Journal of the Eastern Asia Society for Transportation Studies, Vol. 6, pp. 1724 - 1739, 2005

THE ESTIMATION OF DISCRETE CHOICE MODELS WITH LARGE CHOICE SET

Yu-Chin CHEN
Ph. D. Student
Department of Transportation and Communication Management Science
National Cheng Kung University
1, University Road, Tainan
70101, Taiwan, R.O.C.
Fax: +886-6-2753882
E-mail: yuchin.chen@gsb.uchicago.edu

Liang-Shyong DUANN
Professor
Department of Marketing and Logistics
Southern Taiwan University of Technology
1, Nantai St., Yung-Kang, Tainan
71005, Taiwan, R.O.C.
Fax: +886-6-2753882
E-mail: ls.duann@msa.hinet.net

Wan-Pei HU
Ph. D. Student
Department of Transportation and Communication Management Science
National Cheng Kung University
1, University Road, Tainan
70101, Taiwan, R.O.C.
Fax: +886-6-2753882
E-mail: mandy@csu.edu.tw

Abstract: In this paper we discussed the estimation of choice model with correlation and/or heteroscedasticity across alternatives when the number of alternatives was large. We examined the performance of randomly drawn choice set approach and Poisson regression approach. We designed 6 Monte Carlo experiments. 1000 observations with 30 alternatives each were generated for each experiment. We found that the randomly drawn choice set approach is only applicable to independent and homoscedastic alternatives. When there are correlated factors among alternatives, this approach can not get good fit for the correlated factors even when the number of draw is quite large. We proved theoretically and empirically that Poisson regression model can get the same estimates as some logit type models that showed heteroscedasticity and correlation across alternatives. Poisson regression model can also get estimates close to those of logit kernel model with less computing time.

Key Words: large choice set, Poisson regression, logit kernel model, correlation, heteroscedasticity

1. INTRODUCTION

A computational difficulty may rise when a researcher uses a discrete choice model to deal with a choice situation with large choice set, e.g., residential location choice or shopping destination choice. Some researchers (Hansen, 1987; Woodward, 1992; Friedman, *et al*, 1992) followed a suggestion by McFadden (1978) to solve this problem. They estimated logit models by using smaller choice sets randomly drawn from the full choice sets. Decomposing the structure of large choice sets can indeed reduce computational effort. However, McFadden's proof of consistency for multinomial logit (MNL) model does not carry over to more flexible discrete choice models, e.g., nested logit model, logit kernel (LK) model, or multinomial probit (MNP) model. Furthermore, the estimates of randomly drawn choice set approach are statistically less efficient, because this approach disregards some useful information.

Since similarity and heteroscedasticity among alternatives are generally existed in the multi-alternative choice situation, choice models (e.g., LK and MNP) which can deal with this problem are increasing popular. The term logit kernel is chosen because this model has a logit model at the core and is extended by adding different error terms (Ben-Akiva and Bolduc, 1996). LK model is widely used due to its computing efficiency comparing to MNP model. Sometimes it is under different names, e.g., error components logit (Brownstone and Train, 1999) and mixed logit (McFadden and Train, 2000). The primary motivation for using MNP and LK model in the study of multi-alternative choice situations is to avoid the independence of irrelevant alternatives (IIA) problem. MNP model relaxes the IIA assumption by specifying the distribution of unobserved error terms of utility as multivariate normal with a general covariance matrix. LK model solves this problem by specifying the unobserved error term as a combination of Type I Extreme Value distribution and another distribution such as multivariate normal. Very few studies used LK and MNP models in the choice situations with large choice sets due to computing difficulty.

An alternative approach to deal with the large choice set problem is Poisson regression model (Papke, 1991; Wu, 1999; List, 2001). McCullagh and Nelder (1989, Chapter 6) demonstrated the relationship between MNL model and Poisson regression model. Poisson regression is very useful in dealing with large choice sets since it treats each alternative as an observation. Papers dealt with the problem of correlation or heteroscedasticity across alternatives by Poisson regression model are very limited. King (1989) and Hausman, *et al*, (1984) developed two different generalized event count models which permitted dependent variable to be correlated. However, these models were used to analyze the time effects of panel data and they did not have counterpart discrete choice models. The problem of correlation and heteroscedasticity across alternatives is not discussed in their models.

In this paper we discuss the estimation of choice model with correlation and/or heteroscedasticity across alternatives when the number of alternatives is very large. The performance of randomly drawn choice set approach is discussed when the correct choice models are MNL and LK. We prove both theoretically and empirically that Poisson regression model can be used to substitute some logit type models. Some of the models can deal with correlation and heteroscedasticity across alternatives. We also found that Poisson regression model could get estimates close to those of logit kernel model with less computing time. Data sets used in this paper are generated by simulation experiments.

The results of our experiments show that Poisson regression approach is preferred due to its simpler form and less computing time, and randomly drawn choice set approach has limitations to handle correlation and/or heteroscedasticity. These results may provide an insightful reference to researchers in studying large choice set problems.

2. MODEL AND FORMULATION

In this section we will show the formulation of models tested in this paper. First, we describe the basic assumption of logit kernel model and its estimation procedure. Then we illustrate the formulation of nested logit kernel model that can deal with similarities between alternatives. Finally, we develop three logit type models, i.e., fixed correlation logit (FCL) model which considers the correlation between alternatives, Heteroscedastic logit (HL) model which considers the heteroscedasticity among alternatives, and FC-HL model which considers both correlation and heteroscedasticity.

2.1 Logit Kernel (LK) Model

In a multi-alternative choice situation, if there exist *J* alternatives with *j*=1,…,*J* and *N* travelers with *i*=1,…,*N*, then the utility derived by traveler *i* for alternative *j* is given by

$$U_{ij} = \beta' x_{ij} + \varepsilon_{ij} \tag{1}$$

where β is a vector of unknown parameters, x_{ij} is an attribute vector of alternative *j* viewed by traveler *i*, and ε_{ij} is a random error term.

Thus, the utility of traveler *i* for alternative *j* is composed of a deterministic and a stochastic item. The theory of random utility maximization assumes that a traveler will choose the alternative that will yield her/him the highest expected utility. Multinomial logit (MNL) model and MNP model differ in the assumed error structure for ε_{ij}. If ε_{ij}'s are independently and identically distributed (*iid*) with Type I Extreme Value distribution, then the logit probability of traveler *i* choosing alternative *j* is given by (McFadden, 1974)

$$p_{ij} = \exp\left(\beta' x_{ij}\right) \Big/ \sum_{k=1}^{J} \exp\left(\beta' x_{ik}\right) \tag{2}$$

Probit model assumes the error term ε_{ij}'s are distributed multivariate normal with mean 0 and covariance matrix Σ_ε, and the probit probability is (Daganzo, 1979; Bunch, 1991)

$$\begin{aligned} p_{ij} &= \text{Prob}\left[U_{ij} > U_{ik} \text{ for all } j \neq k\right] \\ &= \int_{\varepsilon_j=-\infty}^{\infty} \int_{\varepsilon_1=-\infty}^{\varepsilon_j} \cdots \int_{\varepsilon_{j-1}=-\infty}^{\varepsilon_j} \int_{\varepsilon_{j+1}=-\infty}^{\varepsilon_j} \cdots \int_{\varepsilon_J=-\infty}^{\varepsilon_j} \phi\left(\varepsilon \mid U, \Sigma_\varepsilon\right) d\varepsilon_1 \cdots d\varepsilon_J \end{aligned} \tag{3}$$

where $\phi\left(\varepsilon \mid U, \Sigma_\varepsilon\right)$ is the normal density function with mean *U* and covariance Σ_ε.

LK model appears in this context as an intermediate alternative somewhere between logit and probit model. The main idea behind LK model is to consider one more error components besides Type I Extreme Valued component (Ben-Akiva and Bolduc, 1996). So the basic model is still kept as a logit, while other probit-like components with a multivariate distribution are added to capture the correlation and heteroscedasticity between alternatives. One can specify the correlation and heteroscedasticity using a factor analytic structure. This is a flexible specification that can include all known error structures, e.g., similarities between alternatives as shown below:

$$U_{ij} = \beta' x_{ij} + \left[\eta_j + e_{ij}\right] \tag{4}$$

where η_j is a random term with zero mean whose distribution over alternatives depends in general on underlying parameters and observed data relating to alternative *j*; and e_{ij} is a random term, with zero mean and is *iid* over alternatives, that does not depend on underlying parameters or data.

Let each element of e_{ij} be independent and identical Extreme Value distribution as that of standard logit model and with the value of η_j given, the resulting conditional choice probability is

$$\pi_{ij}\left(\eta\right) = \exp\left(\beta' x_{ij} + \eta_j\right) \Big/ \sum_{k=1}^{J} \exp\left(\beta' x_{ik} + \eta_k\right) \tag{5}$$

Since η_j is not given, the (unconditional) choice probability is the above logit formula integrated over all values of η_j weighted by the normal density of η_j, $\phi(\eta_j \mid \Omega)$.

$$p_{ij} = \int \frac{\exp(\beta' x_{ij} + \eta_j)}{\sum_{k=1}^{J} \exp(\beta' x_{ik} + \eta_k)} \phi(\eta_j \mid \Omega) d\eta = \int \pi_{ij}(\eta) \phi(\eta_j \mid \Omega) d\eta \quad (6)$$

where Ω is the distribution's fixed parameter vector.

The choice probability of MNP model and LK model can not be calculated exactly because integrals that do not have closed forms are involved. These integrals can only be approximated through simulation. For the estimation of MNP model, the use of a maximum simulated likelihood framework combined with a Gewekw–Hajivassiliou-Keane (GHK) choice probability simulator is suggested in the literature (Boulduc, 1999; Geweke, *et al*, 1994). The computing time of MNP model is tremendous when the number of alternatives is large. The simulation process of LK model is much simpler. For given values of the parameter vector Ω, a value of η is randomly drawn from its distribution for each alternative. Then $\pi_{ij}(\eta)$ in equation (6) is calculated through normal MNL estimation procedure. This process is repeated many times. The expectation value of the probability is the average of $\pi_{ij}(\eta)$'s.

$$Sp_{ij} = (1/R) \sum_{r=1}^{R} \pi_{ij}(\eta^r) \quad (7)$$

where *R* is the number of replications, η^r is the result of rth draw, and Sp_{ij} is the simulated probability of traveler *i* choosing alternative *j*.

Let $d_{ij} = 1$, if individual *i* chooses alternative *j* and $d_{ij} = 0$, otherwise. The simulated log-likelihood of LK model can be written as

$$\ln SL = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \ln(Sp_{ij}) \quad (8)$$

We now deal with the problem of similarity between alternatives. The problem of heteroscedasticity among alternatives can be formulated with similar method. It is not reported here for brevity. We adopted the nesting and cross-nesting error structures suggested by Walker, *et al*, (2003) to relax the assumption of IIA.

The nested LK model is specified as follows:

$$U_i = X_i \beta + FT\eta_i + e_i \quad (9)$$

where η_i is an $(M \times 1)$ vector of *iid* random variables, M is the number of nests; F is a $(J \times M)$ matrix of factor loading, $f_{jm} = 1$ if alternative *j* is a member of nest m, and $f_{jm} = 0$ otherwise; T is an $(M \times M)$ diagonal, which contains the standard deviation of each factor.

Nesting LK model assumes alternatives in the same group have common unobserved attributes, i.e., factors. Take a 5-alternative case for example, if the first 2 alternatives belong to one group and the remaining 3 alternatives belong to another group, then the utility and the nesting structure can be written as:

$$\begin{cases} U_{i1} = \beta' x_{i1} + \sigma_1 \eta_{i1} & & + \varepsilon_{i1} \\ U_{i2} = \beta' x_{i2} + \sigma_1 \eta_{i1} & & + \varepsilon_{i2} \\ U_{i3} = \beta' x_{i3} & + \sigma_2 \eta_{i2} & + \varepsilon_{i3} \\ U_{i4} = \beta' x_{i4} & + \sigma_2 \eta_{i2} & + \varepsilon_{i4} \\ U_{i5} = \beta' x_{i5} & + \sigma_2 \eta_{i2} & + \varepsilon_{i5} \end{cases} \quad (10)$$

where $F^T = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix}$ and $T = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}$

If we assume cross-nest structure exists and alternative 3 belongs to both nests, then the factor loading matrix becomes:

$$F^T = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{bmatrix} \tag{11}$$

The utility function of alternative 3 in the above cross-nesting structure becomes:

$$U_{i3} = \beta' x_{i3} + \sigma_1 \eta_{i1} + \sigma_2 \eta_{i2} + \varepsilon_{i3} \tag{12}$$

Under the factor analytic specification, the general form of utility function in LK model can be written as:

$$U_{ij} = \beta' x_{ij} + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m \cdot \eta_{im} + \varepsilon_{ij} \tag{13}$$

There are some LK models specified parameter β's as random variables to capture the unobserved individual heterogeneity, e.g., random parameters logit (Carlsson, 2003) and random coefficient logit (Jain, *et al*, 1994). In this paper we focus on the error structure to simplify the problem.

2.2 Fixed Correlation Logit (FCL) Model

Assuming that the unobserved factors in Eq.(13) are fixed, the utility function and the choice probability of traveler *i* choosing alternative *j* can be expressed as:

$$U_{ij} = \beta' x_{ij} + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m + \varepsilon_{ij} \tag{14}$$

$$p_{ij} = \frac{\exp\left(\beta' x_{ij} + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m\right)}{\sum_{k=1}^{J} \exp\left(\beta' x_{ik} + \sum_{m=1}^{M} f_{km} \cdot \sigma_m\right)} \tag{15}$$

The above model uses fixed unobserved factors to show the similarities between alternatives. It is called “fixed correlation logit (FCL) model” hereafter. It can easily been seen that FCL model is only a simplified special case of general LK model. However, this model does have some merits. First, if similarities between alternatives are due to omission of common variables, it can capture this effect. Second, the fixed effects estimators are consistent regardless of whether the effects are correlated with the explanatory variables or not. The random effects estimator, however, will be inconsistent if the effects are correlated with the explanatory variables (Wooldridge, 2002). Third, the estimation of FCL model can be performed by many econometric software packages, e.g., Alogit (Hague Consulting Group Inc., 1993). In the case of large choice set or complicated nest structure, FCL model may be the only feasible solution. More complicated models are generally very difficult to estimate.

2.3 Heteroscedasticity Logit (HL) Model

We can rewrite Eq.(2) as (McFadden, 1974):

$$p_{ij} = \exp\left[\lambda \cdot \left(\beta' x_{ij}\right)\right] \Big/ \sum_{k=1}^{J} \exp\left[\lambda \cdot \left(\beta' x_{ik}\right)\right] \tag{16}$$

where λ is a scale parameter related to the variance σ_e^2 of error term.

However, λ is unidentifiable and can be fixed to any value without affecting the model. It is usually set equal to unity. The relation between the scale parameter and the variance is

$$\lambda = \pi / \sqrt{6}\sigma_e \tag{17}$$

Since the scale parameter λ is the inverse of the standard deviation of error term, the heteroscedasticity among alternatives can be presented by specifying different scale parameter for each alternative (up to identification). Thus the choice probability of heteroscedastic logit (HL) model can be rewritten as:

$$p_{ij} = \exp\left[\lambda_j \cdot \left(\beta' x_{ij}\right)\right] \Big/ \sum_{k=1}^{J} \exp\left[\lambda_k \cdot \left(\beta' x_{ik}\right)\right] \tag{18}$$

The formulation above was proposed by Munizaga, *et al.* (1997) to model discrete choices between heteroscedastic alternatives. A nested logit model with a unique inclusive value parameter for each alternative (Fig. 1.) is equivalent to the above specification (Louviere, *et al*, 2000, p.142-143). Since HL model has an identical formulation with nested logit, it can be estimated with many econometric software packages, e.g., Alogit.

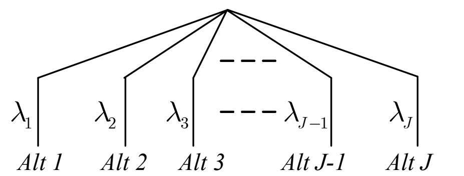

Figure 1. The nest structure of heteroscedastic logit model

It is easy to include the fixed correlation into the HL model. The resulting combination model (FC-HL) can be written as:

$$p_{ij} = \frac{\exp\left[\lambda_j \cdot \left(\beta' x_{ij} + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m\right)\right]}{\sum_{k=1}^{J} \exp\left[\lambda_k \cdot \left(\beta' x_{ik} + \sum_{m=1}^{M} f_{km} \cdot \sigma_m\right)\right]} \tag{19}$$

The estimation of FC-HL model is similar to that of HL model.

3. APPROACHES DEALING WITH THE LARGE CHOICE SET PROBLEM

3.1 Randomly Drawn Choice Set (RDCS) Approach

The computation burden of discrete choice model is tremendous when the number of alternatives is very large. There are several methods proposed to deal with this problem in the literature. One method is called the randomly drawn choice set (RDCS) approach which is constructed from decomposing the structure of full choice sets. This approach estimates the choice model using a randomly drawn choice subset from the full choice set.

McFadden (1978) showed that RDCS approach gave the choice model unbiased estimates. Traveler *i*'s probability of choosing alternative j' given the subset of alternatives *D*, where *D* is randomly drawn from her/his full choice set, can be written as:

$$p_i\left(j' \mid D\right) = \frac{p\left(D \mid j'\right) \cdot \exp\left(\beta' x_{ij}\right)}{\sum_{j \in D} p\left(D \mid j\right) \cdot \exp\left(\beta' x_{ij}\right)} \tag{20}$$

where $p(D \mid j)$ is the probability of drawing the subset D given the alternative j being chosen.

This approach works as follows. Each sampled traveler's choice set is generated by her/his chosen alternative plus randomly drawn subsets of all the other alternatives. For example, if a traveler faced with 20 alternatives, her/his choice set is composed of her/his chosen alternative plus 5 alternatives randomly drawn from the remaining 19 alternatives. One random drawn is performed for each traveler. In this case, $p(D \mid j) = 1/C_5^{20-1}$ for all j, where C_5^{20-1} is the number of combinations of drawing 5 alternatives from (20-1) alternatives.

If each alternative is randomly drawn, $p(D \mid j') = p(D \mid j)$ for all j. Thus, $p(D \mid j')$ will be cancelled out and RDCS approach will get the same estimates as that of the full choice set approach. McFadden (1978) showed that randomly drawn approach was unbiased as long as the drawing rule satisfies the so-called positive and uniform conditioning property. Positive conditioning property requires that the probability of drawing is always non-zero, regardless of which alternative in the full choice set is actually chosen. Uniform conditioning property, a sufficient condition, is more tractable. It requires that the probability of drawing is the same regardless of which alternative is actually chosen. The RDCS approach can be easily applied to the models mentioned in section 2. For example, the choice probability of the FC-HL model by the RDCS approach can be written as:

$$
\begin{aligned}
p_i(j \mid D) &= \frac{p(D \mid j) \cdot \exp\left[\lambda_j\left(\beta' x_{ij} + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m\right)\right]}{\sum_{k \in D} p(D \mid k) \cdot \exp\left[\lambda_k\left(\beta' x_{ik} + \sum_{m=1}^{M} f_{km} \cdot \sigma_m\right)\right]} \\
&\approx \frac{\exp\left[\lambda_j\left(\beta' x_{ij} + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m\right)\right]}{\sum_{k \in D} \exp\left[\lambda_k\left(\beta' x_{ik} + \sum_{m=1}^{M} f_{km} \cdot \sigma_m\right)\right]}
\end{aligned}
\tag{21}
$$

Quigley (1976) developed a method to treat similarities between alternatives. This method can also reduce the number of alternatives in the estimation process. His method is to aggregate all similar alternatives in the same class and define "average" variables to present similar alternatives and include logarithm of the number of alternatives as an explanatory variable in the utility function of that class. In this way, the number of alternatives can be greatly reduced. This method represents the number of similar alternatives as "a proxy for the information available to choice-makers about alternatives". Clearly, this method can only roughly approximate relationships between alternatives, because it replaces each alternative's specific characteristics with "average" attributes. An addition drawback of this method is that cross-similarities among alternatives cannot be allowed. An arbitrary method used by some researchers has artificially reduced the number of alternatives by eliminating alternatives with no or less observations.

3.2 Poisson Regression Approach

McCullagh and Nelder (1989, Chapter 6) showed that the likelihood function of a multinomial logit model is equivalent to that of a Poisson log-linear model. In this section we will prove that the log-likelihood function of FC-HL model is the same as that of its counterpart Poisson regression model. The proof of FC model and HL model is straightforward so is omitted here for brevity.

Let us assume that there are only generic variables in the utility function, i.e., no individual specific variables. In this case, $x_{ij} = x_j$. The log-likelihood function of FC-HL model can be written as:

$$\log L = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \ln p_{ij} = n_i \sum_{j=1}^{J} \ln p_j \tag{22}$$

where

$$p_j = \frac{\exp\left[\lambda_j \left(\beta' x_{ij} + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m\right)\right]}{\sum_{k=1}^{J} \exp\left[\lambda_k \left(\beta' x_{ik} + \sum_{m=1}^{M} f_{km} \cdot \sigma_m\right)\right]} = \frac{\exp\left[\lambda_j \left(\beta' x_{ij} + h_j\right)\right]}{\sum_{k=1}^{J} \exp\left[\lambda_k \left(\beta' x_{ik} + h_k\right)\right]} = \frac{\exp\left[U_j\right]}{\sum_{k=1}^{J} \exp\left[U_k\right]} \tag{23}$$

n_j is the number of travelers choosing alternative j.

h_j is the function of unobserved factors for alternative j.

The basic Poisson regression model assumes that *n* give *x* has a Poisson distribution (Maddala, 1983, p.51). The probability density function of *n* given *x* under the Poisson assumption is determined by the expected mean $E(n \mid x) = u = u(x)$. Hence,

$$f(n \mid x) = \left(u^n \cdot e^{-u}\right) / n! \tag{24}$$

The most common mean function in application is the exponential form and we assume that the expected mean is a function of the utility of alternative j:

$$E(n_j) = u_j = \exp\left[\alpha + \lambda_j \left(\beta' x_j + h_j\right)\right] \tag{25}$$

$$\Rightarrow \ln(u_j) = \alpha + \lambda_j \left(\beta' x_j + h_j\right) \tag{26}$$

where α is the Poisson regression constant.

The log-likelihood function can be written as:

$$\begin{aligned} \ln L_p &= \ln \prod_{j=1}^{J} \left(u_j^{n_j} \cdot e^{-u_j} / n_j!\right) = \sum_{j=1}^{J} \left[-u_j + n_j \cdot \ln u_j - \ln n_j!\right] \\ &= \sum_{j=1}^{J} \left[-\exp\left[\alpha + \lambda_j \left(\beta' x_j + h_j\right)\right] + n_j \left[\alpha + \lambda_j \left(\beta' x_j + h_j\right)\right] - \ln n_j!\right] \end{aligned} \tag{27}$$

To Maximize Eq.(26) as a function of α, we differentiate it with respect to α and set the result equal to zero. Therefore,

$$\sum_{j=1}^{J} \exp\left[\alpha + \lambda_j \left(\beta' x_j + h_j\right)\right] = \sum_{j=1}^{J} n_j = N \tag{28}$$

$$\Rightarrow \exp(\alpha) = N \Big/ \sum_{j=1}^{J} \exp\left[U_j\right] \tag{29}$$

where $U_j = \lambda_j \left(\beta' x_j + h_j\right)$

Thus, the log-likelihood function becomes:

$$\begin{aligned} \ln L_p &= \sum_J \left[-N \cdot \exp\left[U_j\right] \Big/ \sum_J \exp\left[U_j\right] + n_j \left(\ln N - \ln \sum_J \exp\left[U_j\right] + U_j\right) - \ln n_j!\right] \\ &= \sum_J n_j \left[U_j\right] - \sum_J n_j \ln \sum_J \exp\left[U_j\right] + \text{const}_1 = \sum_{j=1}^{J} n_j \ln p_j + \text{const}_1 \end{aligned} \tag{30}$$

The first term in the above expression is identical to the log-likelihood of FC-HL model, and the remaining terms are constant.

We now incorporate heterogeneity into Poisson regression model by sample segmentation and group individual specific variables. In this case, $x_{ij} = x_{jg}$, $g = 1, ..., G$ where vector

$x_{jg} = [x_g, x_j]'$; x_g is a vector of specific variables in group *g* and x_j is a vector of generic variables of alternative *j*. Assuming the expected mean of the Poisson distribution is:

$$E(n_{jg}) = u_{jg} = \exp\left[\alpha_g + \lambda_j \left(\beta' x_{jg} + h_j\right)\right] \tag{31}$$

where α is the vector of Poisson constant with respect to group *g*.

Thus, the log-likelihood function of Poisson regression model can be written as:

$$\begin{aligned}\ln L_p &= \ln \prod_{j=1}^{J} \left(u_j^{n_j} \cdot e^{-u_j} / n_j ! \right) = \sum_{j=1}^{J} \left[-u_j + n_j \cdot \ln u_j - \ln n_j ! \right] \\ &= \sum_{j=1}^{J} \left[-\exp\left[\alpha_g d_g + \lambda_j \left(\beta' x_{jg} + h_j\right)\right] + n_j \left[\alpha_g d_g + \lambda_j \left(\beta' x_{jg} + h_j\right)\right] - \ln n_j ! \right]\end{aligned} \tag{32}$$

We differentiate Eq.(32) with respect to α_g 's, then

$$\sum_{j=1}^{J} \exp\left[\alpha_g + \lambda_j \left(\beta' x_{jg} + h_j\right)\right] = \sum_{j=1}^{J} n_{jg} = n_g \tag{33}$$

This implies $\alpha_g = n_g \Big/ \sum_{j=1}^{J} \exp\left[\alpha_g + \lambda_j \left(\beta' x_{jg} + h_j\right)\right]$. Substituting α_g in the original log-likelihood function, we can get

$$\ln L_p = \sum_G \sum_J \begin{bmatrix} -n_g \cdot \exp\left[U_{jg}\right] \Big/ \sum_J \exp\left[U_{jg}\right] + \\ n_{jg} \left(\ln n_g - \ln \sum_J \exp\left[U_{jg}\right] + U_{jg} \right) - \ln n_{jg} ! \end{bmatrix} = \sum_G \sum_J n_{jg} \ln p_{jg} + \text{const}_2 \tag{34}$$

Again, the FC-HL model is likelihood-equivalent to the Poisson regression model. In the extreme case, each group can contain only one individual so the above result is also applicable to the individual data.

The above proof shows that the estimation result of certain Poisson model will be the same as its counterpart logit type model. In practical application, the model form used in Poisson regression model is the numerator of Eq.(24). The omission of constant denominator term will not affect the result. In the case of grouped aggregate data when all u_j 's are not zero, log-linear model of Eq.(26) and maximum likelihood estimation method can be used to calibrate the model. This simple form is very easy to calibrate. In the case of disaggregate data, Eq.(25) and least squared method will have to be used to calibrate the model. The computing advantage of Poisson regression will be not be as obvious in this situation. Large choice set causes no problem for Poisson regression model since each alternative is treated as an observation as shown in Eq.(24).

4. MONTE CARLO EXPERIMENTS

In this section we examine the performance of RDCS approach when the correct choice models are multinomial logit and logit kernel. We also compare the performance of Poisson regression models to some logit type models, i.e. FCL model, HL model, FC-HL model, and LK model. We performed 6 Monte Carlo experiments for the above purposes. Each experiment has 1000 observations and each observation has 30 alternatives. The chosen alternative of each observation is decided by utility maximizing principle.

Choosing the number of sample size in Monte Carlo experiments is an important issue. The general principal is that the sample size should be large enough to capture error structures used in the experiments and to describe the complex utility function correctly in terms of statistical significance. However, large sample size will generally decrease the computing efficiency. The sample size of 1000 which satisfies the above principle is determined from several trials not reported here. The performance order of the methods discussed in this paper is not affected by the sample size. All estimations and computations were carried out by GAUSS programming language (Aptech Systems Inc., 2002). Normal distribution errors were generated by the random number generator RNDN from GAUSS Library.

We use grouped aggregate data in the following experiments (with the exception of experiment 2 which uses disaggregate data). We assume the choice situation is residential location choice of CBD worker faced with the same network (travel time and cost) and zonal (rent) data. Travel cost and rent are put together into one cost variable. More complicated choice situation, e.g., many to many location choice, and specification, e.g., zonal specific variable, can be easily extended. The zonal size will also have no effect on the result as long as the sample size is large enough to guarantee that the dependent variable is not zero.

Consider the following basic utility function for the Monte Carlo experiments,

$$U_{ij} = \beta_1 x_j^1 + \beta_2 x_j^2 + \varepsilon_{ij} \tag{35}$$

The explanatory variables (x_j^1, x_j^2) in Eq.(35) are generated from an independent normal distribution $N(0,9)$ and they are also independent across individuals and alternatives. Each error term ε is generated from independent Type I Extreme Value distribution with scale parameter of one. The parameter values used to generate the data are $\beta_1 = 0.25$, $\beta_2 = 0.5$. This utility function is used in Experiment 1.

4.1 Experiment 1: MNL (Full Choice Set vs. RDCS) vs. Poisson Regression

This experiment has two purposes. First purpose is to examine the performance of RDCS approach when the correct choice model is multinomial logit. Second purpose is to examine the performance of Poisson regression model comparing to that of multinomial logit model. The utility function of Eq.(34) is used in this experiment. The error terms of all alternatives are assumed to be independent and homoscedastic.

Table 1. The Coefficients (t-values) of Models in Experiment 1

Parameter (true value)	MNL (full choice set)	MNL (RDCS)			Poisson regression
β_1=0.25	0.262(16.59)	0.270(12.22)	0.262(15.12)	0.260(15.85)	0.262(17.01)
β_2=0.5	0.488(38.65)	0.501(25.15)	0.491(32.56)	0.484(35.63)	0.488(44.11)
Constant	—	—	—	—	2.705(51.22)
# of alternatives	30	5	10	15	30

Table 1 shows that the RDCS approach performs well even when the number of randomly drawn alternatives is only 4. Since MNL model is very easy to calibrate, the computing advantage of RDCS approach is quite small. Poisson regression has the same estimates as those of MNL model but with less computing time. The t-values of Poisson regression model is different from those of MNL model due to the existence of extra constant term.

4.2 Experiment 2: Logit Kernel (LK) Models (Full Choice Set vs. RDCS)

The purpose of this experiment is to examine the performance of RDCS approach when the correct choice model is logit kernel. We use disaggregate data in this experiment. The utility function of this model is

$$U_{ij} = \beta_1 x_{ij}^1 + \beta_2 x_{ij}^2 + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m \cdot \eta_{im} + \varepsilon_{ij} \tag{36}$$

We separate 30 alternatives into three nests (M=3) and assume that each nest has the same correlated unobserved factor, i.e., alternative 1 to 10 with σ_1 factor, alternative 11 to 20 with σ_2 factor, and so on. $f_{jm} = 1$ if alternative j is a member of nest m, and $f_{jm} = 0$ otherwise. The parameter values used to generate the data are $\beta_1 = 0.25$, $\beta_2 = 0.5$, $\sigma_1 = 3$, $\sigma_2 = 2$, $\sigma_3 = 1$. η 's are randomly drawn from N(0, 1).

The estimation procedure is showed in Eq.(7) and (8). It requires many random draws. Brownstone and Train (1999) showed that the bias may be negligible with as few as 125-200 pseudo-random draws to simulate the log-likelihood function of logit kernel model. Bhat (2001) reported that Halton generator with only 125 draws was far superior to the pseudo-random generator with as many as 2000 draws. In this paper, we use 500 Halton draws to reduce the simulation variance of maximum simulated likelihood (MSL) estimator.

Table 2. The Coefficients (t-values) of Logit Kernel Models in Experiment 2

Parameter (true value)	LK (full choice set)	LK (RDCS)				Poisson with Random Effects
β_1=0.25	0.264(18.5)	0.249(11.9)	0.258(15.6)	0.256(15.8)	0.257(17.1)	0.154(28.9)
β_2=0.5	0.520(30.2)	0.473(15.9)	0.485(21.4)	0.506(24.8)	0.506(26.8)	0.294(54.7)
σ_1=3	3.697(6.2)	0.731(1.8)	1.231(4.7)	2.018(6.5)	2.197(6.4)	0.714(3.4)
σ_2=2	2.322(4.8)	0.173(0.3)	0.606(1.5)	1.166(3.6)	1.408(4.3)	0.692(3.2)
σ_3=1	1.237(2.1)	0.542(1.0)	0.855(2.4)	0.244(0.2)	0.597(0.9)	0.418(1.2)
P-Constant	—	—	—	—	—	-3.95(-25.7)
# of alternatives	30	5	10	15	20	30
Sample Size	1000	1000	1000	1000	1000	1000
Time(Min)	12.633	5.684	5.937	6.063	6.948	1.573

The results of LK models are shown in Table 2. The computing time of this model is much longer than that of MNL model. The use of RDCS approach can indeed reduce a large portion of the computing time. The coefficients of β_1 and β_2 in the RDCS model are close to those of the full choice set model. However, the estimates of random effect parameters σ_1, σ_2, and σ_3 do not fit well even when the number of random draw is as many as 19 and some of them are insignificant. These results suggest that the RDCS approach may not be suitable when similarities between alternatives exist. The results of Poisson regression model calibrated by least square method is also shown in Table 2. The results showed that Poisson regression model had great computing advantage over logit kernel model. It takes only about one-eighth of the computing time. The ratio of coefficients $\hat{\beta}_1$ and $\hat{\beta}_2$ is very close to the true value 2. However, the absolute value of each coefficient is quite different from their true parameter due to scale difference. If the focus of study is the ratio between parameters, e.g., value of time, Poisson regression model will be a good substitute for logit kernel model due to great computing efficiency.

The purpose of the following three experiments is to demonstrate the relationship between some logit type models (FCL, HL, and FC-HL) and Poisson regression model.

4.3 Experiment 3: Correlation Models (FCL vs. Poisson regression)

The utility function of correlation model is

$$U_{ij} = \beta_1 x_j^1 + \beta_2 x_j^2 + \sum\nolimits_{m=1}^{M} f_{jm} \cdot \sigma_m + \varepsilon_{ij} \tag{37}$$

This is a simplified case of experiment 2. In this experiment, we set all η's to one and use grouped data. The rest are the same as experiment 2. The estimation results are shown in Table 3. The parameter σ_3 is normalized to one to satisfy the identification condition. We can see that Poisson regression model and FCL model have the same estimates.

Table 3. The Coefficients (t-values) of Correlation Models in Experiment 3

Parameter (true value)	FCL	Poisson Regression
β_1=0.25	0.259(14.65)	0.259(15.19)
β_2=0.5	0.495(29.22)	0.495(47.19)
σ_1=3	2.681(12.63)	2.681(19.56)
σ_2=2	1.803(8.11)	1.803(11.21)
σ_3=1	1.0 (—)	1.0 (—)
P-Constant	—	0.2214(1.69)

4.4 Experiment 4: Heteroscedasticity Models (HL vs. Poisson Regression)

The utility function of heteroscedasticity model is

$$U_{ij} = \beta_1 x_j^1 + \beta_2 x_j^2 + \varepsilon_{ij}^* \tag{38}$$

where $\varepsilon_{ij}^* = \varepsilon_{ij} / \lambda_j$; λ is the scale parameter of the error component ε^*.

The heteroscedasticity among alternatives are assumed to be alternative 1 to 10 with scale λ_1, Alternative 11 to 20 with scale λ_2, and so on. The parameter values used to produce the data are $\beta_1 = 0.25$, $\beta_2 = 0.5$, $\lambda_1 = 0.75$, $\lambda_2 = 1.00$, and $\lambda_3 = 1.25$.

The estimation results are shown in Table 4. The parameter λ_2 is constrained to one to satisfy the identification condition. Since the scale is different we have to use the ratio to tell the performance of the model. It can be seen that the ratio of β_1 and β_2 is very close to the 'true' ratio 2 and both models have the same estimates. From Table 3 and 4, we can see that Poisson regression model gets higher t-values in the correlation model but lower t-values in the heteroscedasticity model

Table 4. The Coefficients (t-values) of Heteroscedasticity Models in Experiment 4

Parameter (true value)	HL	Poisson Regression
β_1=0.25	0.228(5.52)	0.228(2.43)
β_2=0.5	0.473(11.33)	0.473(5.76)
λ_1=0.75	0.756(8.99)	0.756(5.70)
λ_2=1.00	1.0 (—)	1.0 (—)
λ_3=1.25	1.065(8.207)	1.065(3.89)
P-Constant	—	1.910(10.69)

4.5 Experiment 5: Correlation and Heteroscedasticity Models (FC-HL vs. Poisson Regression)

The utility function of this model is

$$U_{ij} = \beta_1 x_j^1 + \beta_2 x_j^2 + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m + \varepsilon_{ij}^* \quad (39)$$

The assumption of unobserved correlated factors is the same as those of experiment 3. Moreover, we assume there are heteroscedasticity among alternatives. The heteroscedasticity among alternatives are assumed to be alternative 1 to 15 with scale λ_1, Alternative 16 to 30 with scale λ_2. The parameter values used to generate the data are $\beta_1 = 0.25$, $\beta_2 = 0.5$, $\sigma_1 = 3$, $\sigma_2 = 2$, $\sigma_3 = 1$, $\lambda_1 = 1.25$, and $\lambda_2 = 1.00$.

Table 5. The Coefficients (t-values) of Correlation and Heteroscedasticity combined Models in Experiment 5

Parameter (true value)	FC-HL	Poisson Regression
β_1=0.25	0.158(9.19)	0.158(11.20)
β_2=0.5	0.307(9.73)	0.307(13.98)
σ_1=3	2.530(9.36)	2.530(9.014)
σ_2=2	2.015(8.92)	2.015(7.664)
σ_3=1	1.0 (—)	1.0 (—)
λ_1=1.25	1.356(10.88)	1.356(15.83)
λ_2=1.00	1.0(—)	1.0(—)
P-Constant	—	-0.164(-0.510)

The estimation results are shown in Table 5. The parameters σ_3 and λ_2 are both constrained to one to satisfy the identification condition. It can be seen that the ratio of β_1 and β_2 is very close to the 'true' ratio which is 2 and both models have the same estimates. Poisson regression model gets higher t-values for some parameters but lower t-values for other parameters.

4.6 Experiment 6: Correlation with Random Effect Models (LK vs. Poisson regression)

The purpose of this experiment is to compare the performance of Poisson regression model with general logit kernel model. The utility function of this experiment is

$$U_{ij} = \beta_1 x_j^1 + \beta_2 x_j^2 + \sum_{m=1}^{M} f_{jm} \cdot \sigma_m \cdot \eta_{im} + \varepsilon_{ij} \quad (40)$$

30 alternatives are divided into three groups (*M*=3) and each group has the correlated unobserved factor σ_m. That is, alternative 1 to 10 with factor σ_1, alternative 11 to 20 with factor σ_2, and so on. And η's are *i.i.d.* random normal variables with zero mean and unit variance. The parameter values used to generate the data are $\beta_1 = 0.25$, $\beta_2 = 0.5$, $\sigma_1 = 3$, $\sigma_2 = 2$, $\sigma_3 = 1$.

Table 6 is the estimation results of two random effect models. The estimates of two generic variables are very close to each other and the true parameters (β_1 and β_2). Their t-values are also very close. LK model has more accurate estimates for correlation parameters σ's although the estimates of σ_3 is not significant. The estimates of Poisson regression are less significant and less accurate but still not significantly different from the true parameters. However, Poisson regression model require less computing time. We also extended this experiment from one-group to many-group situation and got similar results. The about results show that Poisson regression model can be used in choice situations with large choice set and complex error structure. Since the estimation of LK model involves tedious trial and error process, the more efficient Poisson regression model can be used as a screen model to find the correct specification. This specification can be used to calibrate the final LK model. This approach will save total computing time.

Table 6. The Coefficients (t-values) of Correlation with Random Effect Models in Experiment 6

Parameter (true value)	LK	Poisson Regression
β_1=0.25	0.254(28.2)	0.254(28.90)
β_2=0.5	0.512(40.8)	0.513(40.81)
σ_1=3	3.122(5.4)	2.448(1.65)
σ_2=2	1.989(1.8)	2.327(1.52)
σ_3=1	0.993(0.5)	1.869(1.06)
P-Constant	—	0.484(0.17)

5. CONCLUSION AND DISCUSSION

We have demonstrated two approaches to deal with the large choice set problem of discrete choice model. The first approach is randomly drawn choice set (RDCS) approach. However, the results of our Monte Carlo simulations show that this approach is only applicable to independent and homoscedastic alternatives. When there are correlated factors among alternatives, this approach can not get good fit for the correlated factors even when the number of draw is quite large.

The other approach is Poisson regression model. We proved theoretically that Poisson regression model could get the same estimates as some logit type models that showed heteroscedasticity and correlation across alternatives. The results of our Monte Carlo experiments showed the same conclusion empirically. We also used Poisson regression model to estimate random effect correlation model. Its estimates are close to those of logit kernel model. We also found that Poisson regression model required less computing time due to its simpler model form. The computing advantage of Poisson regression model is generally dependent on the complex of utility function and number of alternatives.

From our experience, we believe that Poisson regression model is preferred to some logit type model due to simpler model form in the choice situation with large choice set. However, Poisson regression model does have some limitations. First, it is best suited for grouped aggregate data with large choice set and large sample size. For disaggregate data, small choice set, or small sample size, most of its computing advantage will be lost. Second, if the grouped aggregate data is used, the number of groups will put limitation on its specification due to the degree of freedom constraints.

ACKNOWLEDGEMENTS

The financial support of National Science Council, R.O.C. (Grant Number: NSC-91-2211-E-006-089) is greatly appreciated.

REFERENCES

Aptech Systems, Inc., (2002) **GAUSS 4.0.**, Maple Valley, Washington, USA.

Ben-Akiva, M. and Bolduc, D. (1996) Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper, Department of Civil Engineering, MIT.

Bhat, C. R. (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomiallogit model, **Transportation Research B, Vol. 35,** 677–693.

Bolduc, D. (1999) A practical technique to estimate multinomial probit models in transportation, **Transportation Research B, Vol. 33,** 63-79.

Brownstone, D. and Train, K. (1999) Forecasting new product penetration with flexible substitution patterns, **Journal of Econometrics, Vol. 89, No. 1-2,** 109-129.

Bunch, D.S. (1991) Estimability in the multinomial probit model, **Transportation Research B, Vol. 25**, 1-12.

Carlsson, F. (2003) The demand for intercity public transport: the case of business passengers, **Applied Economics, Vol. 35, No. 1,** 41-50.

Daganzo, C. (1979) **Multinomial Probit: The Theory and Its Application to Demand Forecasting**. Academic Press, New York.

Friedman, J., Gerlowski, D. and Silberman, J. (1992) What attracts foreign multinational corporations? Evidence from branch plant location in the united states, **Journal of Regional Science, Vol. 32,** 403-418.

Geweke, J., Keane, M. and Runkle, D. (1994) Alternative computational approaches to inference in the multinomial probit model, **Review of Economics and Statistics, Vol. 76,** 609-632.

Hague Consulting Group, (1993) **ALOGIT 3.2f**, Hague, Netherlands.

Hansen, E. (1987) Industrial location choice in Sao Paulo, Brazil: a nested logit model, **Regional Science and Urban Economics, Vol. 17,** 89-108.

Hausman, J., Hall, B. H. and Griliches, Z. (1984) Econometric models for count data with an application to the Patents-R&D relationship, **Econometrica, Vol. 52, No. 4,** 909-938.

Jain, D., Vilcassim, N. and Chintagunta, P. (1994) A random-coefficients logit brand-choice model applied to panel data, **Journal of Business and Economic Statistics, Vol. 12, No. 3,** 317-328.

King, G. (1989) Variance specification in event count models: from restrictive assumptions to a generalized estimator, **American Journal of Political Science, Vol. 33,** 762-784.

List, J. (2001) US county-level determinants of inbound FDI: Evidence from a two-step modified count data model, **International Journal of Industrial Organization, Vol. 19,** 953-973.

Louviere, J., Hensher, D. and Swait, J. (2000) **Stated Choice Methods.** Cambridge University Press, Cambridge, England.

Maddala. G.S. (1983) **Limited Dependent and Qualitative Variables in Econometrics.** Cambridge University Press, Cambridge, England.

McCullagh, P. and Nelder, J. (1989) **Generalized Linear Models, 2nd edition.** Chapman & Hall, New York.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In P. Zarembka (eds.), **Frontier in Econometrics.** Academic Press, New York.

McFadden, D. (1978) Modelling the choice of residential location. In A. Karquist, , L. Lundqvist and F. Snickars (eds.), **Spatial Interaction Theory and Planning Models.** North-Holland, Amsterdam.

McFadden, D. and Train, K. (2000) Mixed MNL models for discrete response, **Journal of Applied Econometrics, Vol. 15, No. 5,** 447-470.

Munizaga, M.A., Heydecker, B.G. and Ortúzar, J. de D. (1997) On the error structure of discrete choice models, **Traffic Engineering and Control, Vol. 38, No.11**, 593-597.

Papke, L. (1991) Interstate business tax differentials and new firm location, **Journal of Public Economics, Vol. 45,** 47-68.

Quigley, J. M. (1976) Housing demand in the short run: an analysis of polytomous choice, **Explorations in Economic Research, Vol. 3, No. 1,** 76-102.

Walker, J., Ben-Akiva, M. and Bolduc, D. (2003) Identification of the logit kernel (or mixed logit) model. **10th International Conference on Travel Behavior Research,** Switzerland.

Woodward, D. (1992) Location determinants of Japanese manufacturing start-ups in the United States, **Southern Economic Journal, Vol. 58,** 690-708.

Wu, F. (1999) Intrametropolitan FDI firm location in Guangzhou, China: a Poisson and negative Binomial analysis, **Annals of Regional Science, Vol. 33,** 535-555.