[DOI] http://dx.doi.org/10.21487/jrm.2017.05.2.1.79

【연구경향】

An Introduction to Mixed Logit Model

Byung-Jae Lee*

ABSTRACT

The logit models have been quite useful in analyzing the discrete dependent variable. Specifically, multinomial logit(MNL) and conditional logit(CL) models have been the workhorses in analyzing discrete choice data with polychotomous nominal categories. However, there are some limitation in these models which restrict the applicability of these models. Three well-known limitations in these models are the assumptions of homogeneous preference among individuals, independence of irrelevant alternatives(IIA) and no correlation across time and individuals. One of the recent advances in categorical data analysis is mixed logit(MXL) model, or random parameter model. MXL obviates the well-known limitations of standard logit models by allowing heterogeneous preferences, unrestricted substitution patterns, and correlation in unobserved factors over time. The purpose of this paper is to introduce MXL and its applications.

Keywords: Mixed Logit, Multinomial Logit, Conditional Logit

* Yonsei University. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government(NRF-2016S1A3A2925033).

Ⅰ. Introduction

The logit models have been quite useful when the dependent variable is discrete, i.e., when the dependent variable has more than one mutually exclusive category(Agresti 2012; Hilbe 2009; Powers and Xie 2008). More specifically, multinomial logit(MNL) and conditional logit(CL) models have been the workhorses in analyzing discrete choice data with polychotomous nominal categories. Since the methodological innovation by McFadden(1973), plenty of works that applied these models have appeared for analyzing individual choice in many fields, including electoral choice in multi-candidate election setting in political science. The primary advantage of these models is that the researcher could incorporate the alternative-specific variables as well as the variables on the people making choices. However, there are some limitation in these models which restrict the applicability of them. Three well-known limitations in these models are the assumptions of homogeneous preference among individuals, independence of irrelevant alternatives(IIA) and no correlation across time and individuals(Glasgow 2001; Train 2009)

One of the recent advances in categorical data analysis is mixed logit(MXL) model, or random parameter model. MXL obviates the well-known limitations of standard logit models by allowing heterogeneous preferences, unrestricted substitution patterns, and correlation in unobserved factors over time. Like probit, the MXL has been known for many years but has become only very recently available for practical data analysis since the development of simulation techniques.

The main purpose of this paper is to introduce MXL and illustrate its applications with some simple examples. To motivate, random utility, multinomial logit and conditional logit models are introduced in the second and the third section. The fourth section describes the reasoning and estimation procedures for MXL, and the fifth section discusses the applications of MXL. This paper concludes with a brief discussion of the possible research topics that could use MXL models.

Ⅱ. Random Utility and Multinomial Logit Model

Most of the logit models are based on random utility maximization behavior. These models are based on the assumption that individual choice among alternatives can be described by a utility function, which depends on the attributes of the alternatives and the characteristics of the individual(Long and Freese 2014). For example, individual voter selects the candidate or party that yields that highest utility. The decision maker i faces a choice among J alternatives. The utility of person i from alternative j is expressed as follows:

$$U_{ij} = V_{ij} + \epsilon_{ij}$$

where V_{ij} is the systematic portion of utility, ε_{ij} is the unobserved stochastic portion of utility. In turn, V_{ij} is a function of observable attributes of the alternatives, x_{ij} and of the decision maker, z_i. The probability that individual i chooses alternative j is the probability that the utility of alternative j exceeds the utility of all other alternatives.

$$P_{ij} = \Pr(U_{ij} > U_{ik}) \forall K = \Pr(V_{ij} + \epsilon_{ij} > V_{ik} + \epsilon_{ik} \forall K$$

Based on the different assumptions about the distribution of random terms, different choice models follow.

Multinomial logit model(MNL) assumes that the unobserved portion of utility are identically and independently distributed(IID) that follows type I extreme value distribution. It is quite well known that the choice probabilities of MNL have the IID property.

An example of MNL is the choice of transportation mode: car, plane, and train. Here, the satisfaction from each transportation mode, V_J, depends linearly on cost and time.

$$\begin{cases} V_1 = \alpha_1 + \beta x_1 + \gamma z_1 \\ V_2 = \alpha_2 + \beta x_2 + \gamma z_2 \\ V_3 = \alpha_3 + \beta x_3 + \gamma z_3 \end{cases}$$

In this case, the probability of choosing alternative J is increasing with V_j, which in turn depends on cost(x) and time(z). For estimation, we need to transform the satisfaction index, which can take any real value so that it is restricted to the unit interval and can be interpreted as a probability. The multinomial logit model is obtained by applying such a transformation to the $V_{J's}$. For each probability, we get:

$$\begin{cases} P_1 = \dfrac{e^{V_1}}{e^{V_1} + e^{V_2} + e^{V_3}} \\ P_2 = \dfrac{e^{V_2}}{e^{V_1} + e^{V_2} + e^{V_3}} \\ P_3 = \dfrac{e^{V_3}}{e^{V_1} + e^{V_2} + e^{V_3}} \end{cases}$$

Here, the two characteristics of probabilities are satisfied, which are:

- $0 \leq P_j \leq 1 \forall_i = 1, 2, 3$
- $\sum_{j=1}^{3} P_j = 1$

Once fitted, a logit model is useful for predicting the probabilities of choice for an individual(Croissant n.d.).

Ⅲ. Conditional Logit Model and its Limitations

If we make the assumption that the random terms are IID type I extreme value distribution, we obtain the conditional logit model.

$$P_{ij} = \frac{\exp(x_{ij}^T \beta + z_i^T \gamma_j)}{\sum_{j=1}^{m} \exp(x_{ij}^T \beta + z_i^T \gamma_j)}, \ j = 1, \ \ldots, m.$$

where $\mathbf{x}_{ij}$ are alternative-specific regressors and $\mathbf{z}_i$ are case-specific regressors. For identification purpose, one of the γ_j, as in typical multinomial model, need to be set at zero in estimation.

Unlike MNL, Conditional logit(CL) model requires the inclusion of alternative-specific variables, while individual- or case-specific variables are sufficient for multinomial logit model. In case that we have a dataset that includes alternative-specific variables, such as prices and quality measures for all alternatives or evaluation of all candidates by respondents, not just the chosen alternatives, then we can use conditional logit model.[1)]

1) The parameters for conditional logit models are usually estimated with the data in long form(instead of wide form), with one observation providing the data for each alternative for each individual.

As an illustration, here is a conditional logit model using the dataset on mode of transportation(Cameron and Trivedi 2010). The models that incorporate alternative-specific variables require the dataset in which each row of the dataset represents one alternative for one person. So, if we have *N* individuals choosing among *J* alternatives, then the data need to have N×J rows. In the following example, the variable mode distinguishes the modes of transportation(1=train, 2=bus, 3=car), and the variable id indicates different individuals. choice variable indicates which mode is selected(and not selected) by the individual. So, for the first two individuals, the data are as follows:

	id	mode	choice	time	hinc
1	1	Train	Not selected	406	35
2	1	Bus	Not selected	452	35
3	1	Car	Selected	180	35
4	2	Train	Not selected	398	30
5	2	Bus	Not selected	452	30
6	2	Car	Selected	255	30

Here, alternative-specific variable time indicates the amount of travel time for each mode of transportation for each individual. The first row indicates that it would take 406 minutes for individual 1 to travel by train, and the alternative train is not selected. Both individual 1 and 2 selected a car as his/her mode of transportation. The variable hinc, household income, is case-specific variable and the same value are repeated three times for the same individual. Below is the Stata output for a conditional logit model with combination of alternative-specific (time) and individual-specific(hinc) variables. For an easier interpretation, odds ratios are presented instead of coefficients.

〈Table 1〉 Conditional Logit Model for mode of transportation

Variables		Odds Ratio
Time		0.98*** (0.002)
Train		Base category
Bus	income	1.03* (0.017)
	intercept	0.25** (0.144)
Car	income	1.05*** (0.016)
	intercept	0.05*** (0.052)
Log-likelihood		-84.92
N		456

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$, standard errors in parentheses.

In Table 1, the coefficient for time indicates the effect of time on the log odds that an alternative is selected. The coefficient of time is negative, which indicates that the chances of an alternative being selected decrease as the amount of time required to travel using that alternative increases. In other words, the odds ratio for time variable, 0.98, indicates that increasing the time of travel by 1 minute for a given mode of transportation decreases the odds of using that mode by a factor of 0.98(2%), controlling other variables. And a unit increase in household income increases the odds of traveling by bus versus traveling by train by 3.2%, controlling other variables. A unit increase in income increases the odds of traveling by car versus traveling by train by 4.9%, controlling other variables.

Ⅳ. Moving Beyond the Conditional Logit Model[2)]

Although MNL and CL are quite useful in analyzing the choice situations, there are some problems such as IIA property and the assumption of preference homogeneity. MXL considers the coefficients themselves to be random variables and then estimate the multidimensional integrals that define the choice probabilities using Monte Carlo simulation. In other words, the procedure draws the random numbers from the assumed joint probability distributions of the coefficients, and compute the conditional choice probabilities. This process is repeated many times, and the results are averaged. Then, the average is an unbiased estimate of the unconditional choice probabilities. Let me describe these processes in detail.

For the standard logit model, the probabilities are:

$$P_{ij} = \frac{e^{\beta^T x_{ij}}}{\sum_j e^{\beta^T x_{ij}}}$$

Now suppose that the coefficients are individual specific. Then, the probabilities are:

$$P_{ij} = \frac{e^{\beta_i^T x_{ij}}}{\sum_j e^{\beta_i^T x_{ij}}}$$

In this situation, we have two possible estimation strategies. First, estimate the coefficients for each individual in the sample, or second,

2) This section is largely based on chapter 6 of Train 2009 and Croissant n.d.

consider the coefficients as random variables. The first approach is not much interesting, since it would require numerous observations for each individual. And, we are usually interested in the values of the coefficients for a given individual. From the second approach, the random coefficient model is developed.

1. Random Coefficients

The probability that an individual will choose alternative j is:

$$P_{ij}|\beta_i = \frac{e^{\beta_i^T x_{ij}}}{\sum_j e^{\beta_i^T x_{ij}}}$$

This formula describes the probability for individual i to choose alternative j conditional on the vector of individual-specific coefficients β_i. To get the unconditional probability, we just need to compute the average of these conditional probabilities for all values of β_i.

Suppose that, $V_{ij} = \alpha + \beta_i x_{ij}$, i.e., there is only one individual-specific coefficient and that the density of β_i is $f(\beta, \theta)$, where θ is the vector of parameters of the distribution of β. The unconditional probability is calculated then:

$$P_{ij} = E(P_{ij}|\beta_i) = \int_\beta (P_{ij}|\beta) f(\beta, \theta) d\beta$$

which is a one-dimensional integral that can be efficiently estimated by quadrature methods.

If $V_{ij} = \beta_i^T x_{ij}$ where β_i is a vector of length K and $f(\beta, \theta)$ is the joint density of the K individual-specific coefficients, the unconditional probability is:

$$P_{ij} = E(P_{ij}|\beta_i) = \int_{\beta_1}\int_{\beta_2}\dots\int_{\beta_K}(P_{il}|\beta)f(\beta,\theta)d\beta_1 d\beta_2 \dots d\beta_K$$

This is a K-dimensional integral which cannot easily be estimated by quadrature methods. In this kind of situations, the only practically possible method is to use simulations. In simulation, R draws of the parameters are taken from the distribution of β, then the probability is computed for every draw and the unconditional probability, which is the expected value of the conditional probabilities, is estimated by averaging R probabilities.

2. Unrestricted Substitution Patterns

One of the strong assumptions required for MNL is IIA assumption. Although multinomial probit model does not require IIA assumption, the mixed logit model provides more flexible alternatives. Mixed logit does not depend on IIA property or the restrictive substitution patterns of logit. The ratio of mixed logit probabilities, P_{ij}/P_{ik}, depends on all the data, including attributes of alternatives other than j or k. The denominators of the logit formula are inside the integrals and therefore do not cancel out. The percentage change in the probability for one alternative given a change in the m^{th} attribute of another alternative is:

$$\begin{aligned} E_{ijx_{ij}^m} &= -\frac{1}{P_{ij}}\int \beta^m L_{ij}(\beta)L_{ij}(\beta)f(\beta)d\beta \\ &= -\int \beta^m L_{ik}(\beta)\left[\frac{L_{ij}(\beta)}{P_{ij}}\right]f(\beta)d\beta \end{aligned}$$

Here β^m is the m^{th} element of β. The elasticity varies across individuals. A ten-percent reduction for one alternative does not need to imply a ten-percent reduction in other alternatives. Instead, the substitution pattern depends on the specification of the variables and mixing distribution, which can be empirically determined. Note that the percentage change in probability depends on the correlation between $L_{ij}(\beta)$ and $L_{ik}(\beta)$ over different values of β, which is determined by the researcher's specification of variables and mixing distribution.

3. Correlation in unobserved factors over time

Standard logit does not incorporate any unobserved factors that persist over time for a given decision maker. This might be problematic for longitudinal data analysis. By applying a standard logit model to longitudinal data, we are assuming that the unobserved factors that affect a person's choice are new every time the person makes the choice, which is very unrealistic assumption. To take into account both random preference variation and correlation in unobserved factors over time, the utility for respondent i for alternative j at time t is specified as follows:

$$U_{ijt} = \beta_i X_{ijt} + \epsilon_{ijt}$$

where subscript t indicates time dimension. However, we still assume the IID extreme value distribution of ε over time, individual and alternative. The correlation over time over alternatives arises from the common effect of βs, which enter utility in each time period and each alternative.

Assume that βs are $N(\overline{\beta}, \sigma^2)$. Then, the utility equation becomes:

$$U_{ijt} = (\overline{\beta} + \sigma\eta_i)X_{ijt} + \epsilon_{ijt}$$

Here, η is a draw from the standard normal density. The equation becomes

$$U_{ijt} = \overline{\beta}\, X_{ijt} + (\sigma\eta_i X_{ijt} + \epsilon_{ijt})$$

$$U_{ijt} = \overline{\beta}\, X_{ijt} + e_{ijt}$$

where, $e_{ijt} = \sigma\eta_i X_{ijt} + \epsilon_{ijt}$. The second term in the right side of the equation, $\sigma\eta_i X_{ijt}$, is not independent over time or alternatives, while ϵ_{ijt} is independent. Then, the covariance between alternative j and k is,

$$Cov(e_{ijt}, e_{ikt}) = \sigma^2 (X_{ijt} X_{ikt})$$

and the covariance between time t and q is

$$Cov(e_{ijt}, e_{ijq}) = \sigma^2 (X_{ijt} X_{ijq})$$

By specifying the X's properly, we can obtain any pattern of covariance over time and alternatives.

Conditional on β_i, the probability of the sequence of choices by person is simply the product of the logit probability of each individual choice by that individual.

$$L_i(\beta_i) = \prod_t \frac{e^{\beta_i X_{ijt}}}{\sum_j e^{\beta_i X_{ijt}}}$$

since ε_{ijt} is independent over time. Then the unconditional probability of the sequence of choice is simply the integral of this product of logits over the density of β.

$$P_{ij} = \int L_n(\beta) f(\beta|\theta) d\beta$$

4. Simulation

The probabilities for the random parameter logit are integrals with no closed form. Moreover, the degree of integration is the number of random parameters. In practice, these models with random parameters are estimated using simulations, in which the expected value is replaced by an arithmetic mean. More precisely, the computation process is as follows:

1. Begin with an initial hypothesis about the distribution of the random parameters
2. Draw R numbers from this distribution,
3. For each draw β^r, compute the probability $P_{ij}^r = \dfrac{e^{\beta^r x_{ij}}}{\sum_j e^{\beta^r x_{ij}}}$
4. Compute the average of these probabilities: $\overline{P_{ij}} = \sum_{r=1}^{n} \dfrac{P_{ij}}{R}$
5. Compute the log-likelihood for these probabilities
6. Iterate the process until the maximum is reached.

To estimate a model using simulations, one needs to draw pseudo-random numbers from a specified distribution. For this purpose, what we actually need is a function that draws pseudo-random numbers from a uniform distribution between 0 and 1. These numbers are then transformed using the quantile function of the required distribution. For example, suppose we need to draw random numbers from the Gumbell distribution. The cumulative distribution of a Gumbell variable is expressed as $F(x) = e^{e^{-x}}$ the quantile function is obtained by inverting this function:

$$\Rightarrow F^{-1}(x) = -\ln(-\ln x)$$

And R draws from a Gumbell distribution are obtained by computing $F^{-1}(x)$ for R draws from the uniform distribution between 0 and 1. The problem is that there may not be a good coverage of the relevant interval for sufficiently numerous draws to be made. More deterministic methods like Halton draws are more practical and useful. Halton draws are made from Halton sequence.

Halton sequence is the sequence used to generate points in space for numerical methods such as MCMC simulations. An interesting feature of Halton sequences is that although they are actually deterministic, they appear to be random for many purposes. It produces well-spaced draws from the unit interval, and can be used in a way that provides negative correlation between simulated probabilities for individuals. Based on a particular prime number, the sequence is constructed based on finer and finer prime-based divisions of sub-intervals of unit interval. Using a prime number 3 as an example, the process is as follows.

1. Break the interval into equal parts. Then start of sequence is 0, 1/3, 2/3.
2. Break each sub-interval into 3 equal parts. Enter the break points into the sequence in a particular way and enter all lowest break points, then enter all the highest breakpoints. Now the sequence

becomes 1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9

3. Break each sub-sub-interval into 3 equal parts, starting with all the lowest sub-intervals. Enter break points into sequence in particular way, first all the lower breaks, then the upper breaks. Now sequence becomes 1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9, 1/27, 10/27, 19/27, 4/27, 13/27, 22/27, 7/27, 16/27, 25/27, 2/27, 11/27, 20/27, 5/27, 14/27, 23/27, 8/27, 17/27, 26/27
4. Continue this until a sequence of desired length is obtained.

V. Applications

1. Household's choice of electricity supplier

The first example uses the dataset compiled by Huber and Train(2001) on household's choice of electricity supplier. In this experimental data, residential electricity customers were presented with a series of experiments with four alternative electricity suppliers. The following alternative-specific variables are included in the dataset.

- price: price in cents per kWh if fixed price, 0 if TOD or seasonal rates
- contract: contract lengths in years
- local: whether a local company(0 or 1)
- wknown: whether a well-known company(0 or 1)
- tod: time of day rates(0, 1)
- seasonal: seasonal rates(0, 1)

The following is the mixed logit results, with contract, local, wknown, tod and seasonal being assumed to be normally distributed random variables. The coefficient of price is fixed. The following estimation is performed using mixlogit command in Stata, which is a user-written program by Hole(2007).

〈Table 2〉 Mixed Logit Model for Household Electricity

	Variables	Coefficients
Mean	price	-0.95*** (0.07)
	contract	-0.27*** (0.05)
	local	2.14*** (0.23)
	wknown	1.55*** (0.18)
	tod	-9.32*** (0.61)
	seasonal	-9.35*** (0.61)
SD	contract	0.39*** (0.04)
	local	1.87*** (1.87)
	wknown	1.24*** (1.24)
	tod	2.47*** (2.47)
	seasonal	2.26*** (2.26)
Log-likelihood	-1105.2832	
N	4,780	

***p 〈 0.01, ** p 〈 0.05, * p 〈 0.10, standard errors in parentheses.

The upper panel in Table 2 shows the mean coefficients and the lower panel estimated standard deviations of the coefficients. All estimates are highly significant, which means that these coefficients really vary across individuals. The mean coefficients for contract, tod, and seasonal are negative, and the coefficients for local and wknown are positive. The signs in the coefficients indicate the direction of the average effects of these variables on the dependent variable. Customers prefer shorter contract length, non-tod and non-seasonal rates, while they prefer local and well-known companies. Because we assume normal distribution for these coefficients, we can get some information that cannot be obtained in standard logit. For example, using the mean coefficients, we can determine the amount that a customer with average coefficients for fixed price and contract length is willing to pay for an extra year of contract length, which is 0.266/0.958 = 0.277. This means that a customer with mean coefficients for price and contract length is willing to pay 0.277 cents per kWh extra to have a contract that is one year shorter.

2. Train Traveling

The second example is to use the Train data, which is included in R package “mlogit.” The data show the stated preferences for train traveling and were collected in Netherlands in 1987. The unit of observations is individual and the number of observations is 2,929. The dependent variable is either choice 1 or choice 2. The other variables in the dataset are as follows:

- Price: fare of the train(1 or 2)
- Time: travel time in minutes
- Comfort: comfort level(1, 2, or 3 in decreasing order)
- Change: number of changes(0, 1, 2, 3, or 4)

In this application, two mixed logit models, correlated(MXLC) and uncorrelated models(MXLU), are estimated. Time, change, and comfort variables are assumed to be randomly distributed. Censored normal and normal distribution are used for time and change, respectively, while log normal distribution for comfort are used for random draws for each parameter. For comparison, multinomial logit results are also included.

〈Table 3〉 Multinomial Logit, Correlated Mixed Logit and Uncorrelated Mixed Logit Models for Train

	MNL	MXLC	MXLU
intercept	-0.030 (0.04)	-0.032 (0.048)	-0.035*** (0.041)
Price	0.001*** (0.00)	0.001*** (0.000)	0.002*** (0.000)
time	0.028*** (0.003)	0.027*** (0.001)	0.033*** (0.001)
change	0.326*** (0.060)	0.349*** (0.043)	0.410*** (0.03)
comfort	0.947*** (0.947)	-0.130*** (0.018)	-0.077*** (0.006)
Time.time		0.112** (0.001)	
Time.change		-0.200*** (0.061)	
Time.comfort		0.152*** (0.002)	
Change.change		1.607*** (0.059)	
Change.comfort		2.060*** (0.007)	
Comfort.comfort		1.451*** (0.006)	
Sd.time			0.111*** (0.000)
Sd.change			2.189*** (0.03)
Sd.comfort			3.428*** (0.004)
Log-Likelihood	-1723.8	-1556.2	-1589.6
McFadden's R^2	0.151	0.233	0.217

***p 〈 0.01, ** p 〈 0.05, * p 〈 0.10, standard errors in parentheses.

Table 3 shows the results of multinomial logit (MNL), MXLC (correlated mixed logit), and MXLU (uncorrelated mixed logit). In terms of signs of the coefficients, the notable difference is in the sign of comfort variable. In MNL, it does not provide much information other than the positive sign. The signs of the coefficients for comfort variable in two mixed logit models are negative, which is the opposite to the one in MNL. The standard deviation of the coefficient for comfort shows large variation of the effect of comfort on the choice. The distribution in Table 4 shows that the mean is much larger than the median, which indicates that some respondents give very high weights on the comfort, although most people do not.

〈Table 4〉 Distribution of Mean Coefficients for MXLC and MXLU

	time		change		comfort	
	MXLC	MXLU	MXLC	MXLU	MXLC	MLXU
Min	0	0	-	-∞	0	0
1st quartile	0.00	0.000	-0.743	-1.070	0.160	0.092
Median	0.027	0.033	0.349	0.406	0.878	0.926
Mean	0.060	0.063	0.349	0.406	21.238	329.307
3rd quartile	0.103	0.108	1.441	1.882	4.819	9.345
Max	∞	∞	∞	∞	∞	∞

3. Latent Class Mixed Logit Model

So far, we have assumed that the distribution of the coefficients in the model is continuous. However, the coefficients could be discrete and we need more elaborate model in these situations. In latent class model, each respondent is assumed to belong to a class q, in which preferences vary across individuals, but not within classes. In this case, the probability of a particular sequence of choices is given by:

$$S_i = \sum_{q=1}^{Q} H_{iq} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(x_{ijt}^T)\beta_q}{\sum_{j=1}^{J} \exp(x_{ijt}^T)\beta_q} \right]^{y_{ijt}}$$

And the probability of belonging to class q, i.e, H_{iq} in the above equation, is usually specified as:

$$H_{iq} = \frac{\exp(Z_i^T \gamma_q)}{\sum_{q=1}^{Q} \exp(Z_i^T \gamma_q)}$$

where γ_Q is set to zero. The log-likelihood for this model is

$$S_i = \sum_{t=1}^{I} \ln \left\{ \sum_{q}^{Q} H_{iq} \prod_{t=1}^{T} \prod_{j=1}^{J} \left[\frac{\exp(x_{ijt}^T)\beta_q}{\sum_{j=1}^{J} \exp(x_{ijt}^T)\beta_q} \right]^{y_{ijt}} \right\}$$

This can be maximized directly using standard methods, or indirectly using the EM algorithm. Following is an example of the three classes latent model. For the estimation, the same dataset as in the first example, the choice of household's electricity supplier, is used. For the adequate number of classes, the standard methods of model comparison, such as Akaike Information Criteria(AIC) or Bayesian Information Criteria(BIC), can be used. Here, three latent classes model are specified for estimation. The interpretation of the results is same as in the first example, although the coefficients are estimated only for the people in the same class.

〈Table 5〉 Latent Class Mixed Logit Model for Household Electricity

	Variables	Coefficients
Choice1	Price	-0.803*** (0.118)
	Contract	-0.507*** (0.059)
	Local	0.499*** (0.205)
	Wknown	0.357** (0.182)
	Tod	-5.992*** (0.932)
	seasonal	-6.647*** (0.993)
Choice2	Price	-0.211** (0.085)
	Contract	0.023 (0.030)
	Local	3.087*** (0.250)
	Wknown	2.308*** (0.234)
	Tod	-1.878** (0.735)
	seasonal	-1.965** (0.764)
Choice3	Price	-1.140*** (0.141)
	Contract	-0.230*** (0.065)
	Local	1.675*** (0.336)
	Wknown	1.645*** (0.262)
	Tod	-12.530*** (1.387)
	seasonal	-11.750*** (1.138)
Share1		-0.279 (0.291)
Share2		0.000 (0.286)

***p 〈 0.01, ** p 〈 0.05, * p 〈 0.10, standard errors in parentheses.

Table 5 shows that the weights for each variable vary across the classes, although there are also variations within classes. Choice2 has larger mean coefficients for tod and seasonal, while choice2 has larger coefficients for local and wknown and the negative sign of the coefficient for contract. What this table shows is that there are approximately three groups which weight alternative-specific variables differently.

Ⅵ. Discussion

The MXL has many advantages over other choice models. It enables us to incorporate the possibility of preference heterogeneity among individuals, to relax the IIA assumption and to make the incorporation of correlated preferences across time possible. This model has many potential venues of application in political science. For example, it nicely matches the spatial model that assumes random individual weights that were distributed independently of preferences, which can be modeled via MXL. As Mebane et al. illustrate, MXL latent class model can be applied to model "issue public"(Mebane et al. 2014). According to them, issue public that Converse describes means group of voters who vote on the basis of a single or small set of issues and that this issue or set varies among individuals or groups. The proposition of issue public implies different fixed or mean values for the weights for different groups of voters, which can be exactly modeled via MXL latent class models.

The existing statistical methods such as multinomial logit or conditional logit has some limitations to deal with this kind of preference heterogeneity. Although the models of voting behavior existed for decades, little is understood about the heterogeneity in the impact of issue positions, demographic variables, and other factors on vote choices. Both MNP and MXL can be specified to explore random preference variation. However, MNP is limited in the number of random coefficients that can be estimated, and these random coefficients should be normally distributed. MXL can include any number of random coefficients, and these coefficients can follow any distributions. In this sense, MXL could provide useful tools by which we can probe many unanswered questions in voting behavior in particular, natural and social science in general.

References

Agresti, Alan. 2012. *Categorical Data Analysis*. 3rd ed. New York: John Wiley & Sons.

Bhat, Chandra R. 2001. "Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model." *Transportation Research Part B: Methodological* 35(7): 677-693.

Croissant, Yves. n.d. "Estimation of Multinomial Logit in R: The mlogit packages."

Cameron, A. Colin and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata*. Rev. ed. College Station, TX: Stata Press.

Glasgow, Garrett. 2001. "Mixed Logit Models for Multiparty Elections." *Political Analysis* 9(2): 116-136.

Greene, H. William H. and David Hensher. 2003. "A Latent Model for Discrete Choice Analysis: Contrasts with Mixed Logit." *Transportation Research Part B: Methodological* 37(8): 681-698

Hilbe, Joseph. H. 2009. *Logistic Regression Models*. Boca Raton, FL: CRC Press.

Hole, Arne Risa. 2007. "Fitting Mixed Logit Models using Maximum Simulated Likelihood." *The Stata Journal* 7(3): 388-401.

Huber, Joel and Kenneth Train. 2001. "On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths." *Marketing Letters* 12(3): 259-269.

Jaeger, T. Florian. "Categorical Data Analysis: Away from ANOVAs(transform or not) and towards logit mixed models." *Journal of Memory and Language* 59(4): 434-446.

Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. 3rd ed. College Station, TX: Stata Press.

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." in *Frontiers in Econometrics*, ed. Paul Zarembka. New York: Academic Press.

McFadden, Daniel and Kenneth Train. 2000. "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics* 15(5): 447-470.

Mebane, Walter, Jr., John E. Jackson, and Jonathan Wall. 2014. "Preference Heterogeneities in Models of Electoral Behavior." Paper presented at the 2014 Annual Meeting of the Midwest Political Science Association.

Powers, Daniel A. and Yu Xie. *Statistical Methods for Categorical Data Analysis*. 2nd ed. San Diego: Academic Press.

Revelt, David and Kenneth Train. 1998. "Mixed Logit with Repeated Choices: Households choices of Appliances Efficiency Level." *Review of Economics and Statistics* 80(4): 647-657.

Train, Kenneth. 1988. "Recreation Demands Models with Taste Variation." *Land Economics* 74(2): 230-239.

Train, Kenneth. 1999. "Halton Sequences for Mixed Logit." Unpublished manuscript.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation*. 2nd ed. New York: Cambridge University Press.

Train, Kenneth. 2016. "Mixed Logit with a Flexible Mixing Distribution." *The Journal of Choice Modelling* 19: 40-53.

혼합로짓(mixed logit)이란 무엇인가?

이병재*

논문요약

사회과학에서 널리 쓰이는 범주형 데이터, 특히 범주형 종속변수의 분석에 있어 로짓모델은 지난 수 십년간 매우 널리 사용되어왔다. 특히, 다항로짓(multinomial logit)과 조건부로짓(conditional logit)은 종속변수가 3항이상의 명목형 범주로 이루어져 있는 경우 매우 유용한 것이 사실이다. 하지만 이 모델들은 널리 알려진 한계가 있다. 첫째, 개인들 간의 선호도의 동질성의 가정, 둘째, 무관한 대안으로 부터의 독립성(independence of irrelevant alternatives)의 가정, 셋째, 시간별, 개인별 무상관성(uncorrelatedness across time and individual)의 원칙의 가정이 그것이다. 시뮬레이션에 기반한 기법이 확산되기 시작한 이래 범주형 데이터 연구에서 새롭게 활발히 개발, 적용중인 방법중 하나인 혼합로짓(mixed logit)은 개인간 불균질한 선호도(heterogeneous preferences), 교체패턴(substitution patterns)의 자유로운 설정, 시간 및 개인간 상관성 등을 범주형 모델링 과정에 포함시킬 수 있는 가능성을 제시함으로써 전통적으로 사용되던 범주형 데이터 분석의 한계의 극복 뿐만아니라 보다 심화된 데이터 분석을 제시 할 수 있는 방법을 제시한다. 본 논문의 목적은 혼합 로짓을 소개하는 것이다.

주제어: 혼합로짓, 다항로짓, 조건부로짓

투고일: 2017.04.04. 심사일: 2017.04.04. 게재확정일: 2017.04.11.

* 연세대학교