INTERNATIONAL JOURNAL OF INTELLIGENT CONTROL AND SYSTEMS
VOL. 11, NO. 1, MARCH 2006, 36-50

Accounting for Future Utility in Medium-Term and Long-Term Choices

Cynthia CHEN and Xiaoqiang CHEN

Abstract- **In decision making processes, people often anticipate future. Future anticipation can take place in medium-term choice scenarios such as vehicle holding and transaction decisions and long-term ones such as residential and job location choices. Most of the discrete choice models developed in the area of travel behavior are incapable of accounting for expectation of future utility. It remains an open question in terms of how the existing modeling frameworks, such as the popular mixed logit models, can be modified to account for the expectation of future utility. This paper explores this question by constructing two theoretical models: a binary model and multinomial (3-alternative) model. For each model, a series of monte carlo simulations are conducted to explore the properties of the parameter estimates. In particular, two main questions are answered. First, what consequences will result if a without-future-utility model is applied to a dataset generated with the future utility component? Second, how well the existing modeling frameworks can be modified to account for the expectation of future utility in different choice scenarios? The simulation results illustrate three points mainly. First, if a without-future model is applied to a dataset generated with a future utility component, the estimated parameters attenuate to zero. In more complicated choice scenarios involving more than 2 alternatives and correlation between alternatives, the estimated parameters have wrong signs. Second, if a with-future model can be correctly specified on a dataset generated with the future utility component, both the probit model and the mixed logit model can recover true parameters quite well. Third, the use of model forms (probit vs. mixed logit) should be cautious.**

Index Terms—**Future expectation, monte carlo simulation, mixed logit model, probit model, discrete choice model**

1. INTRODUCTION

The field of discrete choice models has seen tremendous development during the past decade. The standard multinomial logit model has been rapidly replaced by the newly developed mixed logit model that explicitly accounts for unobserved heterogeneity that varies with individuals and correlation between alternatives [3, 9, 14].

In terms of the type of variables, we have also seen different types of variables that are being entered into the utility function. In particular, they include individual and household related socio-economic and demographic variables, alternative-specific variables, built environment related variables, and variables related to past experience. The entrance of past experience related variables is particularly notable because they represent our recognition that many of our daily decisions are the result of routines and habits [1, 2, 5]. There have been a number of studies that have used variables related to past experience [1, 2, 12, 13]. Some researchers term models with past experience related variables as dynamic models [10, 15], as technically these models include both the past and the current. Although representing both past and current times, these so-called dynamic models are not adding more complexity to the models that only include the variables observed at the current time. This is because past experience is the perceptions of things that have taken place in the past and therefore treated as deterministic variables.

In addition to past experience, we argue that in decision making processes, people often anticipate future. Take an example of a vehicle purchase decision scenario. A young couple with an existing passenger car and no children may choose to purchase a minivan instead of another passenger car at the current time. A minivan is chosen not because it provides higher utility than a passenger car at the current time, but because it provides higher utility than a passenger car if not only the utility at the current time but also the future expected utility are accounted for. The young couple may be expecting a child in a couple of years and with that in mind, a minivan is a better choice than a passenger car at the current time. For long-term choices such as residential and job location choices, anticipation for future utility is observed more often. A home location may be chosen not because it is close to the current employer, but because it is located in the center of the current employer and future employers [6, 11, 16].

Many of the discrete choice models developed in the area of travel behavior are incapable of accounting for expectation of future utility. It remains an open question in terms of how the existing modeling frameworks, such as the popular mixed logit models, can be modified to account for the expectation of future utility. In this paper, we explore this question by constructing two theoretical models: a binary model and a multinomial (3-alternative) model. For each model, we conducted a series of monte carlo simulations to explore the properties of the parameter estimates. In particular, we intend to answer two main questions. First, what consequences may result if a without-future-utility model is applied to a dataset generated with the future utility component? Second, how well the existing modeling frameworks can be modified to account for the expectation of future utility in different choice scenarios?

The rest of the paper is organized as follows. In the next section, we describe a general dynamic discrete choice framework. Our theoretical frameworks for the simulations are described afterwards. Then, we describe the generation of our artificial datasets, which will be used

Manuscript received Oct. 1, 2005; revised Feb. 25, 2006.
Cynthia Chen, Ph.D. is with Department of Civil Engineering, University Transportation Research Center, City College of New York, New York, 10031. Tel: 212-650-5372. Fax: 212-650-8374. Email: chen@ce.ccny.cuny.edu. Xiaoqiang Chen is with Department of Civil Engineering, University Transportation Research Center, City College of New York, New York, 10031. Tel: 212-650-8290. Fax: 212-650-6965. Email: chenxq@ce.ccny.cuny.edu

in simulations. We then describe the estimation procedures and results. The concluding section follows at last.

2. A DYNAMIC DISCRETE CHOICE FRAMEWORK

A dynamic discrete choice model framework is adopted for this study. When facing K discrete choices at time t, it is assumed that an individual attempts to maximize his/her lifetime utility. The word lifetime is used only to represent future periods. It is unlikely that people attempt to maximize their utility, while anticipating future expected utilities of their entire lifetime. On the other hand, it is highly possible that people would foresee changes that might take place in a couple of years ahead. For example, a household might decide to purchase a minivan with the anticipation that there will be a child coming to the family in a couple of years. In other words, the word lifetime here may only refer to the current time and the near future.

Suppose that the individual's lifetime can be sliced into t = 0,1,…,T stages, a linear-separable utility maximization function can be expressed as follows:

$$Max \quad E[\sum_{k=1} d_k(t)u_k(t) + \sum_{\tau=t+1}^{T} \omega^{\tau-t} \sum_{k=1} d_k(\tau)u_k(\tau) \mid S(t)], \quad (1)$$

where $u_k(t)$ is the utility that the individual derives by choosing alternative k at time t, ω is a discount factor between 0 and 1, $d_k(t) = 1$ indicates alternative k is chosen at time t, $d_k(t) = 0$ indicates alternative k is not chosen at time t, and S(t) is the vector of state variables. State variables may include individuals' socio-economic and demographic characteristics as well as alternative specific attributes. Equation (1) states that the individual attempts to maximize his/her expected utility of the lifetime.

Assuming that an individual's choice is represented by {k, k = 1,2,…,K}, the utility function, $u_k(t)$, may be further detailed as follows:

$$u_k(t) = v_k(t) + \varepsilon_k(t), \quad (2)$$

where

$u_k(t)$ is the random utility that the individual obtains at time t by choosing alternative k,

$v_k(t)$ is the systematic utility that the individual obtains by choosing alternative k at time t, and

$\varepsilon_k(t)$ is the random disturbance term in the utility function.

The probability of choosing alternative k at time t is given by the following equation:

$$\Pr(d_k(t)) = \Pr(u_k(t) + \omega E(u(t+1 \mid d_k(t))) \geq u_j(t) + \omega E(u(t+1 \mid d_j(t))), \quad \forall k, j \in \{1,2,...,K\}, \quad (3)$$

where $E(u(t+1 \mid d_j(t)))$ is the expectation of random utility that the individual obtains at time t+1 in case alternative j is chosen.

A key in estimation is to specify the joint distribution of the εs (see Eq. (2))that exist in both $u_k(t)$ and $u_k(t+1)$. In the case of serial independence, the joint distribution is: $\prod_{t=0}^{T} f(\varepsilon_{t1}, \varepsilon_{t2}, ..., \varepsilon_{tK}; \xi)$, where ξ is the vector of distributional parameters. In the case of correlation between alternatives and serial dependence, the joint distribution is: $f(\varepsilon_{01}, .., \varepsilon_{0K}, .., \varepsilon_{T1}, .., \varepsilon_{TK}; \xi)$. We conduct simulations of both cases in the paper.

3. THEORETICAL MODEL DEVELOPMENT FOR SIMULATIONS

We consider two general cases. In case 1, a binary choice structure is assumed at time t and at time t+1, each alternative at time t branches out two additional alternatives. In case 1, we distinguish between the subcase of serial independence and the subcase of serial dependence. In case 2, a multinomial (3 alternatives) choice structure is assumed at time t and at time t+1, each alternative at time t branches out two additional alternatives. Case 2 represents a more general case than case 1. For simplicity, we also set the discount factor, ω (in theory, ω cannot be separately estimated), to be equal to 1 in the rest of the paper.

3.1 Case 1 (2-Time-Period, 2-Choice at Time t)

Fig. 1. depicts a dynamic decision making framework in a 2-time-period and 2-choice structure. Suppose T = 2, meaning that an individual's lifetime is sliced into two periods. Each of these two time periods may be of different duration. Let us also suppose that the current time period is denoted by t and the future one by t+1. At t, decision makers face two alternatives: *a* and *b*. If they choose alternative *a* at t, they face alternatives *c* and *d* at t+1; if they choose *b* at t, they face alternatives *e* and *f* at t+1.

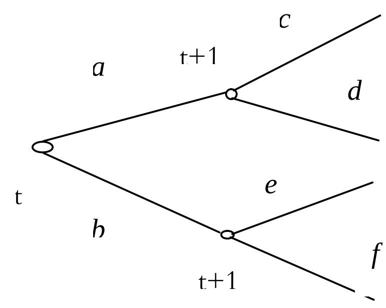

Fig. 1. A 2-Time-Period and 2-Choice Dynamic Decision Making Framework.

As an example, the decision making framework may depict a process that a family decides which vehicle to purchase (e.g., between a passenger car and a minivan) at time t. Time t denotes the current time and time t+1 may denote a couple of years later. Alternative *a* may be purchasing a passenger car and *b* purchasing a minivan. If the person chooses to purchase a passenger car (alternative *a*), he/she may face two alternatives at t+1. The two alternatives at t+1 may include: continue to keep the current fleet (alternative *c*) or replace one of the vehicles with another vehicle (alternative *d*). If the decision maker chooses alternative *b* (purchasing a minivan) at time t, he/she may face two alternatives at t+1: keep the current fleet (alternative *e*) and replace one of the vehicles with another vehicle (alternative *f*).

In the example described above, although the decision making process at time t is about whether to choose alternative *a* or alternative *b* at time t, factors that come into the decision making process include not only attributes associated with alternatives *a* and *b* at time t, but also attributes associated with alternatives *c*, *d*, *e*, and *f* at time t+1. In other words, future utility is clearly accounted for within the decision making process at time t.

In the rest of the section on case 1, we distinguish between two subcases: serial independence and serial dependence. In both cases, we only need to obtain the probability of choosing one alternative, as it is a binary choice model.

3.1.1. Serial Independence

Given the framework shown in Fig. 1., the probabilities of the individual's choosing alternative *a* at time t may be written as follows:

$$\begin{aligned}
&\Pr(d_a(t)=1)\\
&=\Pr(u_a(t)+E(u(t+1)\mid d_a(t))\geq u_b(t)+E(u(t+1)\mid d_b(t)))\\
&=\Pr(u_a(t)+(\Pr(d_c(t+1))\times u_c(t+1)+\Pr(d_d(t+1))\times u_d(t\\
&+1))\geq u_b(t)+(\Pr(d_e(t+1))\times u_e(t+1)+\Pr(d_f(t+1))\times\\
&u_f(t+1)),
\end{aligned}$$

where,

$$\begin{aligned}
&u_a(t)=v_a(t)+\varepsilon_a(t),\\
&u_c(t+1)=v_c(t+1)+\varepsilon_c(t+1),\\
&u_d(t+1)=v_d(t+1)+\varepsilon_d(t+1),\\
&u_b(t)=v_b(t)+\varepsilon_b(t),\\
&u_e(t+1)=v_e(t+1)+\varepsilon_e(t+1),\\
&u_f(t+1)=v_f(t+1)+\varepsilon_f(t+1).
\end{aligned}\tag{4}$$

In Eq. (4), there are a total of six disturbance terms involved: $\varepsilon_a(t)$, $\varepsilon_b(t)$, $\varepsilon_c(t+1)$, $\varepsilon_d(t+1)$, $\varepsilon_e(t+1)$, and $\varepsilon_f(t+1)$. In this simplest case, we assume all these disturbance terms follow an independent and identical standard normal distribution ($\sim N(0,1)$). The lifetime systematic utility components for alternative *a* and alternative *b* are: $v_a' = v_a(t)+\Pr(d_c(t+1))\times(v_c(t+1)) + \Pr(d_d(t+1)\times(v_d(t+1))$ and $v_b' = v_b(t)+\Pr(d_e(t+1))\times(v_e(t+1)) + \Pr(d_f(t+1))\times(v_f(t+1))$ respectively. Since all error components are assumed to be distributed with a standard normal distribution, the lifetime error term for alternative a, ε_a', also has a normal distribution with zero mean and a variance of $1+\Pr(d_c(t+1))^2+\Pr(d_d(t+1))^2$. Similarly, the lifetime error term for alternative b, ε_b', also has a normal distribution with zero mean and a variance of $1+\Pr(d_e(t+1))^2+\Pr(d_f(t+1))^2$. Consequently, Equation (4) can be further written as follows:

$$\begin{aligned}
&\Pr(d_a(t)=1)\\
&=\Pr(\varepsilon_b'-\varepsilon_a'\leq v_a(t)+\Pr(d_c(t+1))\times(v_c(t+1))+\Pr(d_d\\
&(t+1))\times(v_d(t+1))-v_b(t)-\Pr(d_e(t+1))\times(v_e(t+1))-\\
&\Pr(d_f(t+1))\times(v_f(t+1))\\
&=\Phi\left(\frac{\varepsilon_b'-\varepsilon_a'}{\delta}\leq\frac{v_a'-v_b'}{\delta}\right),
\end{aligned}\tag{5}$$

where,

$$\begin{aligned}
&\varepsilon_b'-\varepsilon_a'\sim N(0,2+\Pr(d_c(t+1))^2+\Pr(d_d(t+1))^2+\\
&\Pr(d_e(t+1))^2+\Pr(d_f(t+1))^2),\\
&v_a'-v_b'=v_a(t)+\Pr(d_c(t+1))\times(v_c(t+1))+\Pr(d_d(t+\\
&1))\times(v_d(t+1))-v_b(t)-\Pr(d_e(t+1))\times(v_e(t+1))-\Pr\\
&(d_f(t+1))\times(v_f(t+1)),
\end{aligned}\tag{6}$$

δ is the standard deviation of $\varepsilon_b'-\varepsilon_a'$, and $\delta=[2+\Pr(d_d(t+1))^2+\Pr(d_e(t+1))^2+\Pr(d_f(t+1))^2]^{1/2}$.

In estimation, β (the parameter vector associated with the independent variables in the lifetime systematic utility) and δ cannot be separately identified and only the ratio of the two can be identified [4, 17]. Because of this feature, we can set δ to be 1 and then estimate β afterwards. The estimation of β is done by performing a 1-dimensional standard normal integration.

3.1.2. Serial Dependence

A more complicated case can be established by assuming serial dependence between $\varepsilon_a(t)$ and $\varepsilon_c(t+1)$, $\varepsilon_a(t)$ and $\varepsilon_d(t+1)$, $\varepsilon_b(t)$ and $\varepsilon_e(t+1)$, and $\varepsilon_b(t)$ and $\varepsilon_f(t+1)$. The variance-covariance matrix for these error terms: $\varepsilon_a(t)$, $\varepsilon_b(t)$, $\varepsilon_c(t+1)$, $\varepsilon_d(t+1)$, $\varepsilon_e(t+1)$, and $\varepsilon_f(t+1)$, is:

$$\begin{bmatrix}
\delta_a^2 & 0 & \delta_{a,c} & \delta_{a,d} & 0 & 0\\
0 & \delta_b^2 & 0 & 0 & \delta_{b,e} & \delta_{b,f}\\
\delta_{a,c} & 0 & \delta_c^2 & 0 & 0 & 0\\
\delta_{a,d} & 0 & 0 & \delta_d^2 & 0 & 0\\
0 & \delta_{b,e} & 0 & 0 & \delta_e^2 & 0\\
0 & \delta_{b,f} & 0 & 0 & 0 & \delta_f^2
\end{bmatrix}. \tag{7}$$

Without losing generality, we can rescale the above matrix such that the diagonal elements are all equal to 1:

$$\begin{bmatrix} 1 & 0 & \frac{\delta_{a,c}}{\delta_a^2} & \frac{\delta_{a,d}}{\delta_a^2} & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{\delta_{b,e}}{\delta_a^2} & \frac{\delta_{b,f}}{\delta_a^2} \\ \frac{\delta_{a,c}}{\delta_a^2} & 0 & 1 & 0 & 0 & 0 \\ \frac{\delta_{a,d}}{\delta_a^2} & 0 & 0 & 1 & 0 & 0 \\ 0 & \frac{\delta_{b,e}}{\delta_a^2} & 0 & 0 & 1 & 0 \\ 0 & \frac{\delta_{b,f}}{\delta_a^2} & 0 & 0 & 0 & 1 \end{bmatrix}, \quad (8)$$

where $|\delta_{a,c}/\delta_a^2| \leq 1$, $|\delta_{a,d}/\delta_a^2| \leq 1$, $|\delta_{b,e}/\delta_a^2| \leq 1$, and $|\delta_{b,f}/\delta_a^2| \leq 1$. (9)

Given that $\varepsilon_a^{'} = \varepsilon_a(t) + \Pr(d_c(t+1)) \times (\varepsilon_c(t+1)) \times (\varepsilon_d(t+1))$, the lifetime error term for alternative $a, \varepsilon_a^{'}$, still has a normal distribution with zero mean, but its variance is different from the variance in the subcase of serial independence. In particular, we can derive its variance as follows:

$$\begin{aligned} &Var(\varepsilon_a^{'}) \\ &= Var(\varepsilon_a(t) + \Pr(d_c(t+1)) \times \varepsilon_c(t+1) + \Pr(d_d(t+1)) \times \varepsilon_d(t+1)) \\ &= Var(\varepsilon_a(t)) + Var(\Pr(d_c(t+1)) \times \varepsilon_c(t+1)) + 2Cov(\varepsilon_a(t), \Pr(d_c(t+1)) \times \varepsilon_c(t+1)) + Var(\Pr(d_d(t+1)) \times \varepsilon_d(t+1)) + 2Cov(\varepsilon_a(t), \\ &\Pr(d_d(t+1)) \times \varepsilon_d(t+1)) + 2Cov(\Pr(d_c(t+1)) \times \varepsilon_c(t+1), \Pr(d_d(t+1)) \times \varepsilon_d(t+1)) \\ &= 1 + \Pr(d_c(t+1))^2 + 2\Pr(d_c(t+1))\delta_{a,c}/\delta_a^2 + \Pr(d_d(t+1))^2 + 2\Pr(d_d(t+1))\delta_{a,d}/\delta_a^2. \end{aligned} \quad (10)$$

In addition, $Cov(\Pr(d_c(t+1) \times \varepsilon_c(t+1), \Pr(d_d(t+1) \times \varepsilon_d(t+1)) = 0$, since there is no correlation assumed between alternatives c and d (there is only serial dependence here). Similarly, the lifetime error term for alternative b, $\varepsilon_b^{'}$, also has a normal distribution with zero mean and its variance is: $1 + \Pr(d_e(t+1))^2 + \Pr(d_f(t+1))^2 + 2\Pr(d_e(t+1))\delta_{b,e}/\delta_a^2 + 2\Pr(d_f(t+1))\delta_{b,f}/\delta_a^2$. $\varepsilon_b^{'} - \varepsilon_a^{'}$ has a normal distribution as follows:

$$\begin{aligned} \varepsilon_b^{'} - \varepsilon_a^{'} \sim N(0, &2 + \Pr(d_c(t+1))^2 + \Pr(d_d(t+1))^2 + \Pr(d_e(t+1))^2 + \Pr(d_f(t+1))^2 + 2\Pr(d_c(t+1))\delta_{a,c}/\delta_a^2 + \\ &2\Pr(d_d(t+1))\delta_{a,d}/\delta_a^2 + 2\Pr(d_e(t+1))\delta_{b,e}/\delta_a^2 + 2\Pr(d_f(t+1))\delta_{b,f}/\delta_a^2). \end{aligned} \quad (11)$$

Here, for the subcase of serial dependence in case 1, we denote δ_1 as the standard deviation of $\varepsilon_b^{'} - \varepsilon_a^{'}$ and

$$\begin{aligned} \delta_1 = [&2 + \Pr(d_c(t+1))^2 + \Pr(d_d(t+1))^2 + \Pr(d_e(t+1))^2 \\ &+ \Pr(d_f(t+1))^2 + 2\Pr(d_c(t+1))\delta_{a,c}/\delta_a^2 + 2\Pr(d_d(t+1)) \\ &\delta_{a,d}/\delta_a^2 + 2\Pr(d_e(t+1))\delta_{b,e}/\delta_a^2 + 2\Pr(d_f(t+1)) \\ &\delta_{b,f}/\delta_a^2)]^{1/2}. \end{aligned} \quad (12)$$

In the subcase of serial dependence, although the variance of $\varepsilon_b^{'} - \varepsilon_a^{'}$ is more complicated than that in the subcase of serial independence, the idea behind the estimation procedure is still the same. Because β and the standard deviation of $\varepsilon_b^{'} - \varepsilon_a^{'}$ cannot be separately identified and only the ratio of the two can be identified [4, 17], we can still set the standard deviation of $\varepsilon_b^{'} - \varepsilon_a^{'}$, or δ_1 to be 1 and then estimate β afterwards. Similar to the subcase of serial independence, the estimation of β is done by performing a 1-dimensional standard normal integration.

3.2 Case 2 (2-Time Period and 3-Choice at Time t)

Fig. 2. depicts a dynamic decision making framework in a 2-time-period and 3-choice structure. In this case, at the current time t, the decision maker faces three alternatives: *a'*, *b'*, and *c'*. If the decision maker chooses *a'* at time t, he/she faces *d'* and *e'* at t+1; if the decision maker chooses *b'* at time t, he/she faces *f'* and *g'* at t+1; if the decision maker chooses *c'* at time t, he/she faces *h'* and *i'* at t+1.

As an example, the decision making framework may depict a process that a family decide which car to purchase (e.g., between a passenger car, a SUV, or a minivan) at t. Time t may denote the current time when the decision needs to be made, and time t+1 may denote a couple of years later. Alternative *a'* may be viewed as the alternative of purchasing a passenger car; *b'* may be viewed as purchasing a SUV; and *c'* may be viewed as purchasing a minivan. Any choice chosen at t will result in a decision making process between two alternatives at t+1. The two alternatives at t+1 are denoted as *d'* and *e'*, *f'* and *g'*, and *h'* and *i'*, corresponding to *a'*, *b'*, and *c'* at t respectively.

Similar to case 1 (2-time-period and 2-choice structure), although the decision making process at t is about whether to choose *a'*, *b'*, or *c'*, factors that come into the decision making process include not only attributes associated with *a'*, *b'*, and *c'* at t, but also attributes associated with *d'*, *e'*, *f'*, *g'*, *h'*, and *i'* at t+1.

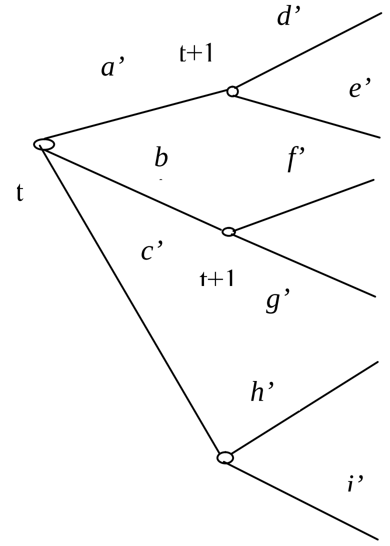

Fig. 2. A 2-Time-Period 3-Choice Dynamic Decision Making Framework

Given the framework shown in Fig. 2., the probability of the individual's choosing an alternative at t may be written as follows:

$$\Pr(d_{k'}(t)=1) = \Pr(u_{k'}(t) + E(u(t+1 \mid d_{k'}(t))) \geq u_{j'}(t) + E(u(t+1 \mid d_{j'}(t))), \forall k', j' \in (a', b', c'). \quad (13)$$

Unlike case 1, in case 2, we will not distinguish between serial independence and serial dependence. As shown in case 1, as long as there is no correlation between different alternatives, serial dependence does not really complicate our estimation procedure. Instead, in case 2, we will discuss a more general case, which allows correlation between alternatives.

Let us assume the lifetime random utilities of alternatives a', b', and c' can be expressed as follows:

$$\begin{aligned}
u'_{a'}(t) &= v_{a'}(t) + \Pr[d_{d'}(t+1)] \times v_{d'}(t+1) + \Pr[d_{e'}(t+1)] \times \\
&\quad v_{e'}(t+1) + \Pr[d_{e'}(t+1)] \times \varepsilon_{e'}(t+1) + \Pr[d_{d'}(t+1)] \times \\
&\quad \varepsilon_{d'}(t+1) + \varepsilon_{a'}(t) = v'_{a'} + \varepsilon'_{a'}, \\
u'_{b'}(t) &= v_{b'}(t) + \Pr[d_{f'}(t+1)] \times v_{f'}(t+1) + \Pr[d_{g'}(t+1)] \\
&\quad \times v_{g'}(t+1) + \Pr[d_{f'}(t+1)] \times \varepsilon_{f'}(t+1) + \Pr[d_{g'}(t+1)] \times \\
&\quad \varepsilon_{g'}(t+1) + \varepsilon_{b'}(t) = v'_{b'} + \varepsilon'_{b'},
\end{aligned}$$

and

$$\begin{aligned}
u'_{c'}(t) &= v_{c'}(t) + \Pr[d_{h'}(t+1)] \times v_{h'}(t+1) + \Pr[d_{i'}(t+1)] \\
&\quad \times v_{i'}(t+1) + \Pr[d_{h'}(t+1)] \times \varepsilon_{h'}(t+1) + \Pr[d_{i'}(t+1)] \\
&\quad \times \varepsilon_{i'}(t+1) + \varepsilon_{c'}(t) = v'_{c'} + \varepsilon'_{c'},
\end{aligned} \quad (14)$$

where,

$$\begin{aligned}
v'_{a'} &= v_{a'}(t) + \Pr[d_{d'}(t+1)] \times v_{d'}(t+1) + \Pr[d_{e'}(t+1)] \times \\
&\quad v_{e'}(t+1), \\
\varepsilon'_{a'} &= \Pr[d_{e'}(t+1)] \times \varepsilon_{e'}(t+1) + \Pr[d_{d'}(t+1)] \times \varepsilon_{d'}(t+1) \\
&\quad + \varepsilon_{a'}(t), \\
v'_{b'} &= v_{b'}(t) + \Pr[d_{f'}(t+1)] \times v_{f'}(t+1) + \Pr[d_{g'}(t+1)] \times \\
&\quad v_{g'}(t+1), \\
\varepsilon'_{b'} &= \Pr[d_{f'}(t+1)] \times \varepsilon_{f'}(t+1) + \Pr[d_{g'}(t+1)] \times \varepsilon_{g'}(t+1) \\
&\quad + \varepsilon_{b'}(t), \\
v'_{c'} &= v_{c'}(t) + \Pr[d_{h'}(t+1)] \times v_{h'}(t+1) + \Pr[d_{i'}(t+1)] \\
&\quad \times v_{i'}(t+1),
\end{aligned}$$

and

$$\begin{aligned}
\varepsilon'_{c'} &= \Pr[d_{h'}(t+1)] \times \varepsilon_{h'}(t+1) + \Pr[d_{i'}(t+1)] \times \varepsilon_{i'}(t+1) \\
&\quad + \varepsilon_{c'}(t).
\end{aligned} \quad (15)$$

We further assume that the variance-covariance matrix of alternatives a', b', c', d', e', f', g', h', and i' can be represented by the following matrix:

$$\begin{bmatrix}
\delta^2_{a'} & . & . & . & . & . & . & . & . \\
\delta_{a'b'} & \delta^2_{b'} & . & . & . & . & . & . & . \\
\delta_{a'c'} & \delta_{b'c'} & \delta^2_{c'} & . & . & . & . & . & . \\
\delta_{a'd'} & \delta_{b'd'} & \delta_{c'd'} & \delta^2_{d'} & . & . & . & . & . \\
\delta_{a'e'} & \delta_{b'e'} & \delta_{c'e'} & \delta_{d'e'} & \delta^2_{e'} & . & . & . & . \\
\delta_{a'f'} & \delta_{b'f'} & \delta_{c'f'} & \delta_{d'f'} & \delta_{e'f'} & \delta^2_{f'} & . & . & . \\
\delta_{a'g'} & \delta_{b'g'} & \delta_{c'g'} & \delta_{d'g'} & \delta_{e'g'} & \delta_{f'g'} & \delta^2_{g'} & . & . \\
\delta_{a'h'} & \delta_{b'h'} & \delta_{c'h'} & \delta_{d'h'} & \delta_{e'h'} & \delta_{f'h'} & \delta_{g'h'} & \delta^2_{h'} & . \\
\delta_{a'i'} & \delta_{b'i'} & \delta_{c'i'} & \delta_{d'i'} & \delta_{e'i'} & \delta_{f'i'} & \delta_{g'i'} & \delta_{h'i'} & \delta^2_{i'}
\end{bmatrix}. \quad (16)$$

The probability of choosing alternative a' at time t can be written as follows:

$$\begin{aligned}
\Pr(d_{a'}(t) = 1) &= \Pr(v'_{a'} + \varepsilon'_{a'} \geq v'_{b'} + \varepsilon'_{b'} \text{ and } v'_{a'} + \varepsilon'_{a'} \geq v'_{c'} + \varepsilon'_{c'}) \\
&= \Pr(\varepsilon'_{b'} - \varepsilon'_{a'} \leq v'_{a'} - v'_{b'}, \text{ and } \varepsilon'_{c'} - \varepsilon'_{a'} \leq v'_{a'} - v'_{c'}).
\end{aligned} \quad (17)$$

Similarly, the probabilities of choosing b' and c' are:

$\Pr(d_{b'}(t) = 1) = \Pr(\varepsilon'_{a'} - \varepsilon'_{b'} \leq v'_{b'} - v'_{a'}$ and $\varepsilon'_{c'} - \varepsilon'_{b'} \leq v'_{b'} - v'_{c'})$ (Eq 18), and

$\Pr(d_{c'}(t) = 1) = \Pr(\varepsilon'_{a'} - \varepsilon'_{c'} \leq v'_{c'} - v'_{a'}$ and $\varepsilon'_{b} - \varepsilon'_{c'} \leq v'_{c'} - v'_{b'})$ (Eq 19) respectively.

3.2.1 Probit Model

From the above equation, we can write the probability of selecting alternative a' as follows:

$$\Pr(d_{a'}(t)=1)$$
$$=\Pr(\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'}\le v_{a'}^{'}-v_{b'}^{'} \quad \text{and} \quad \varepsilon_{c'}^{'}-\varepsilon_{a'}^{'}\le v_{a'}^{'}-v_{c'}^{'})$$
$$=\Pr(\frac{\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'})}}\le\frac{v_{b'}^{'}-v_{a'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'})}} \quad \text{and} \quad \frac{\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'})}}\le\frac{v_{c'}^{'}-v_{a'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'})}}). \tag{20}$$

The variance-covariance matrix of $\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'}$ and $\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'}$ can be denoted as $\Sigma_{a'}$ and $\Sigma_{a'}=\begin{bmatrix} pbr_{b'-a'} & pbr_{c'-a',b'-a'} \\ pbr_{c'-a',b'-a'} & pbr_{c'-a'}\end{bmatrix}$. Since β and $\Sigma_{a'}$ cannot be estimated separately [4，17], we can rescale $\Sigma_{a'}$ into $\hat{\Sigma}_{a'}$ by doing the following:

$$\Sigma_{a'}=\begin{bmatrix} pbr_{b'-a'} & pbr_{c'-a',b'-a'} \\ pbr_{c'-a',b'-a'} & pbr_{c'-a'}\end{bmatrix}$$
$$=pbr_{b'-a'}\times\begin{bmatrix} 1 & pbr_{c'-a',b'-a'}/pbr_{b'-a'} \\ pbr_{c'-a',b'-a'}/pbr_{b'-a'} & pbr_{c'-a'}/pbr_{b'-a'}\end{bmatrix}$$
$$=pbr_{b'-a'}\times\hat{\Sigma}_{a'}.$$

Given $v_{a'}^{'}=v_{a'}(t)+\Pr(d_{d'}(t+1)\times v_{d'}(t+1)+\Pr(d_{e'}(t+1)\times v_{e'}(t+1)$ the calculation of $v_{a'}^{'}$, $v_{b'}^{'}$, and $v_{c'}^{'}$ requires the calculation of $\Pr(d_{d'}(t+1))$, $\Pr(d_{e'}(t+1))$, $\Pr(d_{f'}(t+1))$, $\Pr(d_{g'}(t+1))$, $\Pr(d_{h'}(t+1))$, and $\Pr(d_{i'}(t+1))$, which then requires the calculation of $\operatorname{var}(\varepsilon_{d'}-\varepsilon_{e'})$, $\operatorname{var}(\varepsilon_{d'}-\varepsilon_{e'})$, $\operatorname{var}(\varepsilon_{f'}-\varepsilon_{g'})$, and $\operatorname{var}(\varepsilon_{h'}-\varepsilon_{i'})$. We assume:

$$\operatorname{var}(\varepsilon_{d'}-\varepsilon_{e'})=\operatorname{var}(\varepsilon_{f'}-\varepsilon_{g'})=\operatorname{var}(\varepsilon_{h'}-\varepsilon_{i'})=1.$$

The only effect of this assumption is that $pbr_{b'-a'}$ is rescaled. β is rescaled by $pbr_{b'-a'}$, which is then influenced by the assumption of $\operatorname{var}(\varepsilon_{d'}-\varepsilon_{e'})=\operatorname{var}(\varepsilon_{f'}-\varepsilon_{g'})=\operatorname{var}(\varepsilon_{h'}-\varepsilon_{i'})=1$. Therefore, for the probit model, the probability of selecting *a'* can then be expressed as:

$$\Pr(d_{a'}(t)=1)$$
$$=\Pr(\frac{\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'})}}\le\frac{v_{b'}^{'}-v_{a'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'})}} \quad \text{and} \quad \frac{\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'})}}\le\frac{v_{c'}^{'}-v_{a'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'})}})$$
$$=\int_{-\infty}^{v_{b'}^{'}-v_{a'}^{'}}\int_{-\infty}^{v_{c'}^{'}-v_{a'}^{'}} f(\hat{\Sigma}_{a'})d(\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'})d(\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'}). \tag{21}$$

The procedure to derive the probability of choosing b' and c' is similar to that for the probability of choosing *a'*. The rescaled variance-covariance matrices of $\varepsilon_{a'}^{'}-\varepsilon_{b'}^{'}$ and $\varepsilon_{c'}^{'}-\varepsilon_{b'}^{'}$ and $\varepsilon_{b'}^{'}-\varepsilon_{c'}^{'}$ and $\varepsilon_{a'}^{'}-\varepsilon_{c'}^{'}$ for b' and c' are $\hat{\Sigma}_{b'}$ and $\hat{\Sigma}_{c'}$ respectively. Similarly, the probabilities of selecting b' and c' can be expressed as:

$$\Pr(d_{b'}(t)=1)$$
$$=\Pr(\frac{\varepsilon_{c'}^{'}-\varepsilon_{b'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{c'}^{'}-\varepsilon_{b'}^{'})}}\le\frac{v_{c'}^{'}-v_{b'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{c'}^{'}-\varepsilon_{b'}^{'})}} \quad \text{and} \quad \frac{\varepsilon_{a'}^{'}-\varepsilon_{b'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{a'}^{'}-\varepsilon_{b'}^{'})}}\le\frac{v_{a'}^{'}-v_{b'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{a'}^{'}-\varepsilon_{b'}^{'})}})$$
$$=\int_{-\infty}^{v_{c'}^{'}-v_{b'}^{'}}\int_{-\infty}^{v_{a'}^{'}-v_{b'}^{'}} f(\hat{\Sigma}_{b'})d(\varepsilon_{c'}^{'}-\varepsilon_{b'}^{'})d(\varepsilon_{a'}^{'}-\varepsilon_{b'}^{'}), \quad \text{and} \tag{22}$$

$$\Pr(d_{c'}(t)=1)$$
$$=\Pr(\frac{\varepsilon_{b'}^{'}-\varepsilon_{c'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{b'}^{'}-\varepsilon_{c'}^{'})}}\le\frac{v_{b'}^{'}-v_{c'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{b'}^{'}-\varepsilon_{c'}^{'})}} \quad \text{and} \quad \frac{\varepsilon_{a'}^{'}-\varepsilon_{c'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{a'}^{'}-\varepsilon_{c'}^{'})}}\le\frac{v_{a'}^{'}-v_{c'}^{'}}{\sqrt{\operatorname{var}(\varepsilon_{a'}^{'}-\varepsilon_{c'}^{'})}})$$
$$=\int_{-\infty}^{v_{b'}^{'}-v_{c'}^{'}}\int_{-\infty}^{v_{a'}^{'}-v_{c'}^{'}} f(\hat{\Sigma}_{c'})d(\varepsilon_{b'}^{'}-\varepsilon_{c'}^{'})d(\varepsilon_{a'}^{'}-\varepsilon_{c'}^{'}). \tag{23}$$

3.2.2 *Mixed Logit Model*

In this setting, we adopt the popular mixed logit model to estimate the model in case 2. Mathematically, we can write:

$$u_{a'}^{'}=v_{a'}^{'}+\varepsilon_{a'}^{'}=v_{a'}^{'}+\eta_{a'}+\varphi_{a'},$$
$$u_{b'}^{'}=v_{b'}^{'}+\varepsilon_{b'}^{'}=v_{b'}^{'}+\eta_{b'}+\varphi_{b'},$$
$$u_{c'}^{'}=v_{c'}^{'}+\varepsilon_{c'}^{'}=v_{c'}^{'}+\eta_{c'}+\varphi_{c'}, \tag{24}$$

where $\varphi_{a'}$, $\varphi_{b'}$, and $\varphi_{c'}$ are independently and identically distributed with extreme value distributions, and $\eta_{a'}$, $\eta_{b'}$ and $\eta_{c'}$ have normal distributions and their variance-covariance matrix can be expressed as:

$$\Omega=\begin{bmatrix} ml_{a'} & ml_{a',b'} & ml_{c',a'} \\ ml_{a',b'} & ml_{b'} & ml_{b',c'} \\ ml_{a',c'} & ml_{c',b'} & ml_{c'}\end{bmatrix}$$
$$=ml_{a'}\times\begin{bmatrix} 1 & ml_{a',b'}/ml_{a'} & ml_{c',a'}/ml_{a'} \\ ml_{a',b'}/ml_{a'} & ml_{b'}/ml_{a'} & ml_{b',c'}/ml_{a'} \\ ml_{a',c'}/ml_{a'} & ml_{c',b'}/ml_{a'} & ml_{c'}/ml_{a'}\end{bmatrix}$$
$$=ml_{a'}\times\hat{\Omega} \tag{25}$$

For the same reasons described above, the rescaled variance-covariance matrices: $ml_{a'}$, is rescaled by assuming $\mathrm{var}(\varepsilon_{d'} - \varepsilon_{e'}) = \mathrm{var}(\varepsilon_{f'} - \varepsilon_{g'}) = \mathrm{var}(\varepsilon_{h'} - \varepsilon_{i'}) = 1$. Consequently, the marginal probabilities of selecting alternatives a', b', and c', given $\eta_{a'}$, $\eta_{b'}$ and $\eta_{c'}$ are:

$$\Pr(d_{a'}(t) \mid \eta_{j'}) = \frac{\exp(v'_{a'} + \eta_{a'})}{\sum_{j'} \exp(v'_{j'} + \eta_{j'})}, \; j' \in (a', b', c') \quad (26)$$

$$\Pr(d_{b'}(t) \mid \eta_{j'}) = \frac{\exp(v'_{b'} + \eta_{b'})}{\sum_{j'} \exp(v'_{j'} + \eta_{j'})}, \; j' \in (a', b', c') \quad (27)$$

$$\Pr(d_{c'}(t) \mid \eta_{j'}) = \frac{\exp(v'_{c'} + \eta_{c'})}{\sum_{j'} \exp(v'_{j'} + \eta_{j'})}, \; j' \in (a', b', c') \quad (28)$$

The unconditional probabilities of selecting a', b', and c' are the integral of the marginal probabilities over all possible values of $\eta_{a'}$, $\eta_{b'}$ and $\eta_{c'}$:

$$\Pr(d_{a'}(t)) = \int\!\!\int\!\!\int \frac{\exp(v'_{a'} + \eta_{a'})}{\sum_{j'} \exp(v'_{j'} + \eta_{j'})} f(\hat{\Omega})(d\eta_{j'})^3 \quad (29)$$

$$\Pr(d_{b'}(t)) = \int\!\!\int\!\!\int \frac{\exp(v'_{b'} + \eta_{b'})}{\sum_{j'} \exp(v'_{j'} + \eta_{j'})} f(\hat{\Omega})(d\eta_{j'})^3 \quad (30)$$

$$\Pr(d_{c'}(t)) = \int\!\!\int\!\!\int \frac{\exp(v'_{c'} + \eta_{c'})}{\sum_{j'} \exp(v'_{j'} + \eta_{j'})} f(\hat{\Omega})(d\eta_{j'})^3. \quad (31)$$

Three 3-dimensional integrals in the above equations can be easily reduced to 2-dimentional integrals by setting the lifetime random utility of one of the alternatives to 0. For example, if we set the lifetime random utility of alternative a' to be zero, the unconditional probabilities of selecting alternatives a', b', and c' can be written as follows:

$$\Pr(d_{a'}(t)) = \int\!\!\int \frac{1}{1 + \sum_{j'} \exp(v'_{j'} + \eta_{j'})} f(\tilde{\Omega}) d\eta_{j'} d\eta_{k'}, \; j' \neq a', j' \neq k', \quad (32)$$

$$\Pr(d_{b'}(t)) = \int\!\!\int \frac{\exp(v'_{b'} + \eta_{b'})}{1 + \sum_{j'} \exp(v'_{j'} + \eta_{j'})} f(\tilde{\Omega}) d\eta_{j'} d\eta_{k'}, \; j' \neq a', j' \neq k', \quad (33)$$

and

$$\Pr(d_{c'}(t)) = \int\!\!\int \frac{\exp(v'_{c'} + \eta_{c'})}{1 + \sum_{j'} \exp(v'_{j'} + \eta_{j'})} f(\tilde{\Omega}) d\eta_{j'} d\eta_{k'}, \; j' \neq a', j' \neq k', \quad (34)$$

$$\text{where } M = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \; \tilde{\Omega} = M\hat{\Omega}M'. \quad (35)$$

4. ARTIFICIAL DATASET GENERATION

In both cases 1 and 2, we generate the initial values of **x** (independent variables) at time t from a standard normal distribution. For the subcase of serial independence in case 1, all εs are generated from a standard normal distribution. For the subcase of serial dependence in case 1, the εs are generated with a normal distribution and its variance-covariance matrix of $\varepsilon_a(t)$, $\varepsilon_b(t)$, $\varepsilon_c(t+1)$, $\varepsilon_d(t+1)$, $\varepsilon_e(t+1)$, and $\varepsilon_f(t+1)$ is:

$$\begin{bmatrix} \delta_a^2 & 0 & \delta_{a,c} & \delta_{a,d} & 0 & 0 \\ 0 & \delta_b^2 & 0 & 0 & \delta_{b,e} & \delta_{b,f} \\ \delta_{a,c} & 0 & \delta_c^2 & 0 & 0 & 0 \\ \delta_{a,d} & 0 & 0 & \delta_d^2 & 0 & 0 \\ 0 & \delta_{b,e} & 0 & 0 & \delta_e^2 & 0 \\ 0 & \delta_{b,f} & 0 & 0 & 0 & \delta_f^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 1 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For case 2, the εs are generated with a normal distribution and its variance-covariance matrix of $\varepsilon_{a'}(t)$, $\varepsilon_{b'}(t)$, $\varepsilon_{c'}(t)$, $\varepsilon_{d'}(t+1)$, $\varepsilon_{e'}(t+1)$, $\varepsilon_{f'}(t+1)$, $\varepsilon_{g'}(t+1)$, $\varepsilon_{h'}(t+1)$, and $\varepsilon_{i'}(t+1)$ is:

$$\begin{bmatrix} \delta_{a'}^2 & . & . & . & . & . & . & . & . \\ \delta_{a'b'} & \delta_{b'}^2 & . & . & . & . & . & . & . \\ \delta_{a'c'} & \delta_{b'c'} & \delta_{c'}^2 & . & . & . & . & . & . \\ \delta_{a'd'} & \delta_{b'd'} & \delta_{c'd'} & \delta_{d'}^2 & . & . & . & . & . \\ \delta_{a'e'} & \delta_{b'e'} & \delta_{c'e'} & \delta_{d'e'} & \delta_{e'}^2 & . & . & . & . \\ \delta_{a'f'} & \delta_{b'f'} & \delta_{c'f'} & \delta_{d'f'} & \delta_{e'f'} & \delta_{f'}^2 & . & . & . \\ \delta_{a'g'} & \delta_{b'g'} & \delta_{c'g'} & \delta_{d'g'} & \delta_{e'g'} & \delta_{f'g'} & \delta_{g'}^2 & . & . \\ \delta_{a'h'} & \delta_{b'h'} & \delta_{c'h'} & \delta_{d'h'} & \delta_{e'h'} & \delta_{f'h'} & \delta_{g'h'} & \delta_{h'}^2 & . \\ \delta_{a'i'} & \delta_{b'i'} & \delta_{c'i'} & \delta_{d'i'} & \delta_{e'i'} & \delta_{f'i'} & \delta_{g'i'} & \delta_{h'i'} & \delta_{i'}^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & . & . & . & . & . & . & . & . \\ 0.1 & 1 & . & . & . & . & . & . & . \\ 0 & 0.1 & 1 & . & . & . & . & . & . \\ 0.3 & 0.1 & 0.1 & 1 & . & . & . & . & . \\ 0.3 & 0.1 & 0.1 & 0.2 & 1 & . & . & . & . \\ 0.1 & 0.3 & 0.1 & 0 & 0 & 1 & . & . & . \\ 0.1 & 0.3 & 0.1 & 0 & 0 & 0.2 & 1 & . & . \\ 0.1 & 0.1 & 0.3 & 0 & 0 & 0 & 0 & 1 & . \\ 0.1 & 0.1 & 0.3 & 0 & 0 & 0 & 0 & 0.2 & 1 \end{bmatrix}$$

In the rest of this section, we describe how to update the values of **x** from time t to t+1.

4.1 Case 1

For case 1, we use the following random utility functions:

$$u_a(t) = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \beta_5 x_{5,t} + \varepsilon_{a,t} = v_a(t) + \varepsilon_{a,t},$$

$$u_b(t) = \beta_6 x_{6,t} + \beta_7 x_{7,t} + \beta_8 x_{8,t} + \beta_9 x_{9,t} + \beta_{10} x_{10,t} + \varepsilon_{b,t} = v_b(t) + \varepsilon_{b,t},$$

$$u_c(t+1) = \beta_1 x_{1,t+1} + \beta_2 x_{2,t+1} + \beta_3 x_{3,t+1} + \beta_4 x_{4,t+1} + \beta_5 x_{5,t+1} + \varepsilon_{c,t+1},$$

$$u_d(t+1) = \beta_6 x_{6,t+1} + \beta_7 x_{7,t+1} + \beta_8 x_{8,t+1} + \beta_9 x_{9,t+1} + \beta_{10} x_{10,t+1} + \varepsilon_{d,t+1},$$

$$u_e(t+1) = \beta_1 x_{1,t+1} + \beta_2 x_{2,t+1} + \beta_3 x_{3,t+1} + \beta_4 x_{4,t+1} + \beta_5 x_{5,t+1} + \varepsilon_{e,t+1},$$

and

$$u_f(t+1) = \beta_6 x_{6,t+1} + \beta_7 x_{7,t+1} + \beta_8 x_{8,t+1} + \beta_9 x_{9,t+1} + \beta_{10} x_{10,t+1} + \varepsilon_{f,t+1}. \quad (36)$$

Let m denote the m-th variable (m=1~10) and $Ratio = |v_a(t)| + |v_b(t)|$. If alternative a is chosen at time t, $x_{m,t+1} = x_{m,t} - v_a / Ratio$ for $m = 1, 4$ and 5, $x_{m,t+1} = x_{m,t} + +v_b / Ratio$, for $m = 8, 9$ and 10, and $x_{m,t+1} = x_{m,t}$, for $m = 2, 3, 6$ and 7. If b is chosen at time t, $x_{m,t+1} = x_{m,t} + v_a / Ratio$ for $m = 1, 4$ and 5, $x_{m,t+1} = x_{m,t} - v_b / Ratio$ for $m = 8, 9$ and 10, and $x_{m,t+1} = x_{m,t}$ for $m = 2, 3, 6$ and 7.

4.2 Case 2

For case 2, we apply the following random utility functions:

$$u_{a'}(t) = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 x_{3,t} + \beta_4 x_{4,t} + \beta_5 x_{5,t} + \varepsilon_{a',t},$$

$$u_{b'}(t) = \beta_6 x_{6,t} + \beta_7 x_{7,t} + \beta_8 x_{8,t} + \beta_9 x_{9,t} + \beta_{10} x_{10,t} + \varepsilon_{b',t},$$

$$u_{c'}(t) = \beta_{11} x_{11,t} + \beta_{12} x_{12,t} + \beta_{13} x_{13,t} + \beta_{14} x_{14,t} + \beta_{15} x_{15,t} + \varepsilon_{c',t},$$

$$u_{d'}(t+1) = \beta_1 x_{1,t+1} + \beta_2 x_{2,t+1} + \beta_3 x_{3,t+1} + \beta_4 x_{4,t+1} + \beta_5 x_{5,t+1} + \varepsilon_{d',t+1},$$

$$u_{e'}(t+1) = \beta_6 x_{6,t+1} + \beta_7 x_{7,t+1} + \beta_8 x_{8,t+1} + \beta_9 x_{9,t+1} + \beta_{10} x_{10,t+1} + \varepsilon_{e't+1},$$

$$u_{f'}(t+1) = \beta_6 x_{6,t+1} + \beta_7 x_{7,t+1} + \beta_8 x_{8,t+1} + \beta_9 x_{9,t+1} + \beta_{10} x_{10,t+1} + \varepsilon_{f't+1},$$

$$u_{g'}(t+1) = \beta_{11} x_{11,t+1} + \beta_{12} x_{12,t+1} + \beta_{13} x_{13,t+1} + \beta_{14} x_{14,t+1} + \beta_{15} x_{15,t+1} + \varepsilon_{g',t+1},$$

$$u_{i'}(t+1) = \beta_1 x_{1,t+1} + \beta_2 x_{2,t+1} + \beta_3 x_{3,t+1} + \beta_4 x_{4,t+1} + \beta_5 x_{5,t+1} + \varepsilon_{i',t+1},$$

and

$$u_{h'}(t+1) = \beta_{11} x_{11,t+1} + \beta_{12} x_{12,t+1} + \beta_{13} x_{13,t+1} + \beta_{14} x_{14,t+1} + \beta_{15} x_{15,t+1} + \varepsilon_{h',t+1}. \quad (37)$$

Denote n as the n-th variable (n=1~15). If alternative a' is chosen at t+1, $x_{m,t+1} = x_{m,t} - 1$ for $m = 1, 4$ and 5, $x_{m,t+1} = x_{m,t} + 1$ for $m = 8, 9$ and 10, and $x_{m,t+1} = x_{m,t}$ for $m = 2, 3, 6, 7,$ and $11-15$. If alternative b' is chosen at time t+1, $x_{m,t+1} = x_{m,t} - 1$ for $m = 8, 9$ and 10, $x_{m,t+1} = x_{m,t} + 1$ for $m = 13, 14$ and 15, and $x_{m,t+1} = x_{m,t}$ for $m = 1-7, 11$ and 12. If alternative c' is chosen at time t+1, $x_{m,t+1} = x_{m,t} - 1$ for $m = 13, 14$ and 15, $x_{m,t+1} = x_{m,t} + 1$ for $m = 1, 4$ and 5, and $x_{m,t+1} = x_{m,t}$ for $m = 2, 3,$ and $6-12$.

The values of the 15 parameters are generated from a (N~(0,1)) distribution and shown in Table 1 below. Only the first 10 parameters are used in case 1 and all 15 parameters are used in case 2.

Table 1. Values of Population Parameters

β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}	β_{14}	β_{15}
2.08	-0.66	-0.68	1.6	1.30	-0.5	-0.38	1.5	1.19	2.29	0.7	-1.88	1.24	1.78	2.2

4.3 Estimation Method

Our estimation method for case 1 is relatively simple, as it only involves a binary probit model for both serial independence and serial dependence. Because only the ratio of β and the corresponding standard deviation (δ for serial independence and δ_1 for serial dependence) can be estimated, we simply set $\delta = \delta_1 = 1$ during estimation.

The estimation for case 2 is much more complex and given as follows. The log-likelihood function for an individual is (the notation for individual is suppressed):

$$LogLi = d_{a'}(t) Ln(\Pr(d_{a'}(t) = 1)) + d_{b'}(t) Ln(\Pr(d_{b'}(t) = 1)) + d_{c'}(t) Ln(\Pr(d_{c'}(t) = 1)).$$

To perform the Maximum Likelihood Estimation method, we need to calculate $\Pr(d_{a'}(t) = 1)$, $\Pr(d_{b'}(t) = 1)$, and $\Pr(d_{c'}(t) = 1)$ for the probit and the mixed logit models.

4.3.1 Probit Model

Recall that in a probit model, we have:

$$\Pr(d_{a'}(t)=1) = \int_{-\infty}^{v_{b'}^{'}-v_{a'}^{'}} \int_{-\infty}^{v_{c'}^{'}-v_{a'}^{'}} f(\hat{\Sigma}_{a'}) d(\varepsilon_{b'}^{'}-\varepsilon_{a'}^{'}) d(\varepsilon_{c'}^{'}-\varepsilon_{a'}^{'}),$$

$$\Pr(d_{b'}(t)=1) = \int_{-\infty}^{v_{c'}^{'}-v_{b'}^{'}} \int_{-\infty}^{v_{a'}^{'}-v_{b'}^{'}} f(\hat{\Sigma}_{b'}) d(\varepsilon_{c'}^{'}-\varepsilon_{b'}^{'}) d(\varepsilon_{a'}^{'}-\varepsilon_{b'}^{'}),$$

$$\Pr(d_{c'}(t)=1) = \int_{-\infty}^{v_{b'}^{'}-v_{c'}^{'}} \int_{-\infty}^{v_{a'}^{'}-v_{c'}^{'}} f(\hat{\Sigma}_{c'}) d(\varepsilon_{b'}^{'}-\varepsilon_{c'}^{'}) d(\varepsilon_{a'}^{'}-\varepsilon_{c'}^{'}). \quad (38)$$

As shown from the above equations, the calculation of $\Pr(d_{a'}(t)=1)$, $\Pr(d_{b'}(t)=1)$, and $\Pr(d_{c'}(t)=1)$ requires the estimation of $\hat{\Sigma}_{a'}$, $\hat{\Sigma}_{b'}$, and $\hat{\Sigma}_{c'}$. Let P be the Choleski factor of matrix $\hat{\Sigma}_{a'}$, $\hat{\Sigma}_{a'} = \mathrm{PP'}$, and $\mathrm{P} = \begin{bmatrix} 1 & 0 \\ c_{12} & c_{22} \end{bmatrix}$. The two elements in P can be identified along with the estimation of β. Because the elements of $\hat{\Sigma}_{b'}$ and $\hat{\Sigma}_{c'}$ are all available from $\hat{\Sigma}_{a'}$, we can then write $\hat{\Sigma}_{b'} = f(\hat{\Sigma}_{a'})$ and $\hat{\Sigma}_{c'} = f(\hat{\Sigma}_{a'})$. In other words, $\hat{\Sigma}_{b'}$ and $\hat{\Sigma}_{c'}$ can be retrieved from $\hat{\Sigma}_{a'}$. They are then calculated as follows:

$$\hat{\Sigma}_{b'} = \begin{bmatrix} 1 & 1-c_{12} \\ 1-c_{12} & 1+c_{12}^2+c_2^2-2c_{12} \end{bmatrix}, \quad \text{and}$$

$$\hat{\Sigma}_{c'} = \begin{bmatrix} c_{12}^2+c_{22}^2 & c_{12}^2+c_2^2-c_{12} \\ c_{12}^2+c_2^2-c_{12} & 1+c_{12}^2+c_2^2-2c_{12} \end{bmatrix}.$$

The estimation of $\Pr(d_{a'}(t)=1)$, $\Pr(d_{b'}(t)=1)$, and $\Pr(d_{c'}(t)=1)$ is performed with the GHK simulator [8]. Using $\Pr(d_{a'}(t)=1)$ as an example, the procedures behind the GHK simulator can be briefly described as follows:

1. Calculate $k = \Phi(v_{a'}{}' - v_{b'}{}')$,
2. Draw a standard uniform random number [18],
3. Calculate $\eta_r = \Phi^{-1}(\mu_r \phi(v_{a'}{}' - v_{b'}{}'))$,
4. Calculate $g_r = \Phi(-(v_{c'}{}' - v_{a'}{}' + c_{12}\eta_r)/c_{22})$,
5. Calculate $P_r = kg_r$, and
6. Repeat steps 1~5 R times (R is pre-defined by researchers. We use 100 in this study.), and average $\mathrm{P_r}$ over R give us the simulated $\Pr(d_{a'}(t)=1)$.

4.3.2 Mix Logit Model

Recall that in a mix logit model, we have:

$$\Pr(d_{a'}(t)) = \iint \frac{1}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})} f(\tilde{\Omega}) d\eta_{j'} d\eta_{k'},$$

$$\Pr(d_{b'}(t)) = \iint \frac{\exp(v_{b'}^{'}+\eta_{b'})}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})} f(\tilde{\Omega}) d\eta_{j'} d\eta_{k'},$$

$$\Pr(d_{c'}(t)) = \iint \frac{\exp(v_{c'}^{'}+\eta_{c'})}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})} f(\tilde{\Omega}) d\eta_{j'} d\eta_{k'},$$

$$j' \neq a', j' \neq k'. \quad (39)$$

Similar to the probit model setting, we use the Choleski factor P' and $\tilde{\Omega}$ =PP'. The two elements in P' are estimable along with β. The simulation involves the following steps:

1. Draw R^3 pairs of random numbers: $\eta_{j'}$ and $\eta_{k'}$, with variance and covariance matrix $\tilde{\Omega}$,
2. Calculate $\frac{1}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})}$, $\frac{\exp(v_{b'}^{'}+\eta_{b'})}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})}$, and $\frac{\exp(v_{c'}^{'}+\eta_{c'})}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})}$ for each repetition,
3. Average $\frac{1}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})}$, $\frac{\exp(v_{b'}^{'}+\eta_{b'})}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})}$, and $\frac{\exp(v_{c'}^{'}+\eta_{c'})}{1+\sum_{j'} \exp(v_{j'}^{'}+\eta_{j'})}$ over R, which gives the simulated probabilities.

4.3.3 Results

With a sample comprising 500 observations, the estimation results for case 1 (serial independence and dependence) are shown in Tables 2 and 3. Given that the dataset is generated with the component of future utility, Tables 2 and 3 also show the estimation results if a standard binary probit model with no component of future utility is applied. Note that the true values shown in Tables 2 and 3 are different from each other. This is because, although the values of β are chosen to be the same, the standard deviation is different between the subcase of serial independence (δ) and the subcase of serial dependence (δ_1).

As shown in Tables 2 and 3, for both serial independence and serial dependence, the binary probit model with the component of future utility recovers all 10 parameters very well and all estimated coefficients are found to be statistically significant at 5% significance level. However, if a binary probit model with no component of future utility is applied to a dataset generated with the future utility component, the estimated parameters are found to attenuate toward zero. This attenuation is expected because of the correlation between the systematic utility component and the error component [7].

Table 2. Estimation Results for Case 1 (Serial Independence)
With and Without Future Utility Component

		With Future Utility		Without Future Utility	
		500 Sample		*500 Sample*	
Parameter	True_value	Estimates	Std. error	Estimates	Std. Error[1]
β_1/δ	1.06	1.19	0.12	0.14	0.06
β_2/δ	-0.33	-0.41	0.06	-0.03	0.06
β_3/δ	-0.34	-0.35	0.06	-0.08	0.06
β_4/δ	0.82	0.88	0.10	0.03	0.05
β_5/δ	0.66	0.72	0.08	0.01	0.06
β_6/δ	-0.26	-0.30	0.06	-0.01	0.06
β_7/δ	-0.19	-0.19	0.05	-0.04	0.06
β_8/δ	0.77	0.83	0.09	0.13	0.06
β_9/δ	0.61	0.58	0.07	0.04	0.06
β_{10}/δ	1.17	1.22	0.12	0.17	0.06

[1] If we retain 3 digits after the decimal point, the standard error differs from one variable to another.

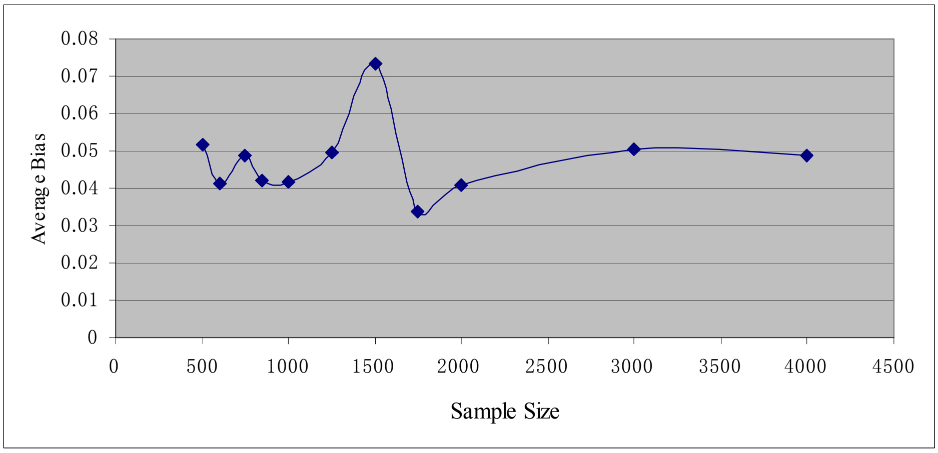

Fig. 3. Average Bias of With-Future Model for Case 1 (Serial Independence) by Sample Size.

Table 3. Estimation Results for Case 1 (Serial Dependence)
With and Without Future Utility Component

		With Future Utility		Without Future Utility	
		500 Sample		*500 Sample*	
Parameter	True_value	Estimates	Std. error	Estimates	Std. Error[1]
β_1/δ_1	0.86	0.74	0.07	0.05	0.06
β_2/δ_1	-0.27	-0.26	0.04	-0.07	0.06
β_3/δ_1	-0.28	-0.24	0.04	-0.04	0.06
β_4/δ_1	0.66	0.63	0.06	-0.01	0.06
β_5/δ_1	0.54	0.41	0.05	0.05	0.06
β_6/δ_1	-0.21	-0.24	0.04	0.05	0.06
β_7/δ_1	-0.16	-0.13	0.05	0.05	0.06
β_8/δ_1	0.62	0.58	0.05	0.10	0.06
β_9/δ_1	0.49	0.44	0.04	0.10	0.06
β_{10}/δ_1	0.94	0.89	0.07	0.09	0.06

[1] If we retain 3 digits after the decimal point, the standard error differs from one variable to another.

To analyze the effect of sample size on the estimates, we also run the with-future binary probit model with 11 different sample sizes, ranging from 500 to 4000. The average bias is calculated as the sum of the absolute values of the difference between the estimated parameter and the true parameter value. As shown in Figs. 3 and 4, the 500 sample size appears to be adequate for both serial independence and serial dependence; increasing sample size does not appear to lower the average bias significantly.

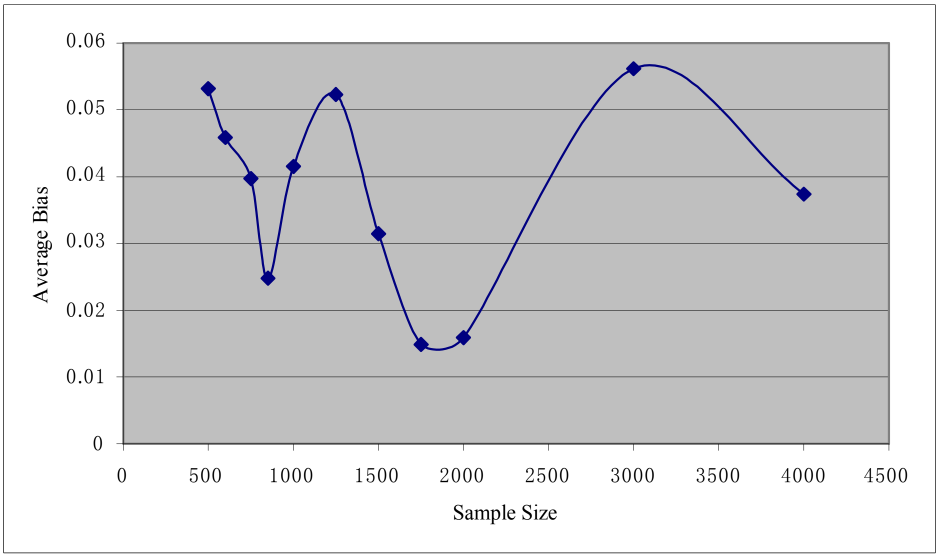

Fig. 4. Average Bias of With-Future Model for Case 1 (Serial Dependence) by Sample Size.

Table 4. Estimation Results of Probit and Mixed Logit Models for Case 2 With and Without Future Utility Component

		w. Future Utility				w.o. Future Utility	
		500 Sample				*500 Sample*	
		Probit		Mixed Logit[2]			
Parameter	True_value	Est.	S.E.	Est.	S.E.	Est.	S.E.[1]
$\beta_1 / pbr_{b'-a;}$	0.97	0.96	0.12	1.04	0.29	0.25	0.05
$\beta_2 / pbr_{b'-a;}$	-0.31	-0.42	0.10	-0.45	0.21	-0.03	0.05
$\beta_3 / pbr_{b'-a;}$	-0.32	-0.37	0.09	-0.41	0.20	-0.14	0.05
$\beta_4 / pbr_{b'-a;}$	0.75	0.67	0.12	0.75	0.27	0.24	0.05
$\beta_5 / pbr_{b'-a;}$	0.61	0.59	0.08	0.66	0.23	0.15	0.05
$\beta_6 / pbr_{b'-a;}$	-0.23	-0.29	0.09	-0.32	0.19	0.09	0.07
$\beta_7 / pbr_{b'-a;}$	-0.18	-0.25	0.09	-0.24	0.18	-0.12	0.07
$\beta_8 / pbr_{b'-a;}$	0.70	0.77	0.11	0.77	0.22	0.006	0.07
$\beta_9 / pbr_{b'-a;}$	0.56	0.61	0.10	0.63	0.21	0.15	0.07
$\beta_{10} / pbr_{b'-a;}$	1.07	0.85	0.12	0.87	0.26	0.11	0.07
$\beta_{11} / pbr_{b'-a;}$	0.33	0.24	0.09	0.22	0.17	-0.03	0.07
$\beta_{12} / pbr_{b'-a;}$	-0.88	-0.88	0.14	-0.77	0.26	-0.46	0.07
$\beta_{13} / pbr_{b'-a;}$	0.58	0.58	0.10	0.48	0.21	0.18	0.06
$\beta_{14} / pbr_{b'-a;}$	0.83	0.91	0.13	0.78	0.25	0.51	0.07
$\beta_{15} / pbr_{b'-a;}$	1.03	0.95	0.13	0.82	0.24	0.53	0.06

[1] If we retain 3 digits after the decimal point, the standard error differs from one variable to another at a larger magnitude. [2] To compare estimates between the mixed logit model and the probit model, the estimates from the mixed logit model need to be divided by $(1+\pi^2/3)$.

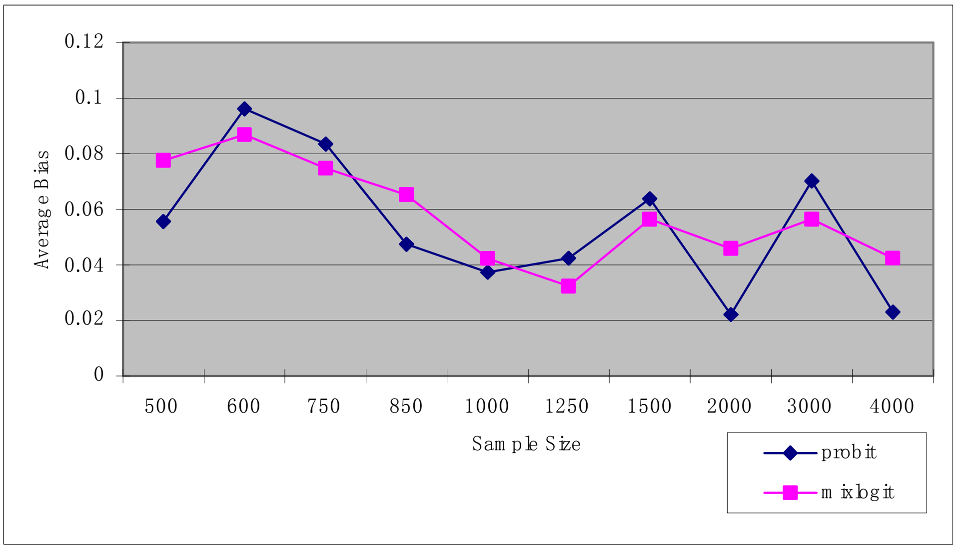

Fig. 5. Average Bias of With-Future Probit and Mixed Logit Models by Sample Size.

Table 4 shows the estimation results of the with-future probit and mixed logit models as well as a without-future probit model applied in case 2. In terms of the estimated parameters, both the probit model and the mixed logit model are able to recover true parameter values quite well. If a without-future model is applied to a dataset generated with the future utility component, the estimated parameters cannot only attenuate to zero but also have wrong signs. What is worthy to note is that the standard errors of the estimated parameters in the mixed logit model are almost twice as large as those in the probit model. This is probably because the dataset is generated with normal distributions while the mixed logit model applies an extreme value distribution.

Fig. 5. shows the average bias of with-future probit and mixed logit model by sample size. The average bias is calculated in the same way as described in case 1. In terms of average bias, the performance of probit model and mixed logit model is comparable with each other. At the 500 sample size, both models have a larger average bias compared to the two models in case 1. This is probably because of the more complicated variance-covariance structure involved in case 2. In addition, it appears that, from Fig. 5, there is a trend of decreasing average bias as the sample size increases. The average bias appears to drop under 0.06 (a bias size that is comparable to the two models in case 1, as shown in Figs. 3 and 4) when the sample size increases to 1000.

5. CONCLUSIONS

Because there is ample evidence showing that people do anticipate future in their decision making processes, it is important to include the future utility component in our modeling practices for certain choice scenarios. These decision making scenarios might include: vehicle holding and transaction decisions, residential location choices, and job location choices. Despite the tremendous development of discrete choice models in the past decade, it remains an open question in terms of how the current modeling frameworks can be modified to account for expectation of the future utility. The purpose of this paper is to provide some answers to this open question.

The paper demonstrates monte carlo simulations for two cases. Case 1 depicts a scenario where the decision maker faces two alternatives at time t and each alternative branches out two alternatives at time t+1. In case 1, we also distinguish between serial independence and serial dependence. Case 2 depicts a scenario where the decision maker faces three alternatives at time t and each alternative branches out two alternatives at time t+1. Case 2 is a general case, as we allow correlations between different alternatives. Furthermore, although case 2 only involves three alternatives, its framework can be treated as general framework for n alternatives. This is because having more alternatives will only involve more complex variance-covariance structures and require more integrals. The principle behind the estimation is the same.

Our simulation results illustrate three points mainly. First, if a without-future model is applied to a dataset generated with the future utility component, the estimated parameters attenuate to zero. In more complicated choice scenarios involving more than 2 alternatives and correlation between alternatives, the estimated parameters also have wrong signs. Second, if a with-future model can be correctly specified on a dataset generated with the future utility component, the above-described simulation methodologies (both the probit model and the mixed logit model) can recover true parameters quite well. Third, the use of a model form (probit vs. mixed logit) should be cautious. As shown in case 2, although the mixed logit model recovers true parameters equally well as the probit model, the standard errors are enlarged.

This study currently assumes that β remains unchanged over time, which can be unrealistic in some cases. Future research effort is to relax this assumption so that β can change over time. In addition to making β vary over time, one can also make β vary with different individuals. In this case, β is treated as a vector of random parameters.

Since the basic properties of dynamic choice scenarios are known, an additional future research effort should be to implement the model with an empirical dataset. As mentioned in Introduction, a number of choice scenarios would be suitable for this type of model, for example, vehicle holding/transaction choices and residential/job location choices. This requires further research.

References

[1] Aarts, H., Verplanken, B., and van Knippenberg, A., "Habit and Information Use in Travel Mode Choices", *Acta Psychologica*, Vol. 96, 1997, pp. 1–14.

[2] Aarts, H., Verplanken, B., and van Knippenberg, A., "Predicting Behavior from Actions in the Past: Repeated Decision Making or a Matter of Habit?", *Journal of Applied Social Psychology*, Vol. 28, No.15, 1998, pp. 1355–1374.

[3] Bhat, C., and Gossen, R., "A Mixed Multinomial Logit Model Analysis of Weekend Recreational Episode Type Choice", *Transportation Research Part B*, Vol. 38, No. 9, 2004, pp. 767-787.

[4] Bunch, D., "Estimability in the Multinomial Probit Model", *Transportation Research Part B*, Vol. 25, No.1, 1991, pp. 1-12.

[5] Chen, C., Garling, T., and Kitamura, R., "Activity Rescheduling: Deliberate or Habitual?", *Transportation Research Part F*, Vol. 7, No.6, 2004, pp.351-371.

[6] Crane, R. (1996) "The Influence of Uncertain Job Location on Urban Form and the Journal to Work", *Journal of Urban Economics*, Vol. 39, 1996, pp.342-356.

[7] W. Greene, *Econometric Analysis*, Fifth Edition, Prentice Hall, New Jersey, 2003

[8] Hajivassiliou, V., McFadden, D., Ruud, P., "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives: Theoretical and Computational Results", *Journal of Econometrics*, Vol. 72, 1996, pp. 85-134.

[9] Hess, S. and Polak, J. "Mixed Logit Modelling of Airport Choice in Multi-airport Regions", *Journal of Air Transport Management*, Vol.11, No.2, 2005, pp. 59-68.

[10] Hocherman, I., Prashker, J., & Ben-Akiva, M., "Estimation and Use of Dynamic Transaction Models of Automobile Ownership", *Transportation Research Record*, Vol. 944, 1983, pp. 134–141.

[11] Kan, K., "Residential Mobility with Job Location Uncertainty", *Journal of Urban Economics*, Vol. 52, 2002, pp. 501-523.

[12] Kitamura, R., Chen, C., Pendyala, R., "Generation of Synthetic Activity-Travel Patterns", *Transportation Research Record*, Vol. 1607, 1997, pp. 154-162.

[13] J., Ma, K. G., Goulias, "Forecasting Home Departure Time, Daily Time Budget, Activity Duration, and Travel Time Using Panel Data", *77th Annual Conference of Transportation Research Board*. 1998, Washington, D. C., U.S.A.

[14] Sándor, Z. and Train, K., "Quasi-random simulation of discrete choice models", *Transportation Research Part B*, Vol. 38, No.4, 2004, pp. 313-327.

[15] Smith, N., Hensher, D., Wrigle, N., 1991. "A Discrete Choice Sequence model: Method and an Illustrative Application to Automobile Transactions", *International Journal of Transport Economics*, Vol. XVIII, No.2, 1991, pp. 123–150.

[16] Wachs, M., Taylor, B.D., Levine, N., & Ong, P., "The Changing Commute: A Case-Study of the Jobs-Housing Relationship Over Time", *Urban Studies*, Vol. 30 No. 10, 1993, pp. 1711-1729.

[17] Weeks, M., "The Multinomial Probit Model Revisited: A Discussion of Parameter Estimability, Idnetification and Specification Testing", *Journal of Economic Surveys*, Vol. 11, No. 3, 1997, pp. 297-320.

[18] http://www.ce.utexas.edu/prof/bhat/fULL_PAPERS.htm

Appendix: Notation List for Symbols Used in the Paper

ω : discount factor (in Eq. (1)),

β: the parameter vector associated with the independent variables in the lifetime systematic utility,

$d_a(t)$: indicator function of choosing a at t in case 1 (in Eq. (4)),

$d_b(t)$: indicator function of choosing b at t in case 1 (in Eq. (4)),

$d_c(t+1)$: indicator function of choosing c at $t+1$ in case 1 (in Eq. (4)),

$d_d(t+1)$: indicator function of choosing d at $t+1$ in case 1 (in Eq. (4)),

$d_e(t+1)$: indicator function of choosing e at $t+1$ in case 1 (in Eq. (4)),

$d_f(t+1)$: indicator function of choosing f at $t+1$ in case 1 (in Eq. (4)),

$u_a(t)$: random utility of choosing a at t in case 1 (in Eq. (4)),

$v_a(t)$: systematic utility of choosing a at t in case 1 (in Eq. (4)),

$\varepsilon_a(t)$: random error term of choosing a at t in case 1 (in Eq. (4)),

$u_b(t)$: random utility of choosing b at t in case 1 (in Eq. (4)),

$v_b(t)$: systematic utility of choosing b at t in case 1 (in Eq. (4)),

$\varepsilon_b(t)$: random error term of choosing b at t in case 1 (in Eq. (4)),

$u_c(t+1)$: random utility of choosing c at $t+1$ in case 1 (in Eq. (4)),

$v_c(t+1)$: systematic utility of choosing c at $t+1$ in case 1 (in Eq. (4)),

$\varepsilon_c(t+1)$: random error term of choosing c at $t+1$ in case 1 (in Eq. (4)),

$u_d(t+1)$: random utility of choosing d at $t+1$ in case 1 (in Eq. (4)),

$v_d(t+1)$: systematic utility of choosing d at $t+1$ in case 1 (in Eq. (4)),

$\varepsilon_d(t+1)$: random error term of choosing d at $t+1$ in case 1 (in Eq. (4)),

$u_e(t+1)$: random utility of choosing e at $t+1$ in case 1 (in Eq. (4)),

$v_e(t+1)$: systematic utility of choosing e at $t+1$ in case 1 (in Eq. (4)),

$\varepsilon_e(t+1)$: random error term of choosing e at $t+1$ in case 1 (in Eq. (4)),

$u_f(t+1)$: random utility of choosing f at $t+1$ in case 1 (in Eq. (4)),

$v_f(t+1)$: systematic utility of choosing f at $t+1$ in case 1 (in Eq. (4) ,

$\varepsilon_f(t+1)$: random error term of choosing f at $t+1$ in case 1 (in Eq. (4)),

$\varepsilon_a^{'}$: lifetime error term of choosing a in case 1 (after Eq. (4)),

$\varepsilon_b^{'}$: lifetime error term of choosing b in case 2 (after Eq. (4)),

δ : standard deviation of $\varepsilon_b^{'} - \varepsilon_a^{'}$ in the subcase of serial independence in case 1 (before Eq. (7)),

δ_1 : standard deviation of $\varepsilon_b^{'} - \varepsilon_a^{'}$ in the subcase of serial dependence in case 1 (in Eq. (12)),

δ_a^2 : variance of $\varepsilon_a(t)$,

δ_b^2 : variance of $\varepsilon_b(t)$,

δ_c^2 : variance of $\varepsilon_c(t+1)$,

δ_d^2 : variance of $\varepsilon_d(t+1)$,

δ_e^2 : variance of $\varepsilon_e(t+1)$,

δ_f^2 : variance of $\varepsilon_f(t+1)$,

$\delta_{a,c}$: covariance of $\varepsilon_a(t)$ and $\varepsilon_c(t+1)$,

$\delta_{a,d}$: covariance of $\varepsilon_a(t)$ and $\varepsilon_d(t+1)$,

$\delta_{b,e}$: covariance of $\varepsilon_b(t)$ and $\varepsilon_d(t+1)$,

$\delta_{b,f}$: covariance of $\varepsilon_b(t)$ and $\varepsilon_f(t+1)$,

$d_{k'}(t)$ and $d_{j'}(t)$: indicator function of choosing s k' and j' at t in case 2, $\forall k', j' \in (a', b', c')$ (in Eq. (13)),

$u_{k'}(t)$ and $u_{j'}(t)$: random utility of choosing k' and j' , $\forall k', j' \in (a', b', c')$ (in Eq. (13)),

$u_{a'}'(t)$: lifetime random utility of choosing a' at t in case 2 (in Eq. (14)),

$v_{a'}'$: lifetime systematic utility of choosing a' at t in case 2 (in Eq. (15)),

$\varepsilon_{a'}'$: lifetime random error term of choosing a' at t in case 2 (in Eq. (15)),

$u_{b'}'(t)$: lifetime random utility of choosing b' at t in case 2 (in Eq. (14)),

$v_{b'}'$: lifetime systematic utility of choosing b' at t in case 2 (in Eq. (15)),

$\varepsilon_{b'}'$: lifetime random error term of choosing b' at t in case 2 (in Eq. (15)),

$u_{c'}'(t)$: lifetime random utility of choosing c' at t in case 2 (in Eq. (14)),

$v_{c'}'$: lifetime systematic utility of choosing c' at t in case 2 (in Eq. (15)),

$\varepsilon_{c'}'$: lifetime random error term of choosing c' t in case 2 (in Eq. (15)),

$d_{a'}(t)$: indicator function of choosing a' at t in case 2 (in Eq. (17)),

$d_{b'}(t)$: indicator function of choosing b' at t in case 2 (in Eq. (18)),

$d_{c'}(t)$: indicator function of choosing c' at t in case 2 (in Eq. (19)),

$d_{d'}(t+1)$: indicator function of choosing d' at $t+1$ in case 2 (in Eq. (15)),

$d_{e'}(t+1)$: indicator function of choosing e' at $t+1$ in case 2 (in Eq. (15)),

$d_{f'}(t+1)$: indicator function of choosing f' at $t+1$ in case 2 (in Eq. (15)),

$d_{g'}(t+1)$: indicator function of choosing g' at $t+1$ in case 2 (in Eq. (15)),

$d_{h'}(t+1)$: indicator function of choosing h' at $t+1$ in case 2 (in Eq. (15)),

$d_{i'}(t+1)$: indicator function of choosing i' at $t+1$ in case 2 (in Eq. (15)),

$v_{a'}(t)$: systematic utility of choosing a' at t in case 2 (in Eq. (15)),

$\varepsilon_{a'}(t)$: random error term of choosing a' at t in case 2 (in Eq. (15)),

$v_{b'}(t)$: systematic utility of choosing b' at t in case 2 (in Eq. (15)),

$\varepsilon_{b'}(t)$: random error term of choosing b' at t in case 2 (in Eq. (15)),

$v_{c'}(t)$: systematic utility of choosing c' at t in case 2 (in Eq. (15)),

$\varepsilon_{c'}(t)$: random error term of choosing c' at t in case 2 (in Eq. (15)),

$v_{d'}(t+1)$: systematic utility of choosing d' at $t+1$ in case 2 (in Eq. (15)),

$\varepsilon_{d'}(t+1)$: random error term of choosing d' at $t+1$ in case 2 (in Eq. (15)),

$v_{e'}(t+1)$: systematic utility of choosing e' at $t+1$ in case 2 (in Eq. (15)),

$\varepsilon_{e'}(t+1)$: random error term of choosing e' at $t+1$ in case 2 (in Eq. (15)),

$v_{f'}(t+1)$: systematic utility of choosing f' at $t+1$ in case 2 (in Eq. (15)),

$\varepsilon_{f'}(t+1)$: random error term of choosing f' at $t+1$ in case 2 (in Eq. (15)),

$v_{g'}(t+1)$: systematic utility of choosing g' at $t+1$ in case 2 (in Eq. (15)),

$\varepsilon_{g'}(t+1)$: random error term of choosing g' at $t+1$ in case 2 (in Eq. (15)),

$v_{h'}(t+1)$: systematic utility of choosing h' at $t+1$ in case 2 (in Eq. (15)),

$\varepsilon_{h'}(t+1)$: random error term of choosing h' at $t+1$ in case 2 (in Eq. (15)),

$v_{i'}(t+1)$: systematic utility of choosing i' at $t+1$ in case 2 (in Eq. (15)),

$\varepsilon_{i'}(t+1)$: random error term of choosing i' at $t+1$ in case 2 (in Eq. (15)),

$pbr_{b'-a'}$: variance of $\varepsilon_{b'}' - \varepsilon_{a'}'$ (after Eq. (20)),

$pbr_{c'-a',b'-a'}$: covariance of $\varepsilon_{c'}' - \varepsilon_{a'}'$ (after Eq. (20)),

$pbr_{c'-a'}$: variance of $\varepsilon_{c'}' - \varepsilon_{a'}'$ (after Eq. (20)),

$\Sigma_{a'}$: variance-covariance matrix of $\varepsilon_{b'}^{'} - \varepsilon_{a'}^{'}$ and $\varepsilon_{c'}^{'} - \varepsilon_{a'}^{'}$ (after Eq. (20)),

$\hat{\Sigma}_{a'}$: rescaled variance-covariance matrix of $\varepsilon_{b'}^{'} - \varepsilon_{a'}^{'}$ and $\varepsilon_{c'}^{'} - \varepsilon_{a'}^{'}$ (after Eq. (20)),

$\hat{\Sigma}_{b'}$: rescaled variance-covariance matrix of $\varepsilon_{a'}^{'} - \varepsilon_{b'}^{'}$ and $\varepsilon_{c'}^{'} - \varepsilon_{b'}^{'}$ (in Eq. (22)),

$\hat{\Sigma}_{c'}$: rescaled variance-covariance matrix of $\varepsilon'_{a'} - \varepsilon'_{c'}$ and $\varepsilon'_{b'} - \varepsilon'_{c'}$ (in Eq. (23)),

Ω : variance-covariance matrix of $\eta_{a'}$, $\eta_{b'}$ and $\eta_{c'}$ (in Eq. (25)),

$\hat{\Omega}$: rescaled variance-covariance matrix of $\eta_{a'}$, $\eta_{b'}$ and $\eta_{c'}$ (in Eq. (25)),

$\tilde{\Omega}$: rescaled variance-covariance matrix of $\eta'_{b'} - \eta'_{a'}$ and $\eta'_{c'} - \eta'_{a'}$ (in Eq. (35)), and

M : transforming matrix (in Eq. (35)).

Dr. Cynthia Chen received her Ph.D. degree in Civil and Environmental Engineering in 2001 from University of California at Davis. She is an assistant professor in the department of civil engineering at City College of New York. Her research activities focus on short-term activity and travel pattern analysis, long-term residential and job location choices, and vehicle transaction choices. Dr. Chen is active in a number of professional organizations. She is currently the chair of the subcommittee of the time use subcommittee at Transportation Research Board. Dr. Chen has also served on NSF panels as well as served as a reviewer for a number of journals, including Transportation Research Part A and Part B, Transportation, and Transportation Research Record. In the Department of Civil Engineering at City College, she is serving on the Ph.D. committee, representing the transportation group.

Xiaoqiang Chen is a Ph.D. candidate in the Department of Civil Engineering at City College of New York, City University of New York. He received his degree of Bachelor of Science from Fuzhou University in 1999 and degree of Master of Science from Tongji University in 2002 respectively. He began to pursue his degree of Doctor of Philosophy at City College of New York from 2004. His research interests include travel behavior analysis, travel demand modeling, and residential location behavior analysis. He is a member of International Chinese Transportation Professionals Association.