Performance Analysis of Mixed Logit Models for Discrete Choice Models

Jaka Nugraha
Department of Statistics
Universitas Islam Indonesia, Indonesia
jnugraha@uii.ac.id

Abstract

Mixed Logit model (MXL) is generated from Multinomial Logit model (MNL) for discrete, i.e. nominal, data. It eliminates its limitations particularly on estimating the correlation among responses. In the MNL, the probability equations are presented in the closed form and it is contrary with in the MXL. Consequently, the calculation of the probability value of each alternative get simpler in the MNL, meanwhile it needs the numerical methods for estimation in the MXL. In this study, we investigated the performance of maximum likelihood estimation (MLE) in the MXL and MNL into two cases, the low and high correlation circumstances among responses. The performance is measured based on differencing actual and estimation value. The simulation study and real cases show that the MXL model is more accurate than the MNL model. This model can estimates the correlation among response as well. The study concludes that the MXL model is suggested to be used if there is a high correlation among responses.

Keywords: Logit; Maximum Likelihood; Monte Carlo simulation; Utility model

Introduction

Discrete choice models (DCM) describe decision maker's choices among alternatives. The decision makers can be people, households, firms, or any other decision-making unit, and the alternatives might represent competing products, courses of action, or any other options or items over which choices must be made. Multinomial Probit Model (MNP), MNL and Mixed Logit Model (MXL) are part of DCM. These models are often used to analyze the discrete choices made by individuals recorded in survey data. These models have been applied in marketing, transportation, political research, energy, housing etc. (Camminatiello and Lucadamo, 2010; Yang, 2010; Lovreglio et al., 2016; Merkert and Beck, 2017).

Each model has advantages and disadvantages. The advantage of MNL is to have a simpler form of equation in estimating the model. The weakness of MNL is that the model is prepared by using irrelevant alternative independence (IIA). Meanwhile, MNP was developed to overcome MNL weakness in terms of IIA. However MNP has an open form of equation, so its model estimation requires simulation and iteration methods such as Newton-Raphson.

The MXL has been developed to overcome the limitations of both of the MNL and the MNP for both of the IIA assumption and the computational aspects. MXL can be derived under a variety of different behavioral specifications, and each derivation provides a particular interpretation. The MXL is defined on the basis of the functional form for its choice probabilities. Any behavioral specification which derived the choice probabilities will take this particular form, is called an MXL (Train, 2003). Refer to vast and popular applications of the MXL, it was stated that the MXL as the model of the future (Bhat, 1998; Bolduc, 1999; Lovreglio et al., 2016; Train, 2016) which proposed a flexible

procedure to represent the distribution of the random parameter model (Train, 2016). However it raises the next issue on how accurate the MXL compared to MNL.

This paper, the profile of MXL for detecting the correlation among alternatives on utilities models is studied. Model was constructed and applied for simulation data on several rank of dependency (correlation) among alternatives and R.3.1.1 software was adopted to run the computational work. Parameters in MXL and MNL are estimated by using the MLE method which has good properties for large samples, especially asymptotically efficient and asymptotic bias in simulated maximum likelihood estimation (Lee, 1995; Horowitz, and Savin, 2001; Lee, 1992). Based on these properties, the estimator accuracy can be measured by using the distance between the estimator and of the parameters at various levels of correlation.

Aim of study is to evaluate the effect of correlation among choice on MXL and MNL models. In order to organize the scheme of study, the specification of MNL and MXL, how to perform parameter estimation using MLE and accuracy of parameters estimation in the MNL and MXL using simulation data are presented. Some examples of the application of MNL and MXL in real cases are also described.

Utility Model

To fit in a discrete choice framework, the set of alternatives, called the choice set, needs to exhibit three characteristics. First, the alternatives must be mutually exclusive from the decision maker's perspective. The decision maker chooses only one alternative from the choice set. Second, the choice set must be exhaustive, in that all possible alternatives are included. The decision maker necessarily chooses one of the alternatives. Third, the number of alternatives must be finite. A decision maker (respondent), denoted as i was faced with a choice among J alternatives. The respondent has a certain level utility (or profit) for each alternative j. U_{ij} for j=1,...,J is the utility that respondent i obtain from alternative j and the real value of U_{ij} that is unkonwn by researcher. The decision maker chooses the alternative that provides the greatest utility. Researcher did not know the value of utility for respondents in each option and looked at the attributes, is denoted Z_{ij}, that exist for each choice and respondents attribute is denoted X_i.

A function that relates these observed factors to the decision maker's utility is denoted V_{ij} and is often called representative utility,

$$V_{ij} = \alpha_j + \beta_j X_j + \gamma_j Z_{ij} \tag{1}$$

where i=1,...,n and j=1,...,J. V_{ij} is assigned as representative utility. Equation (1) is a model constructed and reported by Boulduc (1999). Due to the unkown value of U_{ij}, therefore

$$U_{ij} = V_{ij} + \varepsilon_{ij}. \tag{2}$$

$\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ})'$ is a random variable having density of $f(\boldsymbol{\varepsilon}_i)$, where V_{ij} is observed factor and ε_{ij} is unobserved factor in utility. The density $f(\boldsymbol{\varepsilon}_i)$ is the distribution of the unobserved portion of utility within the population of people who face the same observed portion of utility. Different choice models are derived under different specifications of the density of unobserved factors, $f(\boldsymbol{\varepsilon}_i)$.

Multinomial Logit Model (MNL)

MNL model is derived under the assumption that ε_{ij} is independent and identically distributed (iid) extreme value for all j, Function of extreme value density type I (Gumbel) is

$$f(\varepsilon_{ij}) = \exp(-\varepsilon_{ij}) \cdot \exp(-\exp(-\varepsilon_{ij})) \qquad . \qquad (3)$$

Mean and the variance from this distribution are 0.5772 and $\pi^2/6$, respectively. Cummulative distribution of Extreme Value type I is

$$F(\varepsilon_{ij}) = \exp(-\exp(-\varepsilon_{ij})) \quad . \qquad (4)$$

The Extreme Value distribution and the Standard Normal distribution can be shown in Figure 1.
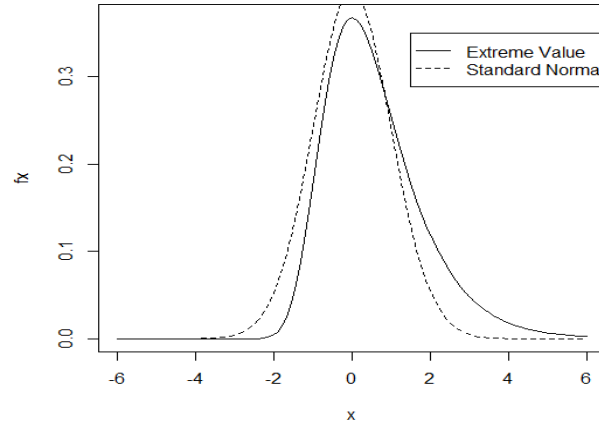

Figure 1: Extreme Value and Standard Normal distribution

Based on Figure 1, it can be seen that the Extreme Value distribution is almost symmetrical distribution with a standard normal distribution. Though this distribution has a thicker or heavy tail compared to the normal distribution. The probability of decision maker i choice k alternative is stated as:

$$\pi_{ik} = \frac{\exp(V_{ik})}{\sum_{j=1}^{J} \exp(V_{ij})}. \qquad (5)$$

This formula is assigned as logit probability (Train, 1998). Parameters $(\alpha_j, \beta_j, \gamma)$ can be estimated by using MLE.

Mixed Logit Model (MXL)

Mixed logit assumes that the unobserved portions of utility are a mixture of an independent and identically distributed (iid) extreme value term and another multivariate distribution selected by the researcher. This general specification allows MXL to avoid imposing the IIA property on the choice probabilities. Further, MXL is a flexible tool for examining heterogeneity in responden behavior through random coefficients specifications. Mixed Logit is a very flexible model that can be approached by several random utility models (McFadden and Train, 2000). In Equation (2), if a δ_j random variable having density $f(\delta_j)$ is added then utility model can served in the following form:

$$U_{ij} = V_{ij} + \delta_j + \varepsilon_{ij}. \qquad (6)$$

Generally, the assumption of $f(\delta_j)$ in standard normal distribution is used, $\delta_j \sim N(0,1)$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_1) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

It is assumed that ε_{ij} has an extreme value distribution and is independent to δ_j. The MXL is a standard integral Logit in respect to density $\boldsymbol{\delta}$. The probabily of i-th respondent chooses the k-th alternative can be formulated into:

$$\pi_{ik} = \int g_{ik}(\boldsymbol{\delta}) f(\boldsymbol{\delta}) d\boldsymbol{\delta} \qquad (7)$$

where $g_{ik}(\boldsymbol{\delta})$ is logit probability that can be writen as:

$$g_{ik}(\boldsymbol{\delta}) = \frac{\exp{(V_{ik}+\delta_k)}}{\sum_{j=1}^{J}\exp{(V_{ij}+\delta_{ij})}} \tag{8}$$

The probability in MXL is a weighted mean to the logit by using weighting density function $f(\boldsymbol{\delta})$. The MXL is a mixed form of logit function and density function $f(\boldsymbol{\delta})$ and the value of probability in Equation (8) can be approached/computed by using simulation. The steps of simulation are as following (Train, 2003) :

a. Take a value of $\boldsymbol{\delta}$ from density $f(\boldsymbol{\delta})$ and label it as $\delta^{(r)}$. At the first take $r=1$.
b. Calculate the logit probability $g_{ik}(\delta^{(r)})$ from equation (8).
c. Repeat the steps 1 and 2 for R–times and evaluate the average

$$\tilde{\pi}_{ik} = \frac{1}{R}\sum_{r=1}^{R}g_{ik}(\delta^{(r)}) \tag{9}$$

Maximum Likelihood Estimation (MLE) on Multinomial Distribution

Let $Y_1, Y_2,...,Y_n$ be random variables having pooled density:

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = f(y_1,...,y_n|\theta_1,...,\theta_J) \tag{10}$$

These functions depend on parameters $\boldsymbol{\theta} = (\theta_1,...,\theta_J)$. As $\mathbf{Y}_i$ is independent each other, then we have

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n}f(y_i,\theta) \tag{11}$$

The Likelihood function, labeled $L(\boldsymbol{\theta}|\boldsymbol{y})$, is a function of the parameters of a statistical model given data, is

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \prod_{i=1}^{n}f(\theta\mid y_i) \tag{12}$$

Suppose $\boldsymbol{\Omega}$ that is a probable/possible set of values for vector parameter $\boldsymbol{\theta}$ and is also assigned as parameter space. In another experiment. it was defined that MLE for $\boldsymbol{\theta}$, denoted as $\widehat{\boldsymbol{\theta}}_{MLE}$ is value of $\boldsymbol{\theta}$ which maximizes the likelihood function $L(\boldsymbol{\theta}|\boldsymbol{y})$ on data $\boldsymbol{y}$ (Greene, 2005).

$$\widehat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\theta}\mathrm{L}(\theta;y) \tag{13}$$

If $Y_1, Y_2,...,Y_n$ are random samples having multinomial density, then

$$f(y_i|\pi_1,...,\pi_J) = \pi_1^{y_{i1}}...\pi_J^{y_{iJ}} \qquad \text{for } i=1,...,n.$$

The π_{ij} is the probability of the i-th decision maker choose the *j-th* alternative as in the equations (5) and (7) that obtained parameter $\boldsymbol{\theta}$. Therefore, $\pi_{ij} = \pi_{ij}(\boldsymbol{\theta})$. The Likelihood function for parameter $\boldsymbol{\theta}$ can be constructed as:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n}\pi_{i1}^{y_{i1}}...\pi_{iJ}^{y_{iJ}}$$

The log-likelihood function is

$$\log L(\boldsymbol{\theta}) = \sum_{i=1}^{n}\log L_i(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J}y_{ij}\log[\pi_{ij}] \tag{14}$$

MLE is the value of $\boldsymbol{\theta}$ that maximizes the $\log L(\boldsymbol{\theta})$ function or

$$\widehat{\boldsymbol{\theta}}_{MLE} = \arg\max_{\boldsymbol{\theta}}\log\mathrm{L}(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}}\sum_{i=1}^{n}\sum_{j=1}^{J}y_{ij}\log[\pi_{ij}]$$

In MXL model, π_{ij} is a double integral that can be calculated/computed by using/implementing the simulation method. For example $\tilde{\pi}_{ij}$ is the value of π_{ij} that was calculated by using the simulation in Equation (9). The simulated likelihood function was obtained by substituting the π_{ij} value into log-likehood function in Equation (14) by simulating the $\tilde{\pi}_{ij}$ value.

$$simlog\ L(\boldsymbol{\theta}) = \sum_{i=1}^{n} simlog L_i(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij}\log[\tilde{\pi}_{ij}]$$

The $\hat{\boldsymbol{\theta}}_{MSL}$ is the $\boldsymbol{\theta}$ value which maximizes the $simlog\ L(\boldsymbol{\theta})$.

$$\hat{\boldsymbol{\theta}}_{MSL} = \arg\max_{\theta}\ simlogL(\boldsymbol{\theta}) = \arg\max_{\theta}\sum_{i=1}^{n}\sum_{j=1}^{J} y_{ij}\log[\tilde{\pi}_{ij}] \tag{15}$$

Simulation Studies

In order to detect the influence of correlation among alternatives, the multinomial data was generated for *J*=3. Three alternatives were chosen as representative for the correlation structure among alternatives for J case in general. First, there is a correlation between alternative j and j', meanwhile there is no correlation among them. Therefore, conclusion from the simulation at case J=3 can be generalized for all value of J. It is assumed that there is a correlation among the first and the second, but there is no correlation for the third alternative.

For example an application in the selection of the mode of transport, there are three alternatives: private car, taxi, and public transport. Taxi is probably related to the private car, in a sense that for someone has no private car then taxi will be chosen and vice versa. Furthermore, the simulated data is generated using the following utility model:

$$U_{i1} = V_{i1} + \delta_{i1} + \varepsilon_{i1};\ U_{i2} = V_{i2} + \delta_{i2} + \varepsilon_{i2};\ U_{i3} = V_{i3} + \varepsilon_{i3} \tag{16}$$

where $\varepsilon_{ij}\sim$*Extreme Value* type I, $\delta_{i1}\sim N(0,1)$, $\delta_{i2}\sim N(0,1)$, $\boldsymbol{\delta} = (\delta_1,\delta_2)\sim N(\mathbf{0},\boldsymbol{\Sigma})$ i=1,...,n;

j=1,2,3 and $\Sigma = \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & 1 \end{pmatrix}$.

It is assumed that $V_{ij} = \alpha_j + \beta_j X_j + \gamma Z_{ij}$. X_i is a characteristic individual/decision maker and Z_{ij} is the attribute of the alternative. Third alternative was assumed as a baseline, $\alpha_3 = \beta_3 = 0$. Data was generated on parameter $\alpha_1 = -1$, $\alpha_2 = 1$, $\beta_1 = 0.5$, $\beta_2 = -0.5$, $\gamma = 1$ and on several value of σ_{12}. Based on these simulation data was estimated by using MNL and MXL. Furthermore, based on several values of σ_{12}, estimators obtained on MXL are compared to those were obtained by MNL.

Results and Discussion

It is clear from Equation (3) that $Var(\varepsilon_{ij}) = \pi^2/6$ and is assumed that ε_{ij} and δ_{ij} are independent. Therefore, the covariance among alternatives on Equation (16) is

$$Cov(U_i) = \begin{pmatrix} \sigma_{12} + \dfrac{\pi^2}{6} & \sigma_{12} & 0 \\ \sigma_{12} & \sigma_{12} + \dfrac{\pi^2}{6} & 0 \\ 0 & 0 & \dfrac{\pi^2}{6} \end{pmatrix} \tag{17}$$

The correlation between alternative 1 and alternative 2 is

$$\rho = Cor(U_{i1}, U_{i2}) = \frac{\sigma_{12}}{\sigma_{12} + \frac{\pi^2}{6}} \geq 0 \quad \text{and} \quad \sigma_{12} = \frac{Cor(U_{i1}, U_{i2}).\frac{\pi^2}{6}}{1 - Cor(U_{i1}, U_{i2})} \tag{18}$$

The equation (18) shows a relationship between covarians (σ_{12}) to the correlation (ρ) value. Data were generated on several values ρ, and they are served in Table 1.

Table 1: Relationship between covarians (σ_{12}) to the correlation (ρ) value.

No.	σ_{12}	ρ	No.	σ_{12}	ρ	No.	σ_{12}	ρ	No.	σ_{12}	ρ
1	0.1	0.06	6	2	0.55	11	7	0.81	16	12	0.88
2	0.2	0.11	7	3	0.65	12	8	0.83	17	13	0.89
3	0.3	0.15	8	4	0.71	13	9	0.85	18	14	0.89
4	0.5	0.23	9	5	0.75	14	10	0.86	19	15	0.90
5	1	0.38	10	6	0.79	15	11	0.87			

Packages of R.3.1.1 software were employed to generate data and estimate the parameter. Some packages in R software are:

a. *randtoolbox* : to generate Halton series that can be used to calculate normal density integral,
b. *micEcon* : to calculate MLE on MNL model,
c. *mnormt* : to generate multivariate normal data,
d. *adapt* : to evaluate maximum point from log-likelihood function on Mixed Logit model.

Before simulation using correlation structure as presented in Table 1, it is important to conduct simulation to observe the influence of sample size towards parameter estimation value. The observation on the effect of sample size to the estimator was conducted for the independent structure $\sigma_{12} = 0$ or $\rho = 0$ for n=50, n=100, n=500, n=1000 and n=5000. The parameter estimation was perforemd by using *geepack* packages in R software. The estimation results are presented in Figure 2 – 4.
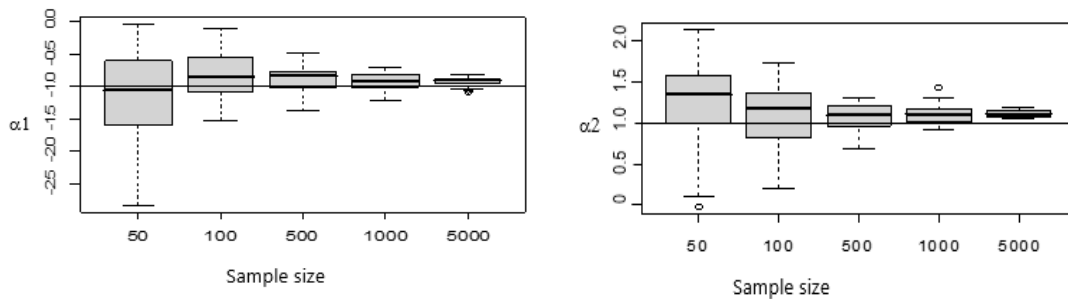

Figure 2.: Box plot of parameter $\alpha_1 = -1$ and $\alpha_2 = 1$
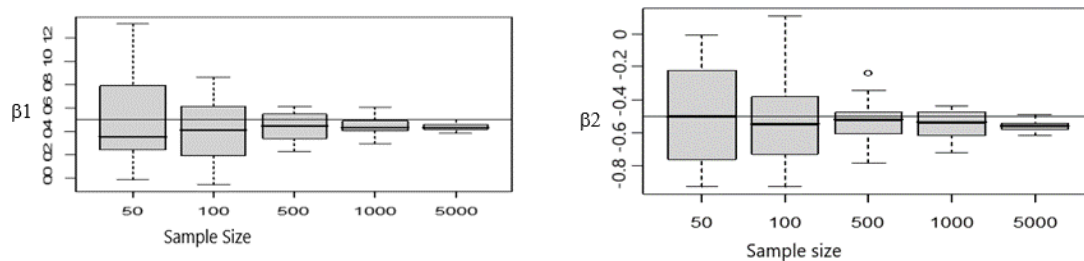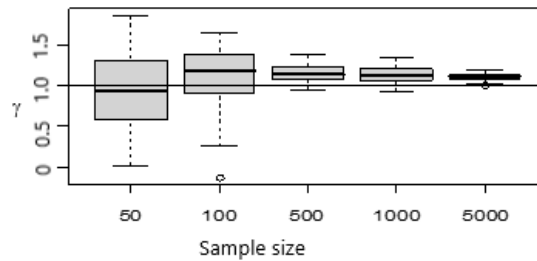

Figure 3.:Box plot of parameter $\beta_1 = 0.5$ and $\beta_2 = -0.5$

Figure 4. : Box plot of parameter $\gamma = 1$

From Figure 2-4 it is seen that for the sample size less than 100, the big variance is resulted. This is due to the wide deviation which resulted from some samples. It is concluded that for n=50 and 100 the resulted estimators are instable. Moreover, the estimator resulted from n=500, 1000 and 5000 are stable enough which there is no such wide deviation from the targeted value. The bigger sample size, the obtained estimator approaching and fit to the real values. Morever, the bigger n, the more stable estimator or the variance of each estimator is smaller.

Eventhough n is big enoughr (n=5000), the bias is still available. The reasons are, first, the generated data are in normal distribution (not from the extreme value distribution), which the normal and extreme value distribution have different variances. Mean and the variance from distribution of Extreme Value type I are 0.5772 and $\pi^2/6$, respectively Generalized estimating equation (GEE) model, as utilized for geepack packages is based on the Extreme Value distribution.

Furthermore, the simulation was performed in order to evaluate the influence of towards estimator of all parameters with the sample size, n=1000. The actual correlation value and its estimation value are presented in Figure 5.
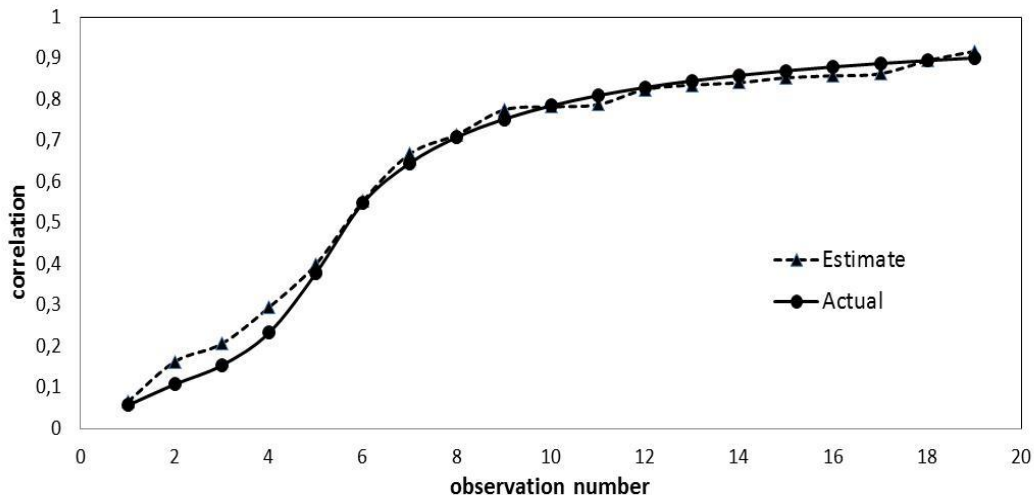

Figure 5: Plot of actual and estimate correlation

From Figure 5., it can be seen that MXL could estimate the correlation parameter, particularly on the correlation value more than 0.4. By using the hypothesis H_0 : $\rho_{actual}=\rho_{estimate}$, it can be concluded that there is insignificant correlation among actual and estimated value on the *p-value* of 0.833413. It suggests that MXL has better ability to estimate the correlation of the parameter.

Further the bias of each parameter $(\alpha_1,\ \alpha_2,\ \beta_1,\ \beta_2, \gamma)$ are illustrated in Figure 6 to Figure 10.
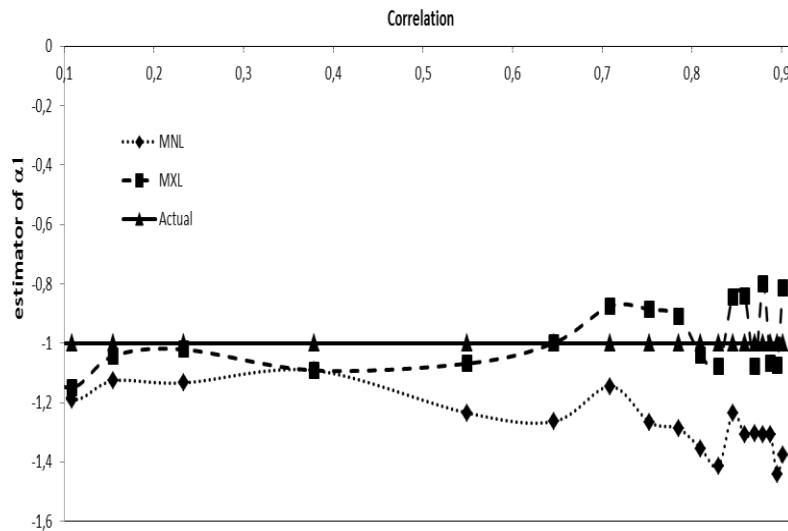

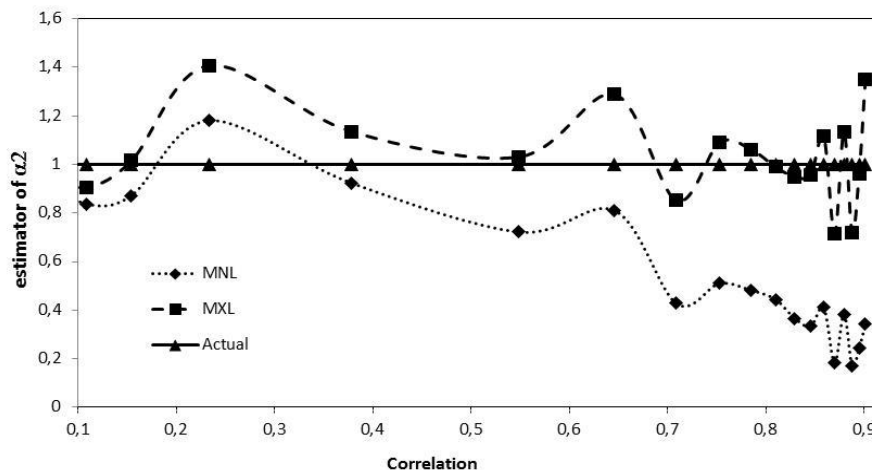Figure 6: Comparison on estimator of α_1 in MNL and MXL

Figure 7: Comparison on estimator of α_2 in MNL and MXL

According to Figures 6 and 7, the α estimator in MNL has a great bias is proportional to the magnitude of the correlation between the response. At the level of correlation is high (more than 0.7), the magnitude of the bias in α estimator reaches more than 50%. Meanwhile, MXL is not affected by the correlation between the response. The estimators in MNL always smaller than the MXL. On the actual parameter $\alpha_1 = -1$, the mean of estimator in both MNL and MXL are -1.2560 and -0.9873 respectively. By using hypothesis H₀: $\alpha_1 = -1$, the significant difference among actual and estimated value was observed in MNL, with the *p-value* of 4.011x10⁻⁹. Meanwhile, MXL gives the insignificant result among them at the *p-value* of 0.626.

For the actual parameter $\alpha_2 = 1$, the estimator mean in MNL and MXL are 0.5578 and 1.0333 respectively. By using hypothesis H₀: $\alpha_2 = 1$, insignificant values of the actual and estimated values are observed using MXL with the *p-value* of 0.441. On tare he MNL, the *p-value* of 3.801x10⁻⁶ is observed, indicating the significant difference among actual and estimated value.
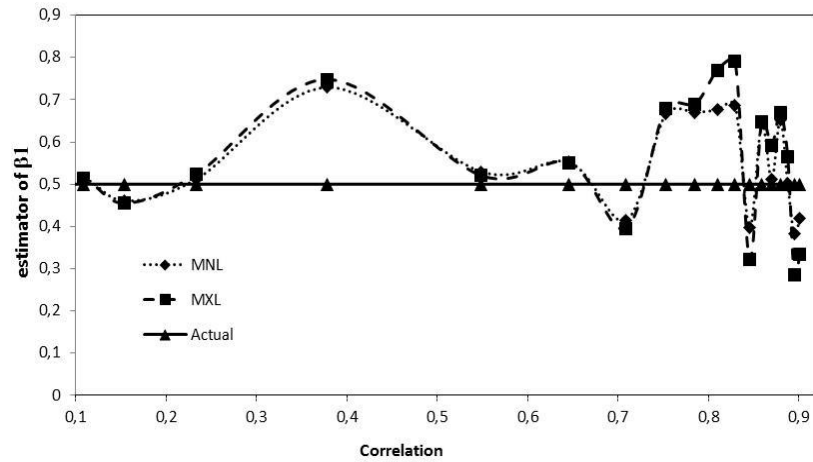
Figure 8: Comparison on estimator of β_1 in MNL and MXL

Based on Figure 8, β_1 estimator in the model MNL and MXL is the same mean of 0.5488 and 0.5557 respectively. By the actual value $\beta_1 = 0.5$, it can be concluded that there is no significant difference between actual and estimated values.

On the $\beta_2 = -0.5$, the almost similar result is obtained with whose observed on the parameter α, MNL gives significant difference of estimator among significant and actual values with *p-value* 5.659×10^{-5}. The mean of estimator β_2 on MNL model is -0.3759, while insignificant estimator among both values is observed in the MXL with *p-value* of 0.415. Mean of estimator β_2 on model MNL is -0.4800.

The results is in line with the parameter γ that has actual value of 1 in that the MNL model produces the predictor value of 0.7347 that gives significantly difference among actual and estimated values (*p-value* of 3.508×10^{-7}), meanwhile the MXL model gives the estimator value of 1.0115, insignificant difference with actual value. *P-value* of MXL model is 0.566.
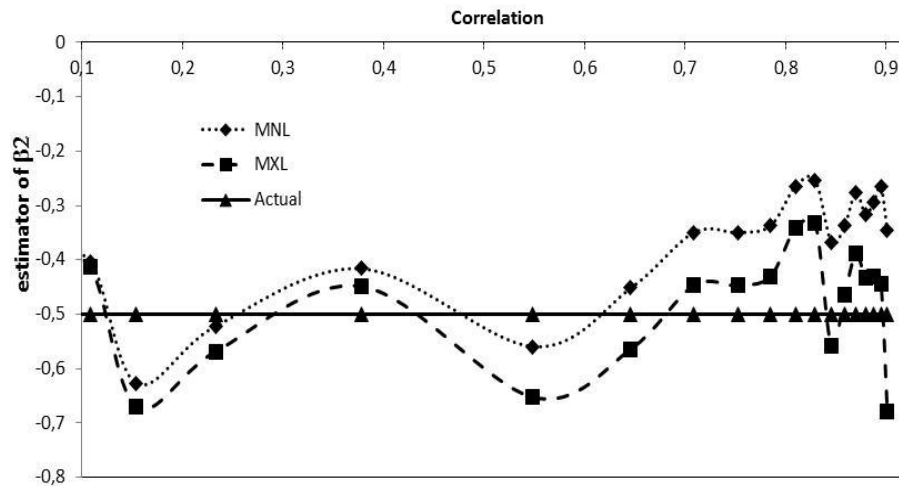

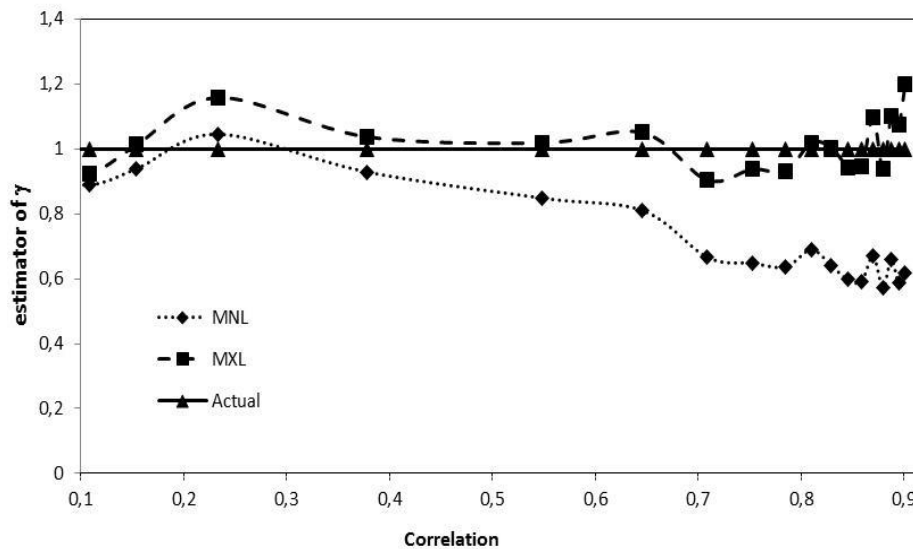Figure 9: Comparison on estimator of β_2 in MNL and MXL

Figure 10: Comparison on estimator of γ in MNL and MXL

Figures 9 and 10 illustrate that the β_2 estimator and γ estimator in MNL have a great bias is proportional to the magnitude of the correlation between (or among) the response. Meanwhile, MXL is not affected by the correlation between (or among) the response. These results are similar to those seen in the α parameter.

Some conclusions which can be derived from Figures 5 to 10 are:
a. Correlation parameters can be well estimated by MXL.
b. In general the bias on MNL model is higher than on MXL.
c. For the intercept parameters (i.e : α_1, α_2). In case of the high correlation (more than 0.7) presents then the MNL produces a higher bias in comparison to the MXL.
d. For coefficient parameter X (i.e : β_1). bias from MNL model is relatively equal to MXL.
For coefficient parameter Z (i.e : γ). In case of high correlation (more than 0.5) then MNL produces a higher bias than MXL.

Applications in Real Cases

In this section, we apply the MNL and MXL in real problems. There are two cases, each represents the state of the low and high correlation circumstances. The first case represents a weak correlation is taken from data "*Electricity*" and the second one represents a high correlation, is taken from the data "*Heating*". Both of them are from *mlogit* package in *R.3.1.1*

Case 1. Data "*Electricity*" are taken from *mlogit* package in R.3.1.1. A sample of residential electricity customers were asked a series of choice experiments. In each experiment, four hypothetical electricity suppliers were described. The person as asked which of the four suppliers he/she would choose (j=1,2,3,4). In the experiments, the characteristics of each supplier were stated. *Pf* is fixed price at a stated cents per kWh for each choice (electricity suppliers). *Cl* is the length of contract that the supplier offered, in years (such as 1 year or 5 years).

As the first supplier (j=1) is stated as base line, the representative utility models can be expressed as

$$V_{ij} = \alpha_j + \gamma_1 f_{ij} + \gamma_2 Cl_{ij} \quad \text{for j=2,3,4.}$$

The results of estimating MNL and MXL are presented in Table 2.

Table 2 : Estimator of the parameter for data *"Electricity"*

Parameter	MNL		MXL	
	Estimate	***p-value***	**Estimate**	***p-value***
α_2	0.081999	0.072198	0.145438	0.031174
α_3	0.085752	0.070240	0.149103	0.032027
α_4	0.174372	0.018980	0.200542	0.017271
γ_1	0.055185	0.001551	0.076797	0.001710
γ_2	-0.058379	0.005354	-0.084129	0.003903
Var(**Pf**)			0.145916	0.000993
Cov(**Cl.Pf**)			0.059773	0.009869
Var(**Cl**)			0.249205	0.000919
log L($\hat{\theta}$)	-5827.6		-5413.1	
Likelihood Ratio	266.6		1095.7	
Time	3 iterations. 0h:0m:0s		13 iterations. 0h:1m:9s	

Based on the results in Table 2, the correlation values obtained is

$$\hat{\rho} = \frac{Cov(Cl.Pf)}{\sqrt{Var(Pf)Var(Cl)}} = 0.313453$$

The value of this correlation still include low correlation, so that the MNL and MXL relatively the same. The calculation process in the MXL takes longer than the MNL.

The Likelihood Ratio statistic (LR) is an alternative for testing hypotheses about θ defined by

$$LR = -2log\frac{L(\theta_0)}{L(\hat{\theta})} \tag{19}$$

where $\hat{\theta}$ denotes the MLE of θ and θ_0 denotes the restriction that H_0 is true. H_0 states that all parameters $\boldsymbol{\theta = 0}$. $L(\hat{\theta})$ is the value of the likelihood function at the estimated parameters dan $L(\theta_0)$ is its value when all parameters are set equal to zero. Statistic *LR* is Chi Square distributed with the degree of freedom equal to tested parameters. The value of *LR* on MXL model is 1095.7 which is higher compared to that of derived by MNL model (266.6). This means that MXL model give more reliable conclusion compared with MNL model. Similar pattern is also obtained from partial test in that the *p-value* of MXL model is less that 0.04 while MNL model gives *p-value* of α_2 and α_3 more are higher than 0.07 parameter.

Train (2003) reported that likelihood ratio index can be utilized for measure the fitness of model and data. The likelihood index is defined as

$$R^2 = 1 - \frac{Log\ L(\hat{\theta})}{Log\ L(\theta_0)} \tag{20}$$

R^2 of MXL is 0.092 which is larger than the MNL which is 0.022. Its means that MXL model better than MNL model.

Case 2. Data *"Heating"* are taken from *mlogit* package in R.3.1.1
The observations consist of single-family houses in California that were newly built and had central air conditioning. The choice is among heating systems. Five types of systems are considered to have been possible: gas central (*gc*), gas room (*gr*), electric central (*ec*), electric room (*er*), heat pump (*hp*). There are two characteristic variable (installation cost and annual operating cost). *Ic* is the installation cost for the five alternatives. *Oc* is the

annual operating cost for the 5 alternatives. As type *ec* is stated as base line, representative utility models can be expressed as:

$$V_{ij} = \alpha_j + \gamma_1 Ic + \gamma_2 Oc_{ij} \quad \text{for } j=2,3,4,5.$$

From 900 observations, the results of estimating MNL and MXL are presented in Table 3.

Table 3. Estimator of the MXL and MNL parameter for Data *"Heating"*

Parameter	MNL		MXL	
	Estimate	p-value	Estimate	p-value
α_2 (*er* intercept)	0.194591	0.166861	0.788838	0.032164
α_3 (*gc* intercept)	0.052133	0.314373	0.203964	0.299783
α_4 (*gr* intercept)	-1.350583	0.039336	-4.139536	0.013045
α_5 (*hp* intercept)	-1.658846	0.021676	-4.526849	0.008003
γ_1	-0.001533	0.044845	0.010199	0.004768
γ_2	-0.006996	0.014967	-0.019823	0.007853
Var(Ic)			0.051509	0.001188
Cov(Ic.Oc)			0.064543	0.001270
Var(Oc)			0.082927	0.000996
$log\ \text{L}(\hat{\theta})$	-1008.2		-1400.9	
Likelihood Ratio	27.99		757.28	
Time	6 iterations. 0h:0m:0s		3 iterations. 0h:0m:16s	

The correlation values obtained

$$\hat{\rho} = \frac{\text{Cov(Ic.Oc)}}{\sqrt{\text{Var(Ic)Var(Oc)}}} = 0.987545$$

Therefore the MXL model is recommended.

The *LR* value on MXL model is 757.28 which is higher compared with that of derived by MNL model (27.99). It means that MXL model gives more reliable conclusion rather than the MNL model. The R^2 of MXL model (0.213) is higher than the MNL model (R^2=0.014) means that the MXL model is better than the MNL model. Based on these two examples, the estimator on MXL model is always higher compared to the estimator on MNL model, $|\theta_{MXL}| > |\theta_{MNL}|$.

Conclusions

Based on data simulation, the conclusions obtained are: MXL can be used to estimate the correlation parameter properly (small bias) and MXL is better than MNL especially for alternatives of the correlation presents among alternatives. Some suggestions for further research are: a) for practitioners who will use the discrete responses in modelling and there is a correlation among alternatives, then the MXL is more appropriate than MNL. If there is no correlation among alternatives then the MNL can be used. b) Statistician can improve the computational method: where the Mixed Logit model requires a long time calculation than MNL model and other estimation methods can also be developed.

Acknowledgement

References

1. Agresti, A. (2002). Categorical Data Analysis. John Wiley & Sons.
2. Bolduc, D. (1999). A practical technique to estimate multinomial probit models in transportation. Transportation Research B, 33, 63–79.

3. Camminatiello, I. and Lucadamo, A. (2010). Estimating multinomial logit model with multicollinear data. Asian Journal of Mathematics and Statistics, 3(2), 93–100.
4. Greene W. (2005). Econometrics Analysis. 5 Editions, Prentice Hall.
5. Horowitz, J.L., and Savin, N.E. (2001). Binary response models : logits, probits and semiparametrics. Journal of Economic Perspectives, 15(4), 43–45.
6. Jain, D.C., Vilcassim, N.J., and Chintagunta, P.K. (1994). A random-coefficients logit brand-choice model applied to panel data. Journal of Business and Economic Statistics, 12(3), 317–328.
7. Lee, L. (1992). On the efficiency of methods of simulated moments and simulated likelihood estimation of discrete choice models. Econometric Theory, 8, 518–552.
8. Lee, L. (1995). Asymptotic bias in simulated maximum likelihood estimation of discrete choice models. Econometric Theory, 11, 437–483.
9. Lovreglio, R.A., Fonzone, A.B., and Dell'Olio, L.C. (2016). A mixed logit model for predicting exit choice during building evacuations. Transportation Research Part A: Policy and Practice, 92, 59–75.
10. McFadden, D. and Train K. (2000). Mixed MNL models for discrete response. Journal of Applied Econometrics, 15(5), 447-470.
11. Merkert, R., and Beck, M. (2017). Value of travel time savings and willingness to pay for regional aviation. Transportation Research Part A: Policy and Practice, 96, 29–42.
12. Revelt, D., and Train, K. (1998). Mixed logit with repeated choices: households choice of appliance efficiency level. Review of Economics and Statistics, 80(4), 647–657.
13. Rodríguez, G. (2001). Generalized Linear Models. Pricenton University.
14. Ruud, P.A. (1996). Approximation and simulation of multinomial probit model: an analysis of covariance matrix estimation. Working Paper, University of California, Berkeley.
15. Train, K. (2003). Discrete Choice Methods with Simulation. Cambrige: UK Press.
16. Train, K. (2016). Mixed logit with a flexible mixing distribution. Journal of Choice Modelling, 19, 40–53.
17. Yang, C.W. (2010). Integrating fuzzy integral with multinomial logit model to evaluate the effects of service quality on travelling airline choice. Information Technology Journal, 9(7), 1449–1457.