class **2** | SE | Cla 2 | SE | Class **3** | SE | Cla 3 | SE |
| gourmet | 0.01 | 0.02 | 0.02 | 0.02 | 0.02 | 0.07 | 0.01 | 0.08 | 0.09 | 0.05 | 0.22 | 0.09 |
| price | -0.04 | 0.02 | -0.05 | 0.02 | **-1.71** | 0.09 | **-1.79** | 0.16 | 0.24 | 0.11 | **0.30** | 0.10 |
| ingredient | **0.1** | 0.02 | **0.11** | 0.02 | **0.46** | 0.06 | **0.51** | 0.10 | **2.17** | 0.19 | **2.19** | 0.15 |
| delivery | 0.02 | 0.02 | 0.02 | 0.02 | 0.14 | 0.1 | 0.13 | 0.07 | -0.03 | 0.16 | -0.08 | 0.12 |
| crust | **-0.04** | 0.01 | **-0.04** | 0.02 | -0.05 | 0.04 | -0.03 | 0.06 | **0.31** | 0.08 | **0.34** | 0.10 |
| size | **0.05** | 0.02 | **0.05** | 0.02 | **0.19** | 0.07 | **0.19** | 0.06 | **0.28** | 0.07 | **0.31** | 0.07 |
| steaming | **0.1** | 0.02 | **0.11** | 0.02 | **0.22** | 0.07 | **0.26** | 0.07 | **0.67** | 0.07 | **0.69** | 0.07 |
| late_hours | **0.04** | 0.01 | **0.05** | 0.02 | 0.06 | 0.06 | 0.05 | 0.05 | 0.07 | 0.1 | 0.09 | 0.07 |
| free_delivery | **0.11** | 0.01 | **0.11** | 0.02 | **0.56** | 0.06 | **0.60** | 0.07 | 0.15 | 0.08 | 0.15 | 0.07 |
| local | **0.14** | 0.01 | **0.14** | 0.02 | -0.01 | 0.07 | -0.01 | 0.06 | 0.1 | 0.12 | 0.12 | 0.07 |
| baking | **0.06** | 0.01 | **0.06** | 0.02 | 0.16 | 0.07 | **0.16** | 0.05 | **0.29** | 0.07 | **0.29** | 0.06 |
| manners | 0.03 | 0.02 | 0.03 | 0.02 | 0.03 | 0.08 | 0.04 | 0.06 | -0.06 | 0.11 | -0.04 | 0.07 |
| vegetarian | 0.02 | 0.02 | 0.02 | 0.02 | **0.15** | 0.04 | **0.16** | 0.05 | 0.04 | 0.11 | 0.01 | 0.07 |
| time_guarant | **0.08** | 0.02 | **0.08** | 0.02 | **0.17** | 0.05 | **0.16** | 0.06 | 0.12 | 0.12 | 0.12 | 0.05 |
| distance | **0.09** | 0.02 | **0.09** | 0.02 | 0.11 | 0.07 | 0.13 | 0.05 | -0.12 | 0.1 | -0.13 | 0.06 |
| variety | **0.07** | 0.03 | **0.07** | 0.02 | 0.03 | 0.07 | 0.04 | 0.05 | 0.07 | 0.1 | 0.07 | 0.06 |
| **Tau** | N.A. | | | | | | | | | | | |
| Gamma | | | | | | | | | | | | |
| Class Probability | 0.51 | 0.03 | 0..52 | 0.03 | 0.14 | 0.02 | 0.13 | 0.02 | 0.12 | 0.02 | 0.12 | 0.02 |
| | | | | | | | | | | | | |
| No. of Parameters | 101 | | 107 | | | | | | | | | |
| LL | -5591 | | -5568 | | | | | | | | | |
| AIC | 11385 | | 11350 | | | | | | | | | |
| BIC | 12117 | | 12127 | | | | | | | | | |
| CAIC | 12218 | | 12234 | | | | | | | | | |

| | MM-MNL(estimate with 3 independent normals) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Class **1** | SE | Cla 1 | SE | Class **2** | SE | Cla 2 | SE | Class **3** | SE | Cla 3 | SE |
| gourmet | -0.03 | 0.04 | 0.002 | 0.02 | -0.12 | 0.07 | -0.02 | 0.05 | **0.37** | 0.08 | 0.12 | 0.09 |
| price | -0.1 | 0.04 | **-0.19** | 0.02 | **-0.86** | 0.1 | **-0.15** | 0.05 | -0.17 | 0.13 | **-0.46** | 0.14 |
| ingredient | **0.12** | 0.03 | **0.10** | 0.02 | **0.29** | 0.07 | **0.80** | 0.07 | **1.02** | 0.13 | **1.20** | 0.17 |
| delivery | 0.02 | 0.03 | 0.03 | 0.02 | 0.19 | 0.07 | 0.07 | 0.04 | 0.14 | 0.08 | 0.04 | 0.13 |
| crust | -0.03 | 0.03 | -0.04 | 0.02 | **0.62** | 0.09 | **0.76** | 0.06 | 0.15 | 0.07 | -0.04 | 0.07 |
| size | 0.06 | 0.03 | **0.06** | 0.02 | **0.31** | 0.07 | **0.16** | 0.04 | **0.26** | 0.09 | 0.17 | 0.09 |
| steaming | **0.11** | 0.03 | **0.12** | 0.01 | **0.37** | 0.06 | **0.19** | 0.04 | **1.43** | 0.17 | **2.11** | 0.20 |
| late_hours | 0.01 | 0.02 | **0.04** | 0.01 | **0.29** | 0.07 | 0.02 | 0.04 | **0.19** | 0.06 | 0.20 | 0.09 |
| free_delivery | **0.22** | 0.05 | **0.15** | 0.01 | **0.26** | 0.06 | 0.08 | 0.04 | **0.28** | 0.07 | 0.10 | 0.08 |
| local | **0.09** | 0.03 | **0.11** | 0.01 | **0.43** | 0.07 | 0.01 | 0.04 | 0.08 | 0.08 | 0.12 | 0.12 |
| baking | 0.01 | 0.03 | **0.06** | 0.02 | **0.32** | 0.06 | **0.11** | 0.04 | **0.35** | 0.11 | 0.19 | 0.08 |
| manners | 0.03 | 0.03 | 0.02 | 0.02 | -0.06 | 0.08 | -0.01 | 0.04 | 0.11 | 0.11 | 0.10 | 0.13 |
| vegetarian | 0.04 | 0.03 | **0.12** | 0.02 | **0.35** | 0.09 | 0.01 | 0.04 | 0.04 | 0.07 | 0.06 | 0.09 |
| time_guarant | **0.14** | 0.04 | **0.08** | 0.02 | 0.07 | 0.08 | 0.08 | 0.04 | **0.19** | 0.07 | 0.02 | 0.08 |
| distance | **0.11** | 0.04 | **0.08** | 0.02 | 0.09 | 0.07 | -0.03 | 0.04 | 0.06 | 0.07 | 0.16 | 0.09 |
| variety | **0.1** | 0.03 | **0.06** | 0.02 | 0.03 | 0.07 | 0.09 | 0.04 | 0.19 | 0.08 | 0.13 | 0.09 |
| **Tau** | N.A. | | | | | | | | | | | |
| Gamma | | | | | | | | | | | | |
| Class Probability | 0.41 | 0.03 | 0.7 | 0.03 | 0.31 | 0.03 | 0.2 | 0.03 | 0.28 | 0.03 | 0.1 | 0.02 |
| | | | | | | | | | | | | |
| No. of Parameters | 98 | | 98 | | | | | | | | | |
| LL | -5310 | | -6279 | | | | | | | | | |
| AIC | **10815** | | 12754 | | | | | | | | | |
| BIC | **11527** | | 13465 | | | | | | | | | |
| CAIC | **11625** | | 13563 | | | | | | | | | |

**Table 22: Empirical analysis results for PIZZA A (by HB procedure)**

| Analysis for the fourth data set - PIZZA A | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * the original results by Keane & Wasi (2013) is highlighted in colour. | | | | | | | | | | | | | | | | | | |
| | MNL | | MIXL (CORR) | | | | Bayesm | | MM-MNL | | | | | | | | BAYESM - 2 MIXTURE | |
| | EST | SE | EST | SE | Nlogit | SE | EST | SE | Cla **1** | SE | Nlogit | SE | Cla **2** | SE | Cla 2 | SE | ALL EST | SE |
| gourmet | 0.02 | 0.02 | 0.01 | 0.05 | -0.02 | 0.05 | 0.03 | 0.06 | 0.02 | 0.07 | 0.04 | 0.03 | 0.14 | 0.47 | -0.04 | 0.07 | -0.02 | 0.09 |
| price | **-0.16** | 0.02 | **-0.38** | 0.06 | **-0.38** | 0.06 | -0.38 | 0.06 | -0.18 | 0.06 | -0.18 | 0.03 | -4.63 | 2.71 | -0.20 | 0.06 | -0.46 | 0.08 |
| ingredient | **0.48** | 0.03 | **1.06** | 0.08 | **1.39** | 0.13 | 1.07 | 0.12 | 0.59 | 0.08 | 0.16 | 0.04 | 13.47 | 7.73 | 1.63 | 0.13 | 1.83 | 0.28 |
| delivery | **0.09** | 0.03 | **0.17** | 0.05 | **0.19** | 0.07 | 0.17 | 0.06 | 0.06 | 0.05 | 0.10 | 0.03 | 3.95 | 2.36 | 0.12 | 0.11 | 0.29 | 0.14 |
| crust | 0.02 | 0.03 | 0.08 | 0.06 | 0.02 | 0.10 | 0.05 | 0.08 | -0.06 | 0.08 | 0.03 | 0.03 | 1.18 | 1.05 | -0.03 | 0.08 | 0.06 | 0.11 |
| size | **0.09** | 0.03 | **0.17** | 0.05 | **0.24** | 0.07 | 0.18 | 0.06 | 0.23 | 0.07 | 0.09 | 0.03 | 0.92 | 0.81 | 0.15 | 0.07 | 0.21 | 0.08 |
| steaming | **0.38** | 0.03 | **0.86** | 0.08 | **0.93** | 0.13 | 0.84 | 0.11 | 0.5 | 0.08 | 0.44 | 0.03 | 9.85 | 5.76 | 0.47 | 0.10 | 1.08 | 0.16 |
| late_hours | **0.04** | 0.02 | 0.07 | 0.05 | 0.03 | 0.06 | 0.08 | 0.05 | 0.12 | 0.06 | 0.04 | 0.03 | -0.97 | 0.72 | 0.08 | 0.08 | 0.02 | 0.07 |
| Tau | N.A. | | N.A. | | | | N.A. | | N.A. | | | | | | | | N.A. | |
| Gamma | | | | | | | | | | | | | | | | | | |
| Class Probability | | | | | | | | | 0.57 | 0.04 | 0.65 | 0.04 | 0.43 | 0.04 | 0.35 | 0.04 | N.A. | |
| No. of Parameters | 8 | | | 44 | 44 | | | | | 33 | 33 | | | | | | | |
| LL | -1657 | | | -1379 | -1378 | | | | | **-1328** | -1526 | | | | | | | |
| AIC | | | | 2847 | 2844 | | | | | **2722** | 3118 | | | | | | | |

**Table 23: Empirical analysis results for PIZZA B (by HB procedure)**

| Analysis for the SEVENTH data set - PIZZA B | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * the original results by Keane & Wasi (2013) is highlighted in colour. | | | | | | | | | | | | | | | | | | | | | | |
| | MNL | | MIXL (CORR) | | | | Bayesm | | MM-MNL | | | | | | | | | | | | bayesm - 3 mixture | |
| | EST | SE | EST | SE | Nlogit | SE | EST | SE | Cla **1** | SE | Nlogit1 | SE | Cla **2** | SE | Nlogit2 | SE | Class **3** | SE | Nlogit3 | SE | ALL EST | SE |
| gourmet | 0.01 | 0.01 | 0.01 | 0.02 | 0.02 | 0.02 | 0.04 | 0.03 | -0 | 0.04 | 0.00 | 0.02 | -0.1 | 0.07 | -0.02 | 0.05 | 0.37 | 0.08 | 0.12 | 0.09 | 0.04 | 0.05 |
| price | -0.17 | 0.01 | -0.3 | 0.03 | -0.21 | 0.02 | -0.39 | 0.05 | -0.1 | 0.04 | -0.19 | 0.02 | -0.9 | 0.1 | -0.15 | 0.05 | -0.17 | 0.13 | -0.46 | 0.14 | -0.69 | 0.09 |
| ingredient | 0.21 | 0.01 | 0.39 | 0.03 | 0.40 | 0.02 | 0.47 | 0.05 | 0.12 | 0.03 | 0.10 | 0.02 | 0.29 | 0.07 | 0.80 | 0.07 | 1.02 | 0.13 | 1.20 | 0.17 | 0.81 | 0.09 |
| delivery | 0.03 | 0.01 | 0.05 | 0.02 | 0.06 | 0.02 | 0.08 | 0.03 | 0.02 | 0.03 | 0.03 | 0.02 | 0.19 | 0.07 | 0.07 | 0.04 | 0.14 | 0.08 | 0.04 | 0.13 | 0.15 | 0.05 |
| crust | 0.08 | 0.01 | 0.16 | 0.03 | 0.11 | 0.03 | 0.15 | 0.04 | -0 | 0.03 | -0.04 | 0.02 | 0.62 | 0.09 | 0.76 | 0.06 | 0.15 | 0.07 | -0.04 | 0.07 | 0.25 | 0.08 |
| size | 0.07 | 0.01 | 0.1 | 0.02 | 0.13 | 0.02 | 0.16 | 0.03 | 0.06 | 0.03 | 0.06 | 0.02 | 0.31 | 0.07 | 0.16 | 0.04 | 0.26 | 0.09 | 0.17 | 0.09 | 0.24 | 0.05 |
| steaming | 0.20 | 0.01 | 0.34 | 0.03 | 0.36 | 0.02 | 0.45 | 0.04 | 0.11 | 0.03 | 0.12 | 0.01 | 0.37 | 0.06 | 0.19 | 0.04 | 1.43 | 0.17 | 2.11 | 0.20 | 0.69 | 0.08 |
| late_hours | 0.04 | 0.01 | 0.07 | 0.02 | 0.07 | 0.02 | 0.09 | 0.02 | 0.01 | 0.02 | 0.04 | 0.01 | 0.29 | 0.07 | 0.02 | 0.04 | 0.19 | 0.06 | 0.20 | 0.09 | 0.13 | 0.04 |
| free_delivery | 0.12 | 0.01 | 0.21 | 0.02 | 0.20 | 0.02 | 0.26 | 0.03 | 0.22 | 0.05 | 0.15 | 0.01 | 0.26 | 0.06 | 0.08 | 0.04 | 0.28 | 0.07 | 0.10 | 0.08 | 0.40 | 0.05 |
| local | 0.08 | 0.01 | 0.13 | 0.02 | 0.14 | 0.02 | 0.17 | 0.03 | 0.09 | 0.03 | 0.11 | 0.01 | 0.43 | 0.07 | 0.01 | 0.04 | 0.08 | 0.08 | 0.12 | 0.12 | 0.18 | 0.05 |
| baking | 0.07 | 0.01 | 0.1 | 0.02 | 0.10 | 0.02 | 0.14 | 0.03 | 0.01 | 0.03 | 0.06 | 0.02 | 0.32 | 0.06 | 0.11 | 0.04 | 0.35 | 0.11 | 0.19 | 0.08 | 0.21 | 0.04 |
| manners | 0.01 | 0.01 | 0.02 | 0.02 | 0.04 | 0.02 | 0.02 | 0.03 | 0.03 | 0.03 | 0.02 | 0.02 | -0.1 | 0.08 | -0.01 | 0.04 | 0.11 | 0.11 | 0.10 | 0.13 | 0.04 | 0.04 |
| vegetarian | 0.09 | 0.01 | 0.11 | 0.03 | 0.14 | 0.02 | 0.15 | 0.03 | 0.04 | 0.03 | 0.12 | 0.02 | 0.35 | 0.09 | 0.01 | 0.04 | 0.04 | 0.07 | 0.06 | 0.09 | 0.17 | 0.04 |
| time_guarant | 0.07 | 0.01 | 0.11 | 0.02 | 0.11 | 0.02 | 0.14 | 0.03 | 0.14 | 0.04 | 0.08 | 0.02 | 0.07 | 0.08 | 0.08 | 0.04 | 0.19 | 0.07 | 0.02 | 0.08 | 0.15 | 0.03 |
| distance | 0.06 | 0.01 | 0.09 | 0.02 | 0.09 | 0.02 | 0.10 | 0.03 | 0.11 | 0.04 | 0.08 | 0.02 | 0.09 | 0.07 | -0.03 | 0.04 | 0.06 | 0.07 | 0.16 | 0.09 | 0.13 | 0.04 |
| variety | 0.06 | 0.02 | 0.09 | 0.02 | 0.08 | 0.02 | 0.11 | 0.03 | 0.1 | 0.03 | 0.06 | 0.02 | 0.03 | 0.07 | 0.09 | 0.04 | 0.19 | 0.08 | 0.13 | 0.09 | 0.14 | 0.04 |
| Tau | N.A. | | N.A. | | | | | | N.A. | | | | | | | | | | | | | |
| Gamma | | | | | | | | | | | | | | | | | | | | | | |
| Class Probability | | | | | | | | | 0.41 | 0.03 | 0.7 | 0.03 | 0.31 | 0.03 | 0.2 | 0.03 | 0.28 | 0.03 | 0.1 | 0.02 | | |
| No. of Parameters | 16 | | | 48 | 152 | | | | | 98 | 98 | | | | | | | | | | | |
| LL | -6747 | | | -5857 | -5867 | | | | | -5310 | -6279 | | | | | | | | | | | |
| AIC | 13525 | | | 11810 | 12038 | | | | | **10815** | 12754 | | | | | | | | | | | |

**Table 24: Simulation results for data from MIXL model**

| Analysis for Simulated data from Mixed logit | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| * Bold estimates are statisticcally significant at 1%, | | | | | | | | | | | | | | | | | | | | | | | | |
| | TRUE | Nlogit - MNL | | Nlogit S-MNL-500 | | Nlogit G-MNL -500 | | Nlogit MIXL-UNCORR-1000 | | Nlogit MIXL-CORR-1000 | | Bayesm 1 Mixture | | Nlogit- CL-2CLS-500 | | | | Nlogit- MM-MNL - 2cls-200 | | | | Bayesm 2 Mixture | |
| | EST | Nlogit | SE | Nlogit | SE | Nlogit | SE | Nlogit | SE | Nlogit | SE | EST | SE | Cla 1 | SE | Cla 2 | SE | Cla 1 | SE | Cla 2 | SE | ALL EST | SE |
| v1 | -0.50 | **-0.36** | 0.06 | **-0.40** | 0.10 | **-0.62** | 0.20 | **-0.69** | 0.11 | **-0.74** | 0.12 | **-0.62** | 0.11 | **-0.52** | 0.09 | **-0.26** | 0.12 | **-0.52** | 0.09 | **-0.2645** | 0.12 | **-0.70** | 0.14 |
| v2 | 0.00 | -0.02 | 0.05 | **-0.12** | 0.05 | -0.06 | 0.10 | 0.02 | 0.09 | -0.02 | 0.10 | 0.02 | 0.09 | -0.12 | 0.09 | 0.11 | 0.09 | -0.12 | 0.09 | 0.1075 | 0.09 | 0.00 | 0.12 |
| v3 | -0.50 | 0.04 | 0.05 | **0.13** | 0.06 | -0.22 | 0.17 | -0.02 | 0.12 | -0.12 | 0.11 | -0.09 | 0.12 | **0.23** | 0.09 | -0.19 | 0.12 | **0.23** | 0.09 | **-0.1901** | 0.11 | -0.09 | 0.16 |
| v4 | 0.00 | -0.08 | 0.05 | 0.03 | 0.06 | -0.03 | 0.13 | -0.03 | 0.11 | -0.06 | 0.11 | -0.06 | 0.10 | 0.01 | 0.09 | **-0.18** | 0.10 | 0.01 | 0.09 | **-0.1815** | 0.10 | -0.08 | 0.14 |
| v5 | -0.50 | **-0.13** | 0.05 | **-0.19** | 0.07 | **-0.63** | 0.17 | **-0.45** | 0.11 | **-0.51** | 0.14 | **-0.45** | 0.11 | **-0.30** | 0.11 | 0.00 | 0.10 | **-0.31** | 0.10 | -0.0037 | 0.09 | **-0.50** | 0.15 |
| v6 | 0.00 | 0.02 | 0.06 | 0.07 | 0.06 | **0.39** | 0.11 | 0.08 | 0.12 | 0.16 | 0.13 | 0.07 | 0.13 | -0.19 | 0.13 | **0.27** | 0.11 | -0.19 | 0.12 | **0.2653** | 0.11 | 0.08 | 0.17 |
| sigma(v1) | 0.50 | N.A. | | N.A. | | **0.37** | 0.13 | **0.59** | 0.10 | **0.58** | 0.10 | 0.55 | | N.A. | | | | 0.0011 | 0.07 | 0.0013 | 0.07 | 0.72 | |
| sigma(v2) | 0.50 | | | | | **0.53** | 0.16 | **0.45** | 0.11 | **0.51** | 0.09 | 0.48 | | | | | | 0.0001 | 0.06 | 0.0007 | 0.07 | 0.62 | |
| sigma(v3) | 0.50 | | | | | **0.81** | 0.19 | **0.83** | 0.13 | **0.61** | 0.11 | 0.64 | | | | | | 0.0001 | 0.06 | 0.0005 | 0.07 | 0.81 | |
| sigma(v4) | 0.50 | | | | | **0.64** | 0.18 | **0.64** | 0.10 | **0.61** | 0.11 | 0.57 | | | | | | 0.0002 | 0.06 | 0.0002 | 0.07 | 0.73 | |
| sigma(v5) | 0.50 | | | | | **0.71** | 0.23 | **0.66** | 0.12 | **0.60** | 0.14 | 0.59 | | | | | | 0.0009 | 0.06 | 0.0005 | 0.07 | 0.76 | |
| sigma(v6) | 0.50 | | | | | **0.72** | 0.19 | **0.73** | 0.12 | **0.90** | 0.14 | 0.72 | | | | | | 0.0028 | 0.07 | 0.0038 | 0.07 | 0.86 | |
| Tau | N.A. | N.A. | | 0 | - | 0.32 | 2.99 | N.A. | | N.A. | | N.A. | | N.A. | | | | | | | | N.A. | |
| Gamma | | | | | | 0.29 | 2.72 | | | | | | | | | | | | | | | | |
| Class Probability | | | | | | N.A. | | | | | | | | 0.56 | 0.13 | 0.44 | 0.13 | 0.55 | 0.12 | 0.45 | 0.12 | | |
| MAE | N.A. | N.A. | | 0.56 | | 0.39 | | **0.31** | | 0.35 | | **0.33** | | 0.53 | | | | 0.53 | | | | 0.33 | |
| RMSE | | | | 0.71 | | 0.62 | | **0.40** | | 0.46 | | **0.42** | | 0.66 | | | | 0.66 | | | | 0.45 | |
| RMSPE | | | | 0.23 | | 0.16 | | **0.15** | | 0.16 | | **0.15** | | 0.22 | | | | 0.22 | | | | 0.15 | |
| | | | | | | | | | | | | | | | | | | | | | | | |
| No. of Parameters | N.A. | 6 | | 7 | | 14 | | 12 | | 27 | | | | 13 | 20 | | | 25 | | | | | |
| LL | | -1064 | | -1058 | | -993 | | -996 | | -983 | | | | -1041 | -1022 | | | -1041 | | | | | |
| AIC | | 2140 | | 2130 | | **2013** | | 2015 | | 2020 | | | | 2109 | | | | 2133 | | | | | |
| BIC | | 2176 | | 2172 | | 2098 | | **2088** | | 2182 | | | | 2186 | 2204 | | | 2282 | | | | | |
| CAIC | | 2182 | | 2179 | | 2112 | | **2100** | | 2209 | | | | 2199 | | | | 2307 | | | | | |

### Table 25: Simulation results for data from MM-MNL model

| True Model Parameters | | | | |
|---|---|---|---|---|
| TRUE | class probability | v1 (sigma=1) | v2 (sigma=1) | v3 (sigma=1) |
| class 1 | 0.40 | 0 | -1 | -2 |
| class 2 | 0.20 | 0 | -2 | -4 |
| class 3 | 0.40 | 0 | -4 | -8 |

**Analysis for Simulated data from Mixed - Mixed logit**

*Bold estimates are statisticcally significant at 1%

| | TRUE | Nlogit MNL | | Nlogit S-MNL-500 | | Nlogit G-MNL -1000 | | Nlogit MIXL-CORR-500 | | Bayesm 1 Mixture | | Nlogit -CL-3cls-500 | | | | | | Nlogit - MM-MNL - 3cls-500 | | | | | | Bayesm 3 Mixture | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | EST | Nlogit | SE | Nlogit | SE | Nlogit | SE | Nlogit | SE | EST (CC) | SE | Cla 1 | SE | Cla 2 | SE | Cla 3 | SE | Cla 1 | SE | Cla 2 | SE | Cla 3 | SE | ALL EST | SE |
| v1 | 0.00 | -0.03 | 0.06 | 0.06 | 0.08 | 0.01 | 0.12 | -0.05 | 0.11 | -0.056 | 0.100 | 0.03 | 0.13 | **0.49** | 0.13 | **-0.97** | 0.17 | 0.13 | 0.13 | **0.40** | 0.13 | **-1.04** | 0.17 | -0.10 | 0.11 |
| v2 | -2.40 | **-1.55** | 0.08 | **-2.56** | 0.30 | **-2.78** | 0.37 | **-2.47** | 0.20 | **-2.381** | 0.200 | **-3.42** | 0.28 | **-1.66** | 0.20 | **-0.73** | 0.14 | **-3.45** | 0.28 | **-1.65** | 0.20 | **-0.71** | 0.14 | **-2.45** | 0.19 |
| v3 | -4.80 | **-2.97** | 0.11 | **-4.95** | 0.55 | **-5.27** | 0.66 | **-4.59** | 0.35 | **-4.445** | 0.350 | **-7.01** | 0.51 | **-1.98** | 0.22 | **-1.95** | 0.18 | **-7.02** | 0.51 | **-1.91** | 0.21 | **-1.85** | 0.18 | **-4.57** | 0.33 |
| sigma(v1) | 1.00 | N.A. | | N.A. | | **0.84** | 0.13 | **0.81** | 0.11 | **0.74** | | | | | | | | 0.03 | 0.11 | 0.05 | 0.09 | 0.02 | 0.13 | **0.75** | |
| sigma(v2) | 1.00 | | | | | **0.87** | 0.18 | **1.35** | 0.15 | **1.31** | | N.A. | | | | | | 0.07 | 0.15 | 0.03 | 0.16 | 0.12 | 0.12 | **1.43** | |
| sigma(v3) | 1.00 | | | | | **0.94** | 0.29 | **1.93** | 0.27 | **2.39** | | | | | | | | 0.04 | 0.27 | 0.00 | 0.16 | 0.00 | 0.17 | **2.70** | |
| Tau | N.A. | N.A. | | 0 | - | 0.61 | 1.36 | N.A. | | N.A. | | | | | | | | N.A. | | | | | | | |
| Gamma | | | | N.A. | | 0.84 | 1.88 | | | | | | | | | | | | | | | | | | |
| Class Probability | | | | | | N.A. | | | | | | 0.49 | 0.06 | 0.28 | 0.06 | 0.23 | 0.05 | 0.48 | 0.06 | 0.27 | 0.06 | 0.25 | 0.05 | N.A. | |
| MAE | N.A. | | | 1.07 | | 0.88 | | **0.81** | | 0.86 | | 0.93 | | | | | | 0.93 | | | | | | **0.77** | |
| RMSE | | | | 1.45 | | 1.34 | | **1.09** | | 1.14 | | 1.20 | | | | | | 1.20 | | | | | | **1.04** | |
| RMSPE | | | | 0.14 | | 0.10 | | **0.10** | | 0.10 | | 0.12 | | | | | | 0.12 | | | | | | **0.10** | |
| | | | | | | | | | | | | | | | | | | | | | | | | | |
| No. of Parameters | N.A. | 3 | | 4 | | 8 | | 9 | | | | 11 | | | | | | 20 | | | | | | | |
| LL | | -1409 | | -1312 | | -1255 | | -1261 | | | | -1269 | | | | | | -1269 | | | | | | | |
| AIC | | 2823 | | 2632 | | **2527** | | 2541 | | | | 2560 | | | | | | 2578 | | | | | | | |
| BIC | | 2844 | | 2659 | | **2580** | | 2600 | | | | 2634 | | | | | | 2712 | | | | | | | |
| CAIC | | 2847 | | 2663 | | **2588** | | 2609 | | | | 2645 | | | | | | 2732 | | | | | | | |

## Appendix B – Additional code

Part A and Part B provide the code for the model estimations in software Nlogit 5 and the *bayesm* package. In part C, the code of simulation is based on an example in the *bayesm* package and used to generate observations from MM-MNL model.

## Part A: Code of software Nlogit 5

The command used in Nlogit 5 is written in a structured way, so here we only provide an example for the estimation of PIZZA delivery experiment A, for the estimation of other datasets, the user just need to modify the code accordingly.

1. Code for MNL model

```
sample ; all$
NAMELIST  ;  xs  =  asc_ts,asc_cf,asc_both,cost_ts,cost_cf,
cost_bot, doctor_r, carrier,indi_cou,risk_ts, risk_cf$
clogit    ; choice = TS,CF,BOTH,NONE
          ; Lhs=mode
          ; rhs=xs
          ; cluster=64
          ; PAR
          ; Output = 1 $
```

2. Code for S-MNL model

```
sample ; all$
NAMELIST  ;  xs  =  asc_ts,asc_cf,asc_both,cost_ts,cost_cf,
cost_bot, doctor_r, carrier,indi_cou,risk_ts, risk_cf$
smnlogit  ; choice = ChoiceA, choiceB
           ; Lhs=mode
           ; rhs=xs
           ; pds=16
           ; alg=bhhh
           ; Halton
           ; pts=1000
           ; maxit=200
           ; Output = 1 $
```

3. Code for G-MNL model

```
sample ; all$
NAMELIST  ; xs = asc_ts,asc_cf,asc_both,cost_ts,cost_cf,
cost_bot, doctor_r, carrier,indi_cou,risk_ts, risk_cf$
GMXlogit  ; choice = ChoiceA, choiceB
           ; Lhs=mode
```

```
            ; rhs=xs
            ;fcn=gourmet(n),price(n),ingredie(n),delivery(
     n),crust(n),size(n),steaming(n),late_hou(n)
            ; pds=16
            ; Halton
            ; draws=400
            ; maxit=100
            ;Tau=1.79
            ;Output = 1 $
```

4. MIXL model (Correlated model)

```
sample ; all$
NAMELIST  ; xs = asc_ts,asc_cf,asc_both,cost_ts,cost_cf,
cost_bot, doctor_r, carrier,indi_cou,risk_ts, risk_cf$
rplogit   ; choice = ChoiceA, choiceB
          ; Lhs=mode
          ; rhs=xs
          ;fcn=gourmet(n),price(n),ingredie(n),delivery(n
     ),crust(n),size(n),steaming(n),late_hou(n)
          ; pds=16
          ; Halton
          ; correlated
          ; pts=500
          ; maxit=200
          ; Output = 1 $
```

5. MIXL model (Uncorrelated model)

```
sample ; all$
NAMELIST  ; xs = asc_ts,asc_cf,asc_both,cost_ts,cost_cf,
cost_bot, doctor_r, carrier,indi_cou,risk_ts, risk_cf$
rplogit   ; choice = ChoiceA, choiceB
          ; Lhs=mode
          ; rhs=xs
          ;fcn=gourmet(n),price(n),ingredie(n),delivery(n
     ),crust(n),size(n),steaming(n),late_hou(n)
          ; pds=16
          ; Halton
          ; pts=500
          ; maxit=200
          ; Output = 1 $
```

6. LC model

```
sample ; all$
NAMELIST  ; xs = asc_ts,asc_cf,asc_both,cost_ts,cost_cf,
cost_bot, doctor_r, carrier,indi_cou,risk_ts, risk_cf$
lclogit   ; choice = TS,CF,BOTH,NONE
```

```
              ; Lhs=mode
              ; rhs=xs
              ; Pts = 4
              ; draws=500
              ; maxit=200
              ; Halton
              ; pds=16
              ; Output = 1 $
```

7. MM-MNL model

```
sample ; all$
NAMELIST  ; xs = asc_ts,asc_cf,asc_both,cost_ts,cost_cf,
cost_bot, doctor_r, carrier,indi_cou,risk_ts, risk_cf$
lcrplogit  ; choice = ChoiceA, choiceB
           ; Lhs=mode
           ; rhs=xs
           ; rpl
          ;fcn=gourmet(n),price(n),ingredie(n),delivery(n
     ),crust(n),size(n),steaming(n),late_hou(n)
           ;pds=16
           ;Halton
           ;pts=50
           ;LCM
           ;pts=2
           ;maxit=100
           ;Output = 3 $
```

## Part B: Code of the *bayesm* package in R

This appendix utilizes the dataset of PIZZA A example and provides the R code of model estimations.

The user is required to download the *bayesm* package in their R programme before using the code. The data structure used in the *bayesm* package and Nlogit 5 are in different format. The users have to transform it in the right format before further processing. An R package (Zhang , 2013) is available for possible structural transformation.

1. Code for MNL model

```
library(bayesm)
p=2            # number of alternatives
nset=16        # number of choice sets
nlgt=178       # number of respondents
```

```
## obtain x variables
x = subset(data4,select = -c(ID,choi_occ,alter,mode))
x=t(t(x))
## obtain y variables
     y=subset(data4,select=mode)
     y=t(t(y))
## prepare for model estiamtion
mnldata4=list(y=y,X=x,p=p)
## total number of iteration is 40000 and keep every 20th draw
mcmc1=list(R=40000,keep=20) #total=2000
out1=rmnlIndepMetrop(Data=mnldata4,Mcmc=mcmc1)
summary(out1$betadraw, burnin=10001)
beta= colMeans(out1$betadraw)
```

2. Code for MIXL model

By making necessary data transformation, now *lgtdata4* is ready for estimation in *bayesm* package.

```
nu=8           #number of parameters
## set the prior
Prior1 = list(nu=nu,V=diag(nu),ncomp=1)
mcmc1=list( R=600000,keep=15) #total=4000
out4mixl=rhierMnlRwMixture(Data=list(p=p,lgtdata=lgtdata4),Mcm
c=mcmc1,Prior=Prior1)
## the first half is discarded as the "burn-in" part and the
## last 2000 results are used to calculate relative results.
summary(out4mixl$nmix,burnin=2000)
## obtain the posterior means
mudraw=out4mixl$nmix$compdraw
itnum=1000
ind=2000-itnum+1
totalit=2000
postmu=NULL
for (i in ind:totalit)
{
     postmu=rbind(postmu,mudraw[[i]][[1]]$mu)
}
## produce the traceplot
matplot(postmu,type='l')
## store the individual parameters
estbeta = apply(out4mixl$betadraw[,,2001:4000],c(1,2),mean)
write.csv(estbeta,'data4-MIXL.csv')
```

3. Code for MM-MNL model

The estimation code is very similar to MIXL model; the only difference is we have to set different prior for the priors.

```
nu=8
## here two mixture components are set for priors.
Prior2 = list(nu=nu,V=diag(nu),ncomp=2)
mcmc1=list(R=60000,keep=15) #total=3000
out4mix2=rhierMnlRwMixture(Data=list(p=p,lgtdata=lgtdata4),Mcm
c=mcmc1,Prior=Prior2)
# the first half is the "burn-in" part.
summary(out4mix2$nmix,burnin=1500)
# make the trace plot for two components.
# first component
itnum=1500
ind=3000-itnum+1
totalit=3000
postmu1=NULL
mudraw42=out4mix2$nmix$compdraw
for (i in ind:totalit)
{
     postmu1=rbind(postmu1,mudraw42[[i]][[1]]$mu)
}
postmu1
par(mfrow=c(2,1))
matplot(postmu1,type='l')
# second component
itnum=1500
ind=3000-itnum+1
totalit=3000
postmu2=NULL
mudraw42=out4mix2$nmix$compdraw
for (i in ind:totalit)
{
     Postmu2=rbind(postmu2,mudraw42[[i]][[2]]$mu)
}
par(mfrow=c(2,1))
matplot(postmu1,type='l')
## store the individual parameter estimates
estbeta = apply(out4mix2$betadraw[,,1501:3000],c(1,2),mean)
write.csv(estbeta,'data4-MM-MNL2.csv')
```

## Part C: Code of simulation code in R

This simulation code is copied from the "*rhierMnlRwMixture*" function in *bayesm* package.

```
### simulating data for MM-MNL model with 3 mixture normals
p=3 # num of choice alterns
ncoef=3
nset=20 # num of choice occasions
nlgt=100 # num of cross sectional units
ncomp=3 # no of mixture components

comps=NULL
comps[[1]]=list(mu=c(0,-1,-2),rooti=diag(rep(1,3)))
comps[[2]]=list(mu=c(0,-1,-2)*2,rooti=diag(rep(1,3)))
comps[[3]]=list(mu=c(0,-1,-2)*4,rooti=diag(rep(1,3)))
pvec=c(.4,.2,.4)

simmnlwX= function(n,X,beta) {
## simulate from MNL model conditional on X matrix
k=length(beta)
Xbeta=X%*%beta
j=nrow(Xbeta)/n
Xbeta=matrix(Xbeta,byrow=TRUE,ncol=j)
Prob=exp(Xbeta)
iota=c(rep(1,j))
denom=Prob%*%iota
Prob=Prob/as.vector(denom)
y=vector("double",n)
ind=1:j
for (i in 1:n)
{yvec=rmultinom(1,1,Prob[i,]); y[i]=ind%*%yvec}
return(list(y=y,X=X,beta=beta,prob=Prob))}

## simulate data
simlgtdata=NULL
ni=rep(nset,nlgt)
for (i in 1:nlgt)
{ betai=as.vector(rmixture(1,pvec,comps)$x)
Xa=matrix(runif(ni[i]*p,min=-1.5,max=0),ncol=p)
X=createX(p,na=1,nd=NULL,Xa=Xa,Xd=NULL,base=1)
outa=simmnlwX(ni[i],X,betai)
simlgtdata[[i]]=list(y=outa$y,X=X,beta=betai)}
```

```
simlgtdata[[2]]
bmat=matrix(0,nlgt,ncoef)
for(i in 1:nlgt) {bmat[i,]=simlgtdata[[i]]$beta}

## save the simulation dataset
write.csv(bmat,'simummmnl-truebeta.csv')
```

# Bibliography


Agresti, A. (2013). *Categorical Data Analysis.* New Jersey: John Wiley & Sons .

Allenby, G. M., & Rossi, P. E. (1999). Marketing models of consumer heterogeneity. *Journal of Econometrics, 89*, 57-78.

Allenby, G. M., Rossi, P. E., & McCulloch, R. E. (2005,January). *SSRN.* Retrieved 4 2014,from SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=655541

Burda, M., Harding, M., & Hausman, J. (2008). A Bayesian mixed logit-probit model for multinomial choice. *Journal of Econometrics* , 147:2,232-246.

Fiebig, D. G., Keane, M. P., & Jordan , L. (2009). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science, 29(3)*, 393-421.

Gelfand, A. E., & Smith, A. F. (1990, June). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association, 85*(410), 398-409.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 721-741.

Geweke, J. (1989). Bayesian Inference in Econometric Models Using Monte Carlo Integration. *Econometrica* , 1317-1339.

Geweke, J., & Keane, M. (2001). Computationally intensive Methods for Integration in Econometrics. *Handbook of Econometrics*, 3463-3568.

Greene, H. W., & Hensher, A. D. (2013). Revealing Additional Dimensions of Preference Heterogeneity in a Latent Class Mixed Multinomial Logit Model. *Journal Applied Economics, 45*(14 (May)), 1897-1902.

Greene, W. H., & Hensher, D. A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research: Part B, 37*, 681-698.

Greene, William H.; Econometric Software, Inc. (2012). NLOGIT 5 Reference Guide. Plainview, NY, USA.

Guindani, M. (2008). *Statistical Computing.* Retrieved April 2014, from University of New Mexico: http://www.math.unm.edu/~michele/Teaching/SC/material/Lecture31.pdf

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 97-109.

Huber, J., & Train , K. (2001). On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths. *Marketing Letters*, 259-269.

Huber, J., & Train , K. (2001). On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths. *Marketing Letters, 12*(3), 257-267.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 679-688.

Keane, M., & Wasi, N. (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics,Vol.28(6)*, 1018-1045.

Kragt, M. E. (2013, February 5-8). Comparing models of unobserved heterogeneity in environmental choice experiments. *57th.* Sydney: Australian Agricultural and Resource Economics Society,2013 Conference (57th).

Kuhfeld, W. F. (2010). Discrete Choice. United States. Retrieved from SAS.

Kutner, M. H., Nachtsheim, J. C., Neter, J., & Li , W. (2004). *Applied Linear Statistical Models.* Blacklick, Ohio: McGraw-Hill College.

Lesaffre, E., & Lawson, A. B. (2012). *Bayesian Biostatistics.* Chichester, West Sussex, United Kingdom: John Wiley & Sons Ltd.

Louviere , J., Islam, T., Wasi, N., Street, D., & Burgess, L. (2008). Designing discrete choice experiments: Do optimal desgins come at a price? *Journal of Consumer Research* , 35, 360-375.

Luce, D. (1959). *Individual Choice Behavior.* New York: John Wiley and Sons.

Marschak, J. (1960). Binary choice constraints on random utility indicators. *Stanford Symposium on Mathematical Methods in the Social Science*, 312-329.

McCulloch, R. E., Polson, N. G., & Rossi, P. E. (2000). A Bayesian analysis of the multinomial probit model with fully identified parameters. *Journal of Econometrics*, 173-193.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, 105-142.

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15. 447-470.

Metropolis , N., Rosenbluth, A. W., & Rosenbluth, M. N. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087-1092.

Revelt, D., & Train, K. (1998). Mixed Logit with Repeated Choices: Households' Choice of Appliance Efficiency Level. *The Review of Economics and Statistics*, 647-657.

Robert, C., & Casella, G. (2010). *Introducing Monte Carlo Methods with R.* Heidelberg : Springer-Verlag Berlin.

Rossi, P. (2012, 5 15). Package 'bayesm'. Chicago, USA.

Rossi, P., Allenby, G., & McCulloch, R. (2006). *Bayesian Statistics and Marketing.* New York: John Wiley and Sons.

Smith , R. L., & Tierney, L. (1996). Exact transition probabilities for the independence Metropolis sampler. *University of Cambridge Statistical Laboratory research report*, 96–16.

Train, K. (2001). *A comparison of hierarchical bayes and maximum simulated likelihood for mixed logit.* Berkerly, department of economics university of california: Working Paper.

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

Zhang , H. (2013). *Estimation of discrete choice models in R: Writing an R package.* Leuven: KU Leuven.

**AFDELING**
Straat nr bus 0000
3000 LEUVEN, BELGIË
tel. + 32 16 00 00 00
fax + 32 16 00 00 00
www.kuleuven.be