A PRACTICAL METHOD TO TEST THE VALIDITY OF THE STANDARD GUMBEL DISTRIBUTION IN LOGIT-BASED MULTINOMIAL CHOICE MODELS OF HUMAN TRAVEL BEHAVIOR

Xin Ye *(corresponding author)*
Tongji University, College of Transportation Engineering,
Key Laboratory of Road and Traffic Engineering of Ministry of Education
4800 Cao'an Road, Shanghai, China, 201804.
Tel: +86-21-59946091
Email: xye@tongji.edu.cn

Venu M. Garikapati
Arizona State University, School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85281
Tel: 480-522-8067; Email: venu.garikapati@asu.edu

Daehyun You
Maricopa Association of Governments
302 N. First Avenue, Suite 300, Phoenix, AZ 85003
Tel: (602) 254-6300; Email: dyou@azmag.gov

Ram M. Pendyala
Arizona State University, School of Sustainable Engineering and the Built Environment
660 S. College Avenue, Tempe, AZ 85281
Tel: 480-965-3589; Email: ram.pendyala@asu.edu

Submitted for Presentation only

Word count: 6450 text + 6 tables/figures x 250 = 7,950 words
96th Annual Meeting of the Transportation Research Board

Committee on Transportation Demand Forecasting (ADB40)

August 2016

ABSTRACT

Most multinomial choice models, particularly in practice (e.g., multinomial logit model), assume an extreme-value Gumbel distribution for the random components of utility functions. The use of this distribution offers a closed-form likelihood expression when the utility maximization principle is applied to model choice behaviors. The maximum likelihood estimation method can be easily applied to estimate model coefficients. However, maximum likelihood estimators are consistent and efficient only if distributional assumptions on the random error terms are valid. It is therefore important to test the validity of underlying distributional assumptions that form the basis of parameter estimation and policy evaluation. In this paper, a practical but strict method is proposed to test the distributional assumption of the random component of utility functions in both the multinomial logit (MNL) model and multiple discrete-continuous extreme value (MDCEV) model. Based on a semi-nonparametric approach, a closed-form likelihood function that nests the MNL or MDCEV model being tested is derived. Then, the traditional likelihood ratio test can be applied to test violations of the standard Gumbel distribution assumption. Simulation experiments are conducted to show that the test yields acceptable Type-I and Type-II error probabilities at commonly available sample sizes. The test is then applied to three real-world discrete and discrete-continuous choice models. For all three models, the proposed test rejects the validity of the standard Gumbel distribution in most utility functions, calling for the development of approaches that overcome adverse effects of violations of distributional assumptions.

Keywords: travel behavior models, discrete choice models, violations of distributional assumptions, test of validity of distributional assumption, multinomial logit model, multiple discrete-continuous extreme value model

1. INTRODUCTION

The Gumbel distribution (also called Type-I extreme value distribution) plays a central role in travel choice models, including both discrete choice models (McFadden, 1974) and multiple discrete-continuous choice models (e.g., Bhat, 2005 and Bhat, 2008). This can be attributed to two main reasons. First, the Gumbel distribution is close to the normal distribution; in the absence of any specific information about the behavioral phenomenon under investigation, it is often assumed in econometric choice models that the random disturbance term which captures the overall impact of unobserved factors is normally distributed. Second, when the Gumbel distribution is assumed for random components in utility functions, a closed-form expression for the likelihood function is obtained when the utility maximization principle is applied. With a neat closed-form expression for the likelihood function, maximum likelihood estimation (MLE) methods can be easily applied to estimate model coefficients consistently and efficiently.

Given these advantages, the Multinomial Logit (MNL) model and Multiple Discrete-Continuous Extreme Value (MDCEV) model, both of which are based on the Gumbel distribution assumption for the random error components, are widely used in practice. Although strides have been made in estimating model formulations that assume a normal distribution for the random error components, namely, the Multinomial Probit Model (Train, 2009) and Multiple Discrete-Continuous Probit (MDCP) model (Bhat et al., 2013), the logit-based models continue to be the model forms of choice for travel demand forecasting. However, the theory of maximum likelihood estimation indicates that the consistency and efficiency of maximum likelihood estimators depend on the validity of the distributional assumption made on the random error components. If the distributional assumption is violated, then the maximum likelihood estimators are neither consistent nor efficient.

In prediction settings, the MNL model ensures that predicted market shares match the observed shares in the sample (Ben-Akiva and Lerman, 1985). In the case of the MNL model, violations of the distributional assumption will therefore not adversely affect the predicted aggregate market shares. In the case of the MDCEV model, however, such a property does not hold. Jäggi et al. (2013) found that predictions from MDCEV models of vehicle fleet composition and usage are quite sensitive to model specification. As the unobserved but significant factors affecting vehicle fleet composition and usage are absorbed into the random error components, they are bound to influence the nature of the distribution of the random error terms. If the model specification results in a situation where there is violation of the standard Gumbel distributional assumption on the random error terms of the MDCEV model, it is reasonable to expect gross inaccuracies in model predictions depending on the severity of the violation.

It is therefore important to validate the assumed distributions on the random error terms prior to applying the MLE method to estimate model coefficients of either discrete or discrete-continuous travel choice models. The objective of this paper is to propose a practical but strict statistical method to test whether the error terms in random utility functions of MNL or MDCEV models follow the assumed Gumbel distribution.

2. LITERATURE REVIEW

Econometricians have been questioning the validity of the distributional assumption on random error components of utility functions ever since McFadden (1974) first proposed the multinomial logit model formulation (e.g., Manski, 1975). Concerns about error distribution violations motivated the development of semi-parametric and semi-nonparametric choice models. The semi-parametric choice

model employs the kernel density method to estimate the distribution of the random errors, and therefore does not rely on any parametric distributional assumptions (e.g., Klein and Spady, 1993; Lee, 1995). The semi-nonparametric (SNP) choice model is based on a polynomial approximation of a probability density function (PDF) that takes a flexible form (Gallant and Nychka, 1987). Because the likelihood function has an explicit analytical expression, the SNP choice modeling method appears to be applied more widely than the semi-parametric approach in practice (e.g., Chen and Randall, 1997; Creel and Loomis, 1997; Crooker and Herriges, 2007). In this paper, the SNP approach is used to derive a statistical test of the validity of the Gumbel distribution in logit models of discrete choice. It would therefore be prudent to first review the SNP binary choice model.

Similar to the binary probit model, the SNP binary choice model is also based on a random utility (U) function, which can be expressed as $\mathrm{U} = \mathrm{V} + \varepsilon$, where "V" is the systematic or deterministic component of the utility function and "ε" is the random component. If a dummy variable "y" indicates whether an alternative is chosen or not, then $\mathrm{P(y = 1) = P(U > 0) = P(V + \varepsilon > 0) = P(\varepsilon > -}V\mathrm{)}$. The probability density function takes the following form:

$$f(\varepsilon) = \frac{\left(\sum_{\mathrm{i}=0}^{\mathrm{K}} \mathrm{a_i}\varepsilon^{\mathrm{i}}\right)^2 \varphi(\varepsilon)}{\int_{-\infty}^{+\infty}\left(\sum_{\mathrm{i}=0}^{\mathrm{K}} \mathrm{a_i}\varepsilon^{\mathrm{i}}\right)^2 \varphi(\varepsilon)\mathrm{d}\varepsilon} \quad (1)$$

In Equation (1), $\varphi(\varepsilon)$ represents the PDF of the standard normal distribution. The denominator ensures that $\int_{-\infty}^{+\infty} f(\varepsilon)d\varepsilon = 1$. Equation (1) can be extended and written in the following form:

$$f(\varepsilon) = \frac{\left(\sum_{\mathrm{i}=0}^{\mathrm{K}}\sum_{\mathrm{j}=0}^{\mathrm{K}} \mathrm{a_i a_j}\varepsilon^{\mathrm{i+j}}\right)\varphi(\varepsilon)}{\int_{-\infty}^{+\infty}\left(\sum_{\mathrm{i}=0}^{\mathrm{K}}\sum_{\mathrm{j}=0}^{\mathrm{K}} \mathrm{a_i a_j}\varepsilon^{\mathrm{i+j}}\right)\varphi(\varepsilon)\mathrm{d}\varepsilon} \quad (2)$$

Then, $$\mathrm{P(y = 1) = P}(\varepsilon > -V) = \frac{\int_{-\mathrm{V}}^{+\infty}\left(\sum_{\mathrm{i}=0}^{\mathrm{K}}\sum_{\mathrm{j}=0}^{\mathrm{K}} \mathrm{a_i a_j}\varepsilon^{\mathrm{i+j}}\right)\varphi(\varepsilon)\mathrm{d}\varepsilon}{\int_{-\infty}^{+\infty}\left(\sum_{\mathrm{i}=0}^{\mathrm{K}}\sum_{\mathrm{j}=0}^{\mathrm{K}} \mathrm{a_i a_j}\varepsilon^{\mathrm{i+j}}\right)\varphi(\varepsilon)\mathrm{d}\varepsilon} = \frac{\sum_{\mathrm{i}=0}^{\mathrm{K}}\sum_{\mathrm{j}=0}^{\mathrm{K}} \mathrm{a_i a_j}\int_{-\mathrm{V}}^{+\infty}\varepsilon^{\mathrm{i+j}}\varphi(\varepsilon)\mathrm{d}\varepsilon}{\sum_{\mathrm{i}=0}^{\mathrm{K}}\sum_{\mathrm{j}=0}^{\mathrm{K}} \mathrm{a_i a_j}\int_{-\infty}^{+\infty}\varepsilon^{\mathrm{i+j}}\varphi(\varepsilon)\mathrm{d}\varepsilon} \quad (3)$$

For the probability value above, recursion formulas may be applied to compute the integral:

$$\int \varepsilon^{\mathrm{i+j}}\varphi(\varepsilon)\mathrm{d}\varepsilon. \quad (4)$$

When K = 0, the SNP model will reduce to a binary probit model. Thus, the SNP binary choice model nests the binary probit model as a special case, and can be used to validate the assumption of normality for the random error component in the binary probit model based on the likelihood ratio test. In the case of the logit model, it is possible to replace $\varphi(\varepsilon)$ in Equation (1) with the PDF of the logistic distribution to be tested. However, in that case, Equation (4) will not have a closed form expression, leading to considerable computational complexity. Another issue is that the SNP choice model is usually limited to a binary choice context rather than a multinomial choice context, possibly due to its computational complexity. It is therefore challenging to extend the original SNP approach to a multinomial choice modeling situation.

Bierens (2008) proposed a new polynomial, called the orthonormal Legendre polynomial, for estimating distributions semi-nonparametrically on the unit interval. In the transportation literature, this approach has been used to test normal and log-normal distributions of random coefficients in mixed logit model (Fosgerau and Bierlaire, 2007). In the method proposed in this paper, the orthonormal Legendre polynomial will be used to test the validity of the Gumbel distribution in both MNL and MDCEV models. It will be shown in this paper that the resultant likelihood function based on the new polynomial will have a closed form expression and nest those of the MNL and MDCEV models. A standard likelihood ratio test can be invoked to verify the validity of the underlying distributional assumption.

3. METHODOLOGY

3.1 The Semi-Nonparametric (SNP) Distribution Nesting the Standard Gumbel Distribution

According to Fosgerau and Bierlaire (2007) and Bierens (2008), the orthonormal Legendre polynomial can be recursively defined as:

$$\mathrm{L}_0 = 1,\ \mathrm{L}_1 = \sqrt{3}(2\mathrm{x} - 1) \tag{5}$$

$$\mathrm{L}_\mathrm{n} = \alpha(2\mathrm{x} - 1)\mathrm{L}_{\mathrm{n}-1} + \beta\mathrm{L}_{\mathrm{n}-2}\ ,\ \mathrm{n} \geq 2 \tag{6}$$

In Equation (6), $\alpha = \frac{\sqrt{4\mathrm{n}^2-1}}{\mathrm{n}}$ and $\beta = -\frac{(\mathrm{n}-1)\sqrt{2\mathrm{n}+1}}{\mathrm{n}\sqrt{2\mathrm{n}-3}}$.

The advantage of using this polynomial is to ensure that:

$$\int_0^1 L_m(x)L_n(x)dx = \begin{cases} 0 & if\ m \neq n \\ 1 & if\ m = n \end{cases}. \tag{7}$$

Then, the polynomial can be used to construct a semi-nonparametric probability density function that extends and nests the PDF of a standard Gumbel distribution as:

$$f(x) = \frac{\{1+\sum_{\mathrm{k}=1}^{\mathrm{K}} \delta_\mathrm{k}\mathrm{L}_\mathrm{k}[\mathrm{G}(\mathrm{x})]\}^2}{1+\sum_{\mathrm{k}=1}^{\mathrm{K}} {\delta_\mathrm{k}}^2}\mathrm{g}(\mathrm{x}), \tag{8}$$

where $\mathrm{g}(\mathrm{x}) = \exp(-\mathrm{e}^{-\mathrm{x}}) \cdot \exp(-\mathrm{x})$, $\mathrm{G}(\mathrm{x}) = \exp(-\mathrm{e}^{-\mathrm{x}})$, and δ_k are parameters. Note that the functional expression "exp(x)" is equivalent to "e^{x}". Based on Equation (7), it is easy to show that $\int_{-\infty}^{+\infty} f(x) = 1$. To test the standard Gumbel distribution, only consider the situation where K = 1 and simplify the formula as:

$$f(x) = \frac{\{1+\delta_1[\gamma_1+\gamma_2\mathrm{G}(\mathrm{x})]\}^2}{1+{\delta_1}^2}\mathrm{g}(\mathrm{x}) = \left[\frac{(1+\delta_1\gamma_1)^2}{1+{\delta_1}^2} + \frac{2(1+\delta_1\gamma_1)\delta_1\gamma_2}{1+{\delta_1}^2}\mathrm{G}(\mathrm{x}) + \frac{(\delta_1\gamma_2)^2}{1+{\delta_1}^2}\mathrm{G}(\mathrm{x})^2\right]\mathrm{g}(\mathrm{x})$$

$$= [\xi_0 + \xi_1\mathrm{G}(\mathrm{x}) + \xi_2\mathrm{G}(\mathrm{x})^2]\mathrm{g}(\mathrm{x}) = \{\textstyle\sum_{\mathrm{m}=0}^{2} \xi_\mathrm{m}[G(x)]^\mathrm{m}\}\mathrm{g}(\mathrm{x})\ . \tag{9}$$

In the formula above, $\xi_0 = \frac{(1+\delta_1\gamma_1)^2}{1+{\delta_1}^2}$, $\xi_1 = \frac{2(1+\delta_1\gamma_1)\delta_1\gamma_2}{1+{\delta_1}^2}$, $\xi_2 = \frac{(\delta_1\gamma_2)^2}{1+{\delta_1}^2}$, $\gamma_1 = -\sqrt{3}$, and $\gamma_2 = 2\sqrt{3}$. The cumulative distribution function (CDF) of the extended distribution is given as $F(\mathrm{x}) = \int_{-\infty}^{\mathrm{x}}\{\sum_{\mathrm{m}=0}^{2} \xi_\mathrm{m}[\mathrm{G}(\varepsilon)]^\mathrm{m}\}\mathrm{g}(\varepsilon)\mathrm{d}\varepsilon$. Since $\int_{-\infty}^{x}[\mathrm{G}(\varepsilon)]^\mathrm{m}\mathrm{g}(\varepsilon)\,d\varepsilon = \int_{-\infty}^{x}[\mathrm{G}(\varepsilon)]^\mathrm{m}\,d\mathrm{G}(\varepsilon)$, and letting z = $\mathrm{G}(\varepsilon)$, $\int_{-\infty}^{x}[\mathrm{G}(\varepsilon)]^\mathrm{m}\mathrm{g}(\varepsilon)\,d\varepsilon = \int_0^{G(x)} \mathrm{z}^\mathrm{m}\,d\mathrm{z} = \frac{[\mathrm{G}(x)]^{\mathrm{m}+1}}{\mathrm{m}+1}$. Thus, one should have that:

$$F(\mathrm{x}) = \textstyle\sum_{\mathrm{m}=0}^{2}\left\{\frac{\xi_\mathrm{m}[\mathrm{G}(\mathrm{x})]^{\mathrm{m}+1}}{\mathrm{m}+1}\right\}. \tag{10}$$

3.2 Test for Validity of Distributional Assumption in the Multinomial Logit (MNL) Model

In a discrete choice model, it is assumed that $\mathrm{U_j} = \mathrm{V_j} + \varepsilon_\mathrm{j}$, where the index of alternatives, j = 1, 2,..., J. In the interest of brevity, the index "i", corresponding to the individual decision-maker, is suppressed in the equation above. In an MNL model, ε_j is independently and identically distributed (i.i.d.) as a standard Gumbel distribution. The proposed method can be used to test whether any random error term, ε_j, follows the standard Gumbel distribution or not. Without any loss of generality, if the error term of the first alternative needs to be tested, it is possible to calculate:

$$q_1(\mathrm{m}) = \frac{\mathrm{e}^{\mathrm{V}_1}}{\mathrm{m}\cdot\mathrm{e}^{\mathrm{V}_1}+\sum_{\mathrm{j}=1}^{\mathrm{J}} \mathrm{e}^{\mathrm{V_j}}}. \tag{11}$$

For the k^{th} alternative not being tested (k > 1), proceed to calculate:

$$q_{\mathrm{k}}(\mathrm{m}) = \frac{e^{V_{\mathrm{k}}}}{(1+\mathrm{m})\cdot\left(\mathrm{m}\cdot e^{V_1} + \sum_{\mathrm{j}=1}^{\mathrm{J}} e^{V_{\mathrm{j}}}\right)} \ . \quad (12)$$

Then, calculate the choice probability for each alternative as:

$$P_{\mathrm{j}} = \sum_{\mathrm{m}=0}^{2} \xi_{\mathrm{m}} q_{\mathrm{j}}(\mathrm{m}), \quad (13)$$

where $\xi_0 = \frac{(1+\delta_1\gamma_1)^2}{1+\delta_1^{\ 2}}$, $\xi_1 = \frac{2(1+\delta_1\gamma_1)\delta_1\gamma_2}{1+\delta_1^{\ 2}}$, $\xi_2 = \frac{(\delta_1\gamma_2)^2}{1+\delta_1^{\ 2}}$, $\gamma_1 = -\sqrt{3}$, and $\gamma_2 = 2\sqrt{3}$.

The log-likelihood function over the entire sample can then be formulated as:

$$LL = \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{\mathrm{j}=1}^{\mathrm{J}} y_{\mathrm{ij}} ln\left(P_{\mathrm{ij}}\right), \quad (14)$$

where y_{ij} are dummy variables indicating whether the j[th] alternative is chosen by the individual "i". If the coefficient δ_1 is fixed at 0, $P_{\mathrm{ij}} = \frac{e^{V_{\mathrm{ij}}}}{\sum_{\mathrm{j}=1}^{\mathrm{J}} e^{V_{\mathrm{ij}}}}$ and the model reduces to an MNL model. Thus, the likelihood ratio test can be applied to test the null hypothesis that the random error in the utility function of the first alternative follows the standard Gumbel distribution (a complete mathematical derivation is given in Appendix A).

3.3 Test of Validity for the Multiple Discrete-Continuous Extreme Value (MDCEV) Model

As per Bhat (2008), the utility function of an MDCEV model takes the following form:

$$\mathrm{U_j} = \frac{\gamma_{\mathrm{j}}}{\alpha_{\mathrm{j}}} \psi_{\mathrm{j}} \left[\left(\frac{t_{\mathrm{j}}}{\gamma_{\mathrm{j}}} + 1 \right)^{\alpha_{\mathrm{j}}} - 1 \right], \quad (15)$$

where α_{j} is a satiation parameter that accounts for diminishing marginal utility and γ_{j} is a translation parameter that accommodates corner solutions (zero consumption of certain alternatives). The baseline utility $\psi_{\mathrm{j}} = e^{V_{\mathrm{j}}+\varepsilon_{\mathrm{j}}}$ and $\mathrm{t_j}$ represents the continuous resource being allocated to the alternative "j". The index of alternatives, j = 1, 2, ..., K, where "K" is the total number of alternatives and "M" represents the total number of alternatives that are allocated resources (M ≤ K). In the interest of brevity, the index denoting the individual "i" is dropped from the equation above. Suppose it is of interest to test the distributional assumption on random error ε_1 of the utility function of the first alternative. If $\mathrm{t_1} > 0$, calculate:

$$q(\mathrm{m}) = \left(\prod_{\mathrm{j}=1}^{\mathrm{M}} \mathrm{c_j}\right) \cdot \left(\sum_{\mathrm{j}=1}^{\mathrm{M}} \frac{1}{\mathrm{c_j}}\right) \cdot (\mathrm{M}-1)! \cdot \frac{\prod_{\mathrm{j}=1}^{\mathrm{M}} e^{V_{\mathrm{j}}}}{\left[\mathrm{m}\cdot e^{V_1} + \sum_{\mathrm{k}=1}^{\mathrm{K}} e^{V_{\mathrm{k}}}\right]^{\mathrm{M}}}. \quad (16)$$

If $\mathrm{t_1} = 0$, calculate $q(\mathrm{m}) = \left(\prod_{\mathrm{j}=1}^{\mathrm{M}} \mathrm{c_j}\right) \cdot \left(\sum_{\mathrm{j}=1}^{\mathrm{M}} \frac{1}{\mathrm{c_j}}\right) \cdot (\mathrm{M}-1)! \cdot \frac{\prod_{\mathrm{j}=2}^{\mathrm{M}+1} e^{V_{\mathrm{j}}}}{(1+\mathrm{m})\left[\mathrm{m}\cdot e^{V_1} + \sum_{\mathrm{k}=1}^{\mathrm{K}} e^{V_{\mathrm{k}}}\right]^{\mathrm{M}}}$, (17)

where $\mathrm{c_j} = \frac{1-\alpha_{\mathrm{j}}}{\mathrm{t_j}+\gamma_{\mathrm{j}}}$. Then, one can compute the likelihood value for each individual as:

$$P = \sum_{\mathrm{m}=0}^{2} \xi_{\mathrm{m}} q(\mathrm{m}), \quad (18)$$

where $\xi_0 = \frac{(1+\delta_1\gamma_1)^2}{1+\delta_1^{\ 2}}$, $\xi_1 = \frac{2(1+\delta_1\gamma_1)\delta_1\gamma_2}{1+\delta_1^{\ 2}}$, $\xi_2 = \frac{(\delta_1\gamma_2)^2}{1+\delta_1^{\ 2}}$, $\gamma_1 = -\sqrt{3}$, and $\gamma_2 = 2\sqrt{3}$.

The log-likelihood function for the entire sample can be formulated as $LL = \sum_{\mathrm{i}=1}^{\mathrm{N}} ln\left(P_{\mathrm{i}}\right)$, which can be maximized to estimate model coefficients as well as the parameter δ_1. Similar to the case of the MNL model, if the parameter δ_1 is fixed at 0, the model will reduce to the MDCEV model. Therefore, the likelihood ratio test can be applied to test the null hypothesis that the random error term of the utility function of the first alternative follows the standard Gumbel distribution (a complete mathematical derivation is given in Appendix B).

4. Simulation Experiments

Before applying the methods for models estimated on travel survey data sets in an empirical context, the suitability of the methods was confirmed through the use of simulation experiments. This section describes results of the simulation experiments that were conducted to demonstrate the applicability of the proposed testing methods. The second objective of the simulation experiments is to determine the sample sizes required for controlling the probability of Type-I and Type-II errors in statistical testing. Travel survey sample sizes are often limited and it would be of value for modelers to be aware of the sample sizes required to apply the test proposed in this paper.

4.1 Simulation Experiments for Test of MNL Model

The experiment is designed with four alternatives; four random utility values are computed as:

$U_1 = 0.4 - 0.5 \times x_1 + \varepsilon_1$,

$U_2 = -0.5 - 0.4 \times x_2 + \varepsilon_2$,

$U_3 = -0.6 - 0.3 \times x_3 + \varepsilon_3$,

$U_4 = - 0.5 \times x_4 + \varepsilon_4$.

In the equations above, x_1, x_2, x_3, and x_4 follow an independent uniform distribution between 0 and 10. ε_2, ε_3, and ε_4 follow an independent standard Gumbel distribution [i.e. G(0,1)], while ε_1 follows either G(0,1) or a distribution other than G(0,1). Then, four dummy choice variables, say y_1, y_2, y_3, and y_4, are generated in accordance with the utility maximization principle:

$y_j = [U_j \geq \max(U_1, U_2, U_3, U_4)]$.

Both MNL and SNP models are first estimated and then the χ^2 statistic is computed as $2 \times (LL_{SNP} - LL_{MNL})$, where LL_{SNP} and LL_{MNL} are the log-likelihood values of the SNP and MNL models at convergence. At one degree of freedom, the critical χ^2 value is 3.84 at a 0.05 significance level. If the value of the χ^2 statistic is greater than 3.84, the null hypothesis that ε_1 follows the standard Gumbel distribution is rejected at a 95 percent confidence level. Two types of simulation experiments are conducted to investigate the probability of making Type-I and Type-II errors in the hypothesis tests. The Type-I error refers to the case where the null hypothesis is rejected when it is true, while the Type-II error refers to the case where the null hypothesis is not rejected when it is false.

To determine the probability of making a Type-I error, the true standard Gumbel distribution is generated for ε_1. Then simulation experiments are repeated 100 times and the frequency of rejection of the null hypothesis (even though it is true) is recorded. This frequency of erroneously rejecting the null hypothesis is used to estimate the probability of making a Type-I error. The first two rows of Table 1 show the estimated probabilities of making a Type-I error when the sample size is 200 (small) and 4000 (large) respectively. It is found that the Type-I error probability is less than 0.05 for both sample sizes. Results of the simulation experiment thus demonstrate that the probability of making a Type-I error is very small for the proposed testing method.

For examining the probability of making Type-II errors, ε_1 should follow a distribution other than the standard Gumbel distribution, G(0,1). Two types of normal distributions are chosen for ε_1 in this simulation experiment. One is the standard normal distribution, i.e., N(0,1), and the other is a normal distribution adjusted to have the same expectation and standard deviation as that of a standard Gumbel distribution. Figure 1 compares the PDFs of the three distributions. Based on a visual examination, the standard normal distribution seems closer to the standard Gumbel distribution than the adjusted normal distribution. Distributions that are very similar to one another are chosen for this experiment to examine the statistical power of the proposed test. The statistical power can

be defined as the probability of correctly rejecting the null hypothesis, i.e., P(Reject H_0|H_0 is wrong), which is equal to [1 - P(Type-II error)]. Through simulation experiments, it is found that the probability of making Type-II errors (or statistical power of the test) depends on both the sample size and how close the erroneous distribution is to the standard Gumbel distribution with respect to expectation and standard deviation.

Table 1. Simulation Results for MNL models

Distribution	Sample Size	Accepted	Rejected
$\varepsilon_1 \sim G(0, 1)$	200	1.00	0.00 (Type-I Error)
$\varepsilon_1 \sim G(0, 1)$	4000	0.96	0.04 (Type-I Error)
$\varepsilon_1 \sim N(0, 1)$	4000	0.03 (Type-II Error)	0.97
$\varepsilon_1 \sim \gamma + N(0, 1) \bullet \frac{\pi}{\sqrt{6}}$	200000	0.02 (Type-II Error)	0.98

γ is the Euler constant ≈ 0.577216; number of repetitions = 100; 0.05 level of significance is used

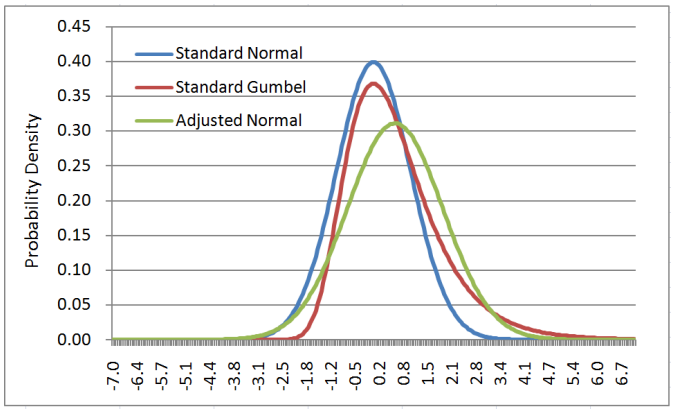

Figure 1. Comparison of Distributions Being Tested

A sample size of 4000 is found to be adequate to provide a statistical power of 0.97 and a Type-II error probability less than 0.05 in distinguishing the standard normal distribution from the standard Gumbel distribution. However, when the erroneous distribution is the adjusted normal distribution with the same expectation and standard deviation as that of a standard Gumbel distribution, the sample size needs to be as high as 200000 to obtain satisfactory statistical power and Type-II error probability. Although the standard normal distribution appears more similar (visually) to the standard Gumbel distribution than the adjusted normal distribution, it is actually much more difficult to distinguish the adjusted normal distribution (than the standard normal distribution) from the standard Gumbel distribution presumably because the adjusted normal distribution has the same expectation and standard deviation as the standard Gumbel distribution.

4.2 Simulation Experiments for Test of MDCEV Model

In the MDCEV model specification, either parameter α_j or γ_j needs to be normalized for identification purposes. Alternative normalization approaches will result in two different model profiles, namely, the "α" profile and "γ" profile. Simulation experiments are conducted for both profiles.

Similar to previous experiment for the MNL model, the experiment is designed with four alternatives and corresponding random utility values, computed as:

$u_1 = -1.0 + 0.9 \times x_1 + \varepsilon_1$;

$u_2 = -0.6 + 0.8 \times x_2 + \varepsilon_2$;

$u_3 = -0.7 + 0.4 \times x_3 + \varepsilon_3$;

$u_4 = 0.6 \times x_4 + \varepsilon_4$.

Once again, x_1, x_2, x_3, and x_4 follow an independent uniform distribution between 0 and 10, i.e., U(0,10). ε_2, ε_3, and ε_4 follow an independent standard Gumbel distribution, i.e., G(0,1), while ε_1 follows either G(0,1) or a distribution other than G(0,1). The continuous resource budget T = Trunc [U(0,1000)]+10, where Trunc[ ] is a function to convert a real number to an integer by truncating its decimal part. The utility function for the α-profile MDCEV model is $U_j = \frac{1}{\alpha_j}\psi_j\left[\left(\frac{t_j}{1}+1\right)^{\alpha_j}-1\right]$, where $\psi_j = e^{u_j}$, α_1 = 0.5, α_2 = 0.6, α_3 = 0.7 and α_4 = 0.8. Then, four resource allocation variables, t_1, t_2, t_3, and t_4, are calculated by applying the efficient algorithm proposed in Pinjari and Bhat (2011).

Table 2 furnishes the simulation results for α–Profile MDCEV models. When the sample size is 500 (relatively small) or 4000 (relatively large), it is found that the proposed test provides satisfactory Type-I error probabilities. In the case of Type-II errors, it is found that a sample size of 1000 is adequate to realize a satisfactorily small Type-II error probability when ε_1 follows the standard normal distribution. A sample with just 3000 observations is adequate to distinguish the adjusted normal distribution from the standard Gumbel distribution. In the case of the MNL model, the simulation experiment results reported in the previous section suggest that the sample size required is 4000 and 200000 to yield satisfactory Type-II error probabilities. The difference in statistical power between the MNL and MDCEV models is likely due to the different nature of the two types of models. Whereas the MNL model focuses exclusively on a single discrete choice from among a set of alternatives, the MDCEV model utilizes information about multiple discrete choices and the allocation of a continuous budget to each of the chosen alternatives. As the MDCEV model utilizes more information about choices and resource allocation behavior, a smaller sample is presumably needed to avoid making a Type-II error.

Table 2. Simulation Experiment Results for MDCEV models

Distribution	Sample Size	Accepted	Rejected
α-Profile MDCEV Model			
$\varepsilon_1 \sim G(0, 1)$	500	0.97	0.03 (Type-I Error)
$\varepsilon_1 \sim G(0, 1)$	4000	0.95	0.05 (Type-I Error)
$\varepsilon_1 \sim N(0, 1)$	1000	0.05 (Type-II Error)	0.95
$\varepsilon_1 \sim \gamma + N(0, 1) \bullet \frac{\pi}{\sqrt{6}}$	3000	0.05 (Type-II Error)	0.95
γ-Profile MDCEV Model			
$\varepsilon_1 \sim G(0, 1)$	500	0.97	0.03 (Type-I Error)
$\varepsilon_1 \sim G(0, 1)$	4000	0.95	0.05 (Type-I Error)
$\varepsilon_1 \sim N(0, 1)$	750	0.02 (Type-II Error)	0.98
$\varepsilon_1 \sim \gamma + N(0, 1) \bullet \frac{\pi}{\sqrt{6}}$	750	0.03 (Type-II Error)	0.97

Number of Repetitions = 100; 0.05 significance level is used

The experiment for γ–profile MDCEV models also considers four alternatives with random utility values computed as:

$u_1 = 0.4 - 0.5 \times x_1 + \varepsilon_1$;

$u_2 = -0.5 - 0.4 \times x_2 + \varepsilon_2$;

$u_3 = -0.6 - 0.3 \times x_3 + \varepsilon_3$;

$u_4 = - 0.5 \times x_4 + \varepsilon_4$.

x_1, x_2, x_3, and x_4 follow an independent uniform distribution between 0 and 10, i.e., U(0, 10). ε_2, ε_3, and ε_4 follow an independent standard Gumbel distribution, i.e., G(0,1), while ε_1 follows either G(0,1) or a distribution other than G(0,1). The continuous resource budget T = Trunc[U(0,500)] + 10. The utility function is $U_j = \gamma_j \psi_j ln\left[\frac{t_j}{\gamma_j} + 1\right]$, where $\psi_j = e^{u_j}$, $\gamma_1 = 2.0$, $\gamma_2 = 1.0$, $\gamma_3 = 0.5$ and $\gamma_4 = 1.5$. Four resource allocation variables t_1, t_2, t_3, and t_4 are calculated by applying the efficient forecasting algorithm of Pinjari and Bhat (2011). Table 2 provides the simulation experiment results for γ–Profile MDCEV models.

Similar to the α-profile MDCEV models, γ-profile MDCEV models have satisfactory Type-I error probabilities when the sample size is 500 (relatively small) and 4000 (relatively large), respectively. As for the statistical power, γ-profile MDCEV models require even fewer observations (750) than the α-profile MDCEV models to distinguish the normal or adjusted normal distribution from the standard Gumbel distribution.

5. CASE STUDIES

The simulation experiments demonstrated the efficacy of the statistical test developed in this research effort. The proposed testing methods were applied to MNL and MDCEV models estimated on real-world travel survey data sets to examine the extent to which violations of the standard Gumbel distribution occur in different empirical contexts.

5.1 MNL Model of Long-distance Travel Mode Choice

The data set used for the MNL model case study is obtained from the "AER" package available in R statistical platform (Greene, 2011). The data set includes 210 observations of mode choice for long-distance travel among four alternative modes: Air, Train, Bus, and Car. The MNL model is estimated using a specification that includes a number of explanatory variables. All of the explanatory variables exhibit behaviorally intuitive and statistically significant coefficients in the MNL model. The model estimation results furnish goodness-of-fit statistics that are consistent with those typically seen for MNL models in research studies and practice. The proposed method is applied to test the validity of the standard Gumbel distribution in each utility function. Estimation results are presented in Table 3. It is seen that the test does not reject the null hypothesis that the random error follows the standard Gumbel distribution for the first, third, and fourth alternatives (air, bus and car). However, the test does reject the standard Gumbel distributional assumption in the case of the second utility function (train mode). The χ^2 statistic is 8.933 and the corresponding p-value is 0.003. As per the simulation result in Table 1, the Type-I error probability is almost zero when the sample size is as small as 200. A Type-I error occurs when the standard Gumbel distribution is rejected even though the random error truly follows the standard Gumbel distribution.

Table 3. Test Results of the MNL model (N = 210)

Model	MNL Model		Test (Air Utility)		Test (Train Utility)		Test (Bus Utility)		Test (Car Utility)	
Variable	Coeff.	t-test	Coeff.	t-test	Coeff.	t-test	Coeff.	t-test	Coeff.	t-test
Constant for Air	8.038	5.58	7.822	4.86	7.618	5.39	7.955	5.54	6.425	4.35
Air travel time (min)	-0.030	-4.18	-0.031	-4.20	-0.031	-4.50	-0.030	-4.17	-0.024	-3.40
Party size in mode choice	-0.951	-3.66	-0.962	-3.62	-0.864	-3.45	-0.937	-3.62	-0.837	-3.49
Air waiting time (min)	-0.103	-5.72	-0.105	-5.75	-0.101	-5.69	-0.103	-5.71	-0.094	-5.41
δ_1	--	--	0.133	0.40	--	--	--	--	--	--
Constant for Train	4.409	5.03	4.429	5.04	4.253	6.45	4.349	4.98	3.185	3.61
Train travel time (min)	-0.005	-2.94	-0.005	-2.99	-0.005	-3.80	-0.005	-2.97	-0.004	-2.55
Train cost ($)	-0.024	-1.84	-0.024	-1.85	-0.019	-2.04	-0.024	-1.83	-0.021	-1.75
Household annual income (K $)	-0.048	-3.66	-0.048	-3.65	-0.036	-3.96	-0.047	-3.62	-0.042	-3.45
Train waiting time (min)	-0.064	-3.83	-0.064	-3.82	-0.045	-4.17	-0.064	-3.80	-0.058	-3.64
δ_1	--	--	--	--	-0.745	-3.43	--	--	--	--
Constant for Bus	4.905	3.85	4.919	3.86	4.461	3.68	5.033	4.19	3.643	2.88
Bus travel time (min)	-0.006	-3.26	-0.006	-3.30	-0.006	-3.55	-0.006	-3.40	-0.004	-2.29
Bus waiting time (min)	-0.151	-5.17	-0.151	-5.16	-0.141	-5.09	-0.138	-4.13	-0.147	-5.18
δ_1	--	--	--	--	--	--	-0.195	-1.07	--	--
Car travel time (min)	-0.006	-5.13	-0.007	-5.14	-0.007	-5.70	-0.006	-5.15	-0.005	-3.68
δ_1	--	--	--	--	--	--	--	--	-0.588	-1.16
$LL(\beta)$	-160.092	--	-159.963	--	-155.626	--	-159.751	--	-159.339	--
χ^2 Statistic	--	--	0.258	--	**8.933**	--	0.683	--	1.506	--
p-value	--	--	0.611	--	**0.003**	--	0.409	--	0.220	--

Therefore, it can be concluded that the standard Gumbel distribution is not valid for the random error term in the utility equation of the train mode. It is also interesting to see that the additional coefficient "δ_1" is significant and considerably changes the magnitude of coefficients in the second utility function.

5.2 MDCEV Model of Activity Engagement and Time Allocation

The proposed method is next applied to an MDCEV model of activity (stop) engagement and time allocation for home-based work tours, where the primary purpose of the tour is to go to the workplace. The model predicts the secondary activities that will be undertaken during the tour, along with the time allocated to each activity. The model is estimated on 2009 National Household Travel Survey (NHTS) data from the Greater Phoenix Metropolitan Area (Garikapati et al. 2014). In the interest of brevity, model estimation results are suppressed in this paper. Activities within a home-based work tour are classified into 11 types. Explanatory variables include commuters' demographic and socio-economic characteristics, commute and work status, built environment attributes associated with residential and work locations, and accessibility measures. The sample size of the estimation sample is 1968 and the log-likelihood value at convergence is -17528.658.

The proposed method is applied to test the validity of the standard Gumbel distribution for the random error term in the utility function of each activity type. The test results, including the log-likelihood value at convergence, χ^2 statistics and corresponding p-values, are listed in Table 4. It is seen that the likelihood ratio tests reject the distributional assumption for *all* of the utility functions in the MDCEV model. It is interesting to see that the χ^2 statistic values for two outside goods (going to work and returning home) are considerably lower than those for other goods, which implies that the distributional assumption is even more invalid for alternatives that are not outside goods. Outside goods are those that are consumed (chosen) by all observations in the data set. In this case, going to work and returning home are activities that must be undertaken in the context of home-based work tours and are therefore treated as outside goods.

Table 4. Test Results for the MDCEV Model of Home-Based Work Tour Activity Engagement and Time Allocation

Activity Type	*LL*(b)	χ^2 Statistics	p-value
Work Epoch (outside good 1)	-17218.635	620.047	0.000
Home Epoch (outside good 2)	-17377.888	301.539	0.000
Other Escort Epoch 1	-14957.450	5142.417	0.000
Other Escort Epoch 2	-14871.097	5315.122	0.000
School Escort Epoch 1	-14987.671	5081.973	0.000
School Escort Epoch 2	-14860.288	5336.741	0.000
Shopping Epoch 1	-15118.892	4819.533	0.000
Maintenance Epoch 1	-15071.536	4914.243	0.000
Meal Epoch 1	-14987.993	5081.330	0.000
Social Visit Epoch 1	-14850.375	5356.567	0.000
Other Discretionary Epoch 1	-14925.381	5206.554	0.000

5.3 MDCEV Model of Household Vehicle Fleet Composition and Utilization

The proposed method is finally applied to a MDCEV model of household vehicle fleet composition and utilization. The model was estimated on 2009 National Household Travel Survey (NHTS) data from

the Greater Phoenix Metropolitan Area (You et al. 2014). Model estimation results are suppressed in the interest of brevity. Household vehicles are categorized into 14 types, defined by a cross-classification of body type and vintage. Explanatory variables include household composition and structure, economic status, and built environment attributes of the household location. The estimation sample has 4262 observations and the log-likelihood value at convergence is -77020.486.

The proposed method is applied to test the validity of the standard Gumbel distribution for the random error term in the utility function of each vehicle type. Results of the test are shown in Table 5. It is seen once again that the likelihood ratio tests reject the distributional assumption for all utility functions of the MDCEV model. As in the previous case, the χ^2 statistic for the outside good is much lower than those for other goods, which implies that the distributional assumption is more invalid for alternatives that are not treated as outside goods. In this model, non-motorized vehicle is treated as the outside good as all individuals are assumed to walk for at least some minimal duration over the course of a day. Thus, the non-motorized vehicle is chosen by all observations in the data set.

Table 5. Test Results for the MDCEV Model of Vehicle Fleet Composition and Utilization

Vehicle Type	LL(b)	Chi-Sq. Statistics	p-value
Non-motorized vehicle (outside good)	-76839.481	362.018	0.000
Car (years) 0–5	-73185.231	7670.517	0.000
6–11	-72898.926	8243.128	0.000
≥12	-72234.349	9572.282	0.000
Van (years) 0–5	-71472.336	11096.308	0.000
6–11	-71493.089	11054.802	0.000
≥12	-71239.748	11561.484	0.000
SUV (years) 0–5	-72159.229	9722.522	0.000
6–11	-71842.624	10355.732	0.000
≥12	-71363.351	11314.278	0.000
Pickup (years) 0–5	-71736.722	10567.536	0.000
6–11	-71832.517	10375.946	0.000
≥12	-71591.310	10858.360	0.000
Motorbike	-71387.042	11266.896	0.000

6. CONCLUSIONS AND DISCUSSIONS

In this paper, a practical but statistically rigorous method is proposed to test the validity of the standard Gumbel distribution assumption that is often associated with the random error components in multinomial travel-related choice models, including both discrete and discrete-continuous models. The method is based on the use of the orthonormal Legendre polynomial to derive a closed-form likelihood expression that nests the likelihood functions of the multinomial logit (MNL) and multiple discrete-continuous extreme value (MDCEV) models. The standard likelihood-ratio test can then be applied to test the validity of the Gumbel distribution innate to the logit-based choice models. The efficacy of the proposed method is first examined via simulation experiments. Results of the simulation experiments show that acceptably low Type-I and Type-II error probabilities in the application of the test may be realized at typically available travel survey sample sizes, except for the case of the Type-II error probability for the multinomial logit model (which needs a sample size on the

order of 200000 to ensure Type-II error probability less than 0.05). The proposed method is then applied to three real-world case studies, including a multinomial logit model of long-distance travel mode choice, a multiple discrete-continuous choice model of activity-time allocation in home-based work tours, and another multiple discrete-continuous choice model of vehicle fleet composition and utilization. For all three models, the proposed test shows that the assumption of a standard Gumbel distribution for the random error components is rejected at a high level of confidence.

In theory, the violation of underlying distributional assumptions will lead to the estimation of parameters that are statistically inconsistent and inefficient. However, the extent to which departures from the standard Gumbel distribution affect the magnitudes of model coefficients and model sensitivity to policy scenarios remains unclear. This question could be addressed in future research using simulation experiments in which a robust model based on valid distributional assumptions is developed. Then, the impacts of modifying the distribution on the random error component on parameter estimates in logit-based models can be explicitly quantified.

The three real-world models considered in this paper are typical discrete or discrete-continuous choice models with a rich set of explanatory variables and exhibiting goodness-of-fit statistically typically encountered in practice. Given that violations of the standard Gumbel distribution are occurring even in the context of these models, it is prudent to identify approaches that could potentially overcome any ill-effects of the distributional assumption violations. Three strategies are noted below:

1. *Adopting an alternative parametric distribution*
   The normal distribution is usually a preferred distribution for the error term that captures the effects of unobserved factors (as opposed to the Gumbel distribution). Thus, the corresponding multinomial probit (MNP) or multiple discrete-continuous probit (MDCP) models (Bhat et al., 2013) are likely to be better alternatives to MNL and MDCEV models. In addition, heteroskedastic versions of the logit model or discrete-continuous choice model (Bhat, 1995; Sikder and Pinjari, 2013) may also prove to be better alternatives. However, tests should be conducted to determine whether violations of the chosen alternative parametric distribution are occurring. There are methods currently available to test for violations of the normal distribution in the econometric literature (e.g., Bera et al., 1984).
2. *Adopting mixed logit model with random coefficients of known distribution*
   Modelers may still use the standard Gumbel distribution for the random error components, but could identify a random coefficient that follows a certain parametric distributional assumption. Then, this coefficient may be scaled up to approximate the random utility values and minimize the impact of the Gumbel random disturbance. Details of this method are described in Train (2009). Note that the existence of such a random coefficient is a necessary condition to apply this method. The test proposed by Fosgerau and Bierlaire (2007) may be used to test the distributional assumption on a random coefficient in a mixed logit model.
3. *Developing a robust multinomial choice model free from distributional assumptions*
   Statistical or econometric models estimated using maximum likelihood methods necessarily involve the making of distributional assumptions. Modelers have longed for the development of robust choice models free from distributional assumptions for several decades (e.g., Manski, 1975; Gallant and Nychka, 1987; Klein and Spady, 1993); however, most practical distribution-free or semi-parametric choice methods have been limited to the analysis and modeling of binary choice variables, rendering their application to multinomial choice contexts computationally challenging.

The extension of such distribution-free or semi-parametric approaches to the modeling of multinomial choice variables would constitute a worthy research endeavor.

ACKNOWLEDGEMENT

This research is funded by the startup grant of the "Thousand Young Talent" program from the Central Organization Department of China.

REFERENCES

Ben-Akiva, M. and S.R. Lerman (1985). Discrete choice analysis: theory and application to travel demand, MIT Press, Cambridge, M.A.

Bera, A., Jarque, C., and Lee, L. (1984). Testing the Normality Assumption in Limited Dependent Variable Models. International Economic Review, 25, pp. 563-578.

Bhat, C.R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. Transportation Research Part B: Methodological, Volume 29, Issue 6, December 1995, pp. 471-483

Bhat, C.R. (2005). A Multiple Discrete–Continuous Extreme Value Model: Formulation and Application to Discretionary Time-use Decisions. Transportation Research Part B, Vol. 39, No. 8, 2005, pp. 679–707.

Bhat, C.R. (2008). The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions. Transportation Research Part B, 2008, Vol. 42, No. 3, pp. 274–303.

Bhat, C.R., M. Castro and M. Khan (2013). A new estimation approach for the multiple discrete–continuous probit (MDCP) choice model. Transportation Research Part B, 55, pp. 1-22.

Bierens, H.J. (2008). Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results. Econometric Theory, 24(3), pp. 749-794.

Chen, H.Z. and A. Randall (1997). Semi-nonparametric estimation of binary response models with an application to natural resource valuation. Journal of Econometrics, 76(1-2), pp. 323-340.

Creel, M. and J. Loomis (1997). Semi-nonparametric Distribution-Free Dichotomous Choice Contingent Valuation. *Journal of Environmental Economics and Management*, 32(3), pp. 341-358.

Crooker, J.R. and J.A. Herriges (2007). Parametric and Semi-Nonparametric Estimation of Willingness-to-Pay in the Dichotomous Choice Contingent Valuation Framework, Environmental and Resource Economics, 27(4), pp. 451-480.

Fosgerau, M. and M. Bierlaire (2007). A practical test for the choice of mixing distribution in discrete choice models. Transportation Research Part B: Methodological, 41, pp. 784–794.

Gallant, A.R. and D. Nychka (1987). Semi-Non-Parametric Maximum Likelihood Estimation, Econometrica, 55, pp. 363–390.

Garikapati, V.M., D. You, R.M. Pendyala, P.S. Vovsha, V. Livshits and K. Jeon (2014). Multiple Discrete-Continuous Model of Activity Participation and Time Allocation for Home-Based Work Tours. Transportation Research Record: Journal of the Transportation Research Board, No. 2429, Transportation Research Board of the National Academies, Washington, D.C., 2014, pp. 90–98.

Greene, W.H. (2011). Econometric Analysis (7th Edition), Prentice Hall. The web linkage of the data Table F18-2: http://people.stern.nyu.edu/wgreene/Text/Edition7/TableF18-2.csv or the Table F21-2 from http://pages.stern.nyu.edu/~wgreene/Text/tables/tablelist5.htm

Jäggi, B., C. Weis and K.W. Axhausen (2013). Stated response and multiple discrete-continuous choice models: Analyses of residuals. Journal of Choice Modelling, 6, pp. 44-59.

Klein, R.W. and R.H. Spady (1993). An Efficient Semiparametric Estimator for Binary Response Models. Econometrica, 61(2), pp. 387-421.

Lee, L.F. (1995). Semiparametric maximum likelihood estimation of polychotomous and sequential choice models. Journal of Econometrics, 65, pp. 381-428

Manski, C.F. (1975). Maximum score estimation of the stochastic utility model of choice. Journal of Econometrics, 3(3), pp. 205–228.

Mcfadden, D. (1974). Conditional logit analysis of qualitative choice behavior. in P. Zarembka (ed.), Frontiers in Econometrics, pp. 105-142, Academic Press: New York.

Pinjari, A.R. and C.R. Bhat (2011). Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: application to residential energy consumption analysis. Technical paper, Department of Civil and Environmental Engineering, University of South Florida.

Sikder, S. and A.R. Pinjari (2013). The benefits of allowing heteroscedastic stochastic distributions in multiple discrete-continuous choice models. Journal of Choice Modelling, 9, pp. 39-56.

Train, K.E. (2009). Discrete choice methods with simulation. Cambridge University Press.

You, D., V.M. Garikapati, R.M. Pendyala, C.R. Bhat, S. Dubey, K. Jeon, and V. Livshits (2014). Development of Vehicle Fleet Composition Model System for Implementation in Activity-Based Travel Model. Transportation Research Record: Journal of the Transportation Research Board, No. 2430, pp. 145–154.

Appendix A: Mathematical Derivation of the Test for MNL Model

Suppose there are "J" alternatives in the choice set and their random utility functions are U_1, U_2, ..., U_J. And the utility U_j is expressed as the sum of the systematic component V_j and the random component ε_j , i.e. $U_j = V_j + \varepsilon_j$. Suppose the standard Gumbel distributional assumption of ε_1 needs to be tested. All the other ε_j ($j > 1$) really follow the standard Gumbel distribution. One can derive the probability that the alternative 1 is chosen as:

$$P(y=1) = P(U_1 > U_2, U_1 > U_3, \dots, U_1 > U_J)$$

$$= P(V_1 + \varepsilon_1 > V_2 + \varepsilon_2, V_1 + \varepsilon_1 > V_3 + \varepsilon_3, \dots, V_1 + \varepsilon_1 > V_J + \varepsilon_J)$$

$$= P(\varepsilon_2 < V_{12} + \varepsilon_1, \varepsilon_3 < V_{13} + \varepsilon_1, \dots, \varepsilon_J < V_{1J} + \varepsilon_1).$$

In the formulae above, V_{ij} represents V_i - V_j. Since ε_j are assumed to be independently distributed in an MNL model, then $P(y=1) = \int_{-\infty}^{+\infty}\left[\prod_{j=2}^{J} G(V_{1j} + \varepsilon_1)\right] f(\varepsilon_1)d\varepsilon_1$.

By plugging the PDF of the semi-nonparametric distribution in Equation (9), one can obtain that

$$P(y=1) = \int_{-\infty}^{+\infty}\left[\prod_{j=2}^{J} G(V_{1j} + \varepsilon_1)\right]\{\sum_{m=0}^{2} \xi_m [G(\varepsilon_1)]^m\} g(\varepsilon_1)d\varepsilon_1.$$

Then, it can be simplified as $P(y=1) = \sum_{m=0}^{2} \xi_m q_1(m)$, where

$$q_1(m) = \int_{-\infty}^{+\infty}\left[\prod_{j=2}^{J} G(V_{1j} + \varepsilon_1)\right][G(\varepsilon_1)]^m g(\varepsilon_1)d\varepsilon_1$$

$$= \int_{-\infty}^{+\infty}\left[\prod_{j=2}^{J} \exp\left(-e^{-V_{1j}-\varepsilon_1}\right)\right][\exp(-e^{-\varepsilon_1})]^m \exp(-e^{-\varepsilon_1})\exp(-\varepsilon_1)d\varepsilon_1$$

$$= -\int_{-\infty}^{+\infty}\left[\prod_{j=2}^{J} \exp\left(-e^{-V_{1j}-\varepsilon_1}\right)\right][\exp(-e^{-\varepsilon_1})]^m \exp(-e^{-\varepsilon_1})d[\exp(-\varepsilon_1)].$$

Let $w = \exp(-\varepsilon_1)$. Since ε_1 ranges from -∞ to +∞, "w" should range from +∞ to 0. Then

$$q_1(m) = -\int_{+\infty}^{0}\left[\prod_{j=2}^{J} \exp\left(-w \cdot e^{-V_{1j}}\right)\right][\exp(-w)]^m \exp(-w)dw$$

$$= -\int_{+\infty}^{0} \exp\left(-w\sum_{j=2}^{J} e^{-V_{1j}} - mw - w\right) dw.$$

Let $\theta = -\sum_{j=2}^{J} e^{-V_{1j}} - m - 1$, then $q_1(m) = -\int_{+\infty}^{0} \exp(\theta w)\, dw = -\frac{1}{\theta}\int_{+\infty}^{0} \exp(\theta w)\, d(\theta w)$.

Let $\eta = \theta w$. Since "w" ranges from +∞ to 0 and $\theta < 0$, η should range from -∞ to 0. Then, $q_1(m) = -\frac{1}{\theta}\int_{-\infty}^{0} \exp(\eta)\, d\eta = -\frac{1}{\theta} = -\frac{1}{-\sum_{j=2}^{J} e^{-V_{1j}} - m - 1} = \frac{1}{\sum_{j=2}^{J} e^{V_j - V_1} + m + 1} = \frac{e^{V_1}}{(m+1)e^{V_1} + \sum_{j=2}^{J} e^{V_j}} = \frac{e^{V_1}}{m \cdot e^{V_1} + \sum_{j=1}^{J} e^{V_j}}$.

For the alternative other than the 1st one (say the 2nd alternative), its choice probability:

$$P(y=2) = P(U_2 > U_1, U_2 > U_3, \dots, U_2 > U_J)$$

$$= P(V_2 + \varepsilon_2 > V_1 + \varepsilon_1, V_2 + \varepsilon_2 > V_3 + \varepsilon_3, \dots, V_2 + \varepsilon_2 > V_J + \varepsilon_J)$$

$$= P(\varepsilon_1 < V_{21} + \varepsilon_2, \varepsilon_3 < V_{23} + \varepsilon_2, \dots, \varepsilon_J < V_{2J} + \varepsilon_2).$$

Since $F(\varepsilon_1) = \sum_{m=0}^{2}\left\{\xi_m \frac{[G(\varepsilon_1)]^{m+1}}{m+1}\right\}$ according to Equation (10), $P(y=2) =$

$\int_{-\infty}^{+\infty}\sum_{m=0}^{2}\left\{\xi_m \frac{[G(V_{21}+\varepsilon_2)]^{m+1}}{m+1}\right\}\left[\prod_{j=3}^{J} G(V_{2j} + \varepsilon_2)\right] g(\varepsilon_2)d\varepsilon_2$. It can be simplified as

$P(y=2) = \sum_{m=0}^{2} \xi_m q_2(m)$, where

$$q_2(m) = \frac{1}{m+1}\int_{-\infty}^{+\infty}[G(V_{21}+\varepsilon_2)]^{m+1}\left[\prod_{j=3}^{J} G(V_{2j}+\varepsilon_2)\right] g(\varepsilon_2)d\varepsilon_2$$

$$= \frac{1}{m+1}\int_{-\infty}^{+\infty} exp\left[-e^{-V_{21}-\varepsilon_2+\ln(m+1)}\right]\left[\prod_{j=3}^{J} exp(-e^{-V_{2j}-\varepsilon_2})\right]\exp(-e^{-\varepsilon_2})\cdot e^{-\varepsilon_2}d\varepsilon_2$$

$$= -\frac{1}{m+1}\int_{-\infty}^{+\infty} exp\left[-e^{-V_{21}-\varepsilon_2+\ln(m+1)}\right]\left[\prod_{j=3}^{J} exp(-e^{-V_{2j}-\varepsilon_2})\right]\exp(-e^{-\varepsilon_2})de^{-\varepsilon_2}$$

Let $w = e^{-\varepsilon_2}$. Since ε_2 ranges from -∞ to +∞, "*w*" should range from +∞ to 0, then

$$q_2(m) = -\frac{1}{m+1}\int_{+\infty}^{0} exp\left[-w\cdot e^{-V_{21}+\ln(m+1)}\right]\left[\prod_{j=3}^{J} exp(-w\cdot e^{-V_{2j}})\right]\exp(-w)dw$$

$$= -\frac{1}{m+1}\int_{+\infty}^{0} exp\left[-w\cdot e^{-V_{21}+\ln(m+1)}\right]\left[exp\left(-w\cdot\sum_{j=3}^{J} e^{-V_{2j}}\right)\right]\exp(-w)dw.$$

Let $\theta = -e^{-V_{21}+\ln(m+1)} - \sum_{j=3}^{J} e^{-V_{2j}} - 1$, then

$$q_2(m) = -\frac{1}{m+1}\int_{+\infty}^{0}\exp(\theta w)\,dw = -\frac{1}{\theta(m+1)}\int_{+\infty}^{0}\exp(\theta w)\,d(\theta w).$$

Let $\eta = \theta w$. Since "w" ranges from +∞ to 0 and $\theta < 0$, η should range from -∞ to 0. Then,

$$q_2(m) = -\frac{1}{\theta(m+1)}\int_{-\infty}^{0}\exp(\eta)\,d(\eta) = -\frac{1}{\theta(m+1)} = \frac{1}{m+1}\cdot\frac{1}{e^{-V_{21}+\ln(m+1)}+\sum_{j=3}^{J} e^{-V_{2j}}+1}$$

$$= \frac{1}{m+1}\cdot\frac{e^{V_2}}{(m+1)\cdot e^{V_1}+\sum_{j=3}^{J} e^{V_j}+e^{V_2}} = \frac{e^{V_2}}{(m+1)\left(m\cdot e^{V_1}+\sum_{j=1}^{J} e^{V_j}\right)}.$$

Without loss of generality, $q_k(m) = \frac{e^{V_k}}{(m+1)\left(m\cdot e^{V_1}+\sum_{j=1}^{J} e^{V_j}\right)}$, where k > 1.

Appendix B: Mathematical Derivation of the Test for MDCEV Model

Assume the utility for the alternative "j" takes the following form as in Bhat (2008): $U_j = \frac{\gamma_j}{\alpha_j}\psi_j\left[\left(\frac{t_j}{\gamma_j}+1\right)^{\alpha_j}-1\right]$, where $\psi_j = e^{V_j+\varepsilon_j}$ and the alternative index j = 1, 2,..., K. Suppose the random error ε_1 in the 1st alternative needs to be tested. "K" represents the total number of alternatives and "M" represents the total number of alternatives being allocated resource (M ≤ K). Each individual is supposed to allocate continuous resource by maximizing the overall utility value subject to the budget:

Maximize $\sum_{j=1}^{K}\frac{\gamma_j}{\alpha_j}\psi_j\left[\left(\frac{t_j}{\gamma_j}+1\right)^{\alpha_j}-1\right]$ subject to $\sum_{j=1}^{K} t_j = T$.

It is equivalent to maximizing $\sum_{j=1}^{K}\frac{\gamma_j}{\alpha_j}\psi_j\left[\left(\frac{t_j}{\gamma_j}+1\right)^{\alpha_j}-1\right] - \lambda(\sum_{j=1}^{K} t_j - T)$. For the observation where the resource allocation $t_1 > 0$, according to KT conditions, one should have

$$\psi_1\left(\frac{t_1}{\gamma_1}+1\right)^{\alpha_1-1} = \lambda,\ \psi_2\left(\frac{t_2}{\gamma_2}+1\right)^{\alpha_2-1} = \lambda,\ \ldots,\ \psi_M\left(\frac{t_M}{\gamma_M}+1\right)^{\alpha_M-1} = \lambda;\ \psi_{M+1}\left(\frac{t_{M+1}}{\gamma_{M+1}}+1\right)^{\alpha_{M+1}-1} <$$

$$\lambda,\ldots,\ \psi_K\left(\frac{t_K}{\gamma_K}+1\right)^{\alpha_K-1} < \lambda.$$

In the above formulae, $t_j > 0$ when j ≤ M and $t_j = 0$ when M < j ≤ K. Then, one will have

$$\psi_1\left(\frac{t_1}{\gamma_1}+1\right)^{\alpha_1-1} = \psi_j\left(\frac{t_j}{\gamma_j}+1\right)^{\alpha_j-1},\ \ j = 2, 3, \ldots, M.$$

Take a logarithm function on both sides and one should obtain that:

$V_1 + \varepsilon_1 = V_j + \varepsilon_j$, j = 2, 3, ..., M;

$V_1 + \varepsilon_1 > V_j + \varepsilon_j$, j = (M+1), (M+2), ..., K, where $V_j = x_j\beta_j + (\alpha_j - 1)\ln\left(\frac{t_j}{\gamma_j} + 1\right)$.

As per Bhat (2005), the likelihood value $P = |J| \cdot \int_{-\infty}^{\infty}\left[\prod_{j=2}^{M} g(V_{1j} + \varepsilon_1)\right]\left[\prod_{j=M+1}^{K} G(V_{1j} + \varepsilon_1)\right] f(\varepsilon_1)d\varepsilon_1$.

To obtain $f(\varepsilon_1)$, one can plug the SNP density function: $f(x) = \{\sum_{\mathrm{m}=0}^{2} \xi_{\mathrm{m}}[G(x)]^m\}\mathrm{g(x)}$ into the equation above and then have that $P = |J| \int_{-\infty}^{\infty}\left[\prod_{j=2}^{M} g(V_{1j} + \varepsilon_1)\right]\left[\prod_{j=M+1}^{K} G(V_{1j} + \varepsilon_1)\right]\{\sum_{\mathrm{m}=0}^{2} \xi_{\mathrm{m}}[G(\varepsilon_1)]^m\}\mathrm{g}(\varepsilon_1)d\varepsilon_1$

$= \sum_{m=0}^{2} \xi_m |J| \int_{-\infty}^{\infty}\left[\prod_{j=2}^{M} g(V_{1j} + \varepsilon_1)\right]\left[\prod_{j=M+1}^{K} G(V_{1j} + \varepsilon_1)\right][G(\varepsilon_1)]^m\mathrm{g}(\varepsilon_1)d\varepsilon_1 = \sum_{m=0}^{2} \xi_m q_1(m)$,

where $q_1(m) = \int_{-\infty}^{+\infty} |J|\left(\prod_{j=2}^{M} g(V_{1j} + \varepsilon_1)\right)\left(\prod_{j=M+1}^{K} G(V_{1j} + \varepsilon_1)\right)[G(\varepsilon_1)]^m\mathrm{g}(\varepsilon_1)d\varepsilon_1$

$=|J| \int_{-\infty}^{\infty}\left[\prod_{j=2}^{M} g(V_{1j} + \varepsilon_1)\right]\left[\prod_{j=M+1}^{K} G(V_{1j} + \varepsilon_1)\right][G(\varepsilon_1)]^m\mathrm{g}(\varepsilon_1)d\varepsilon_1$

$=|J| \int_{-\infty}^{\infty}\left\{\prod_{j=2}^{M}\left[G(V_{1j} + \varepsilon_1) \cdot e^{-V_{1j}-\varepsilon_1}\right]\right\}\left[\prod_{j=M+1}^{K} G(V_{1j} + \varepsilon_1)\right][G(\varepsilon_1)]^m G(\varepsilon_1)e^{-\varepsilon_1}d\varepsilon_1$

$=|J| \int_{-\infty}^{\infty}\left\{\prod_{j=2}^{M}\left[G(V_{1j} + \varepsilon_1) \cdot e^{-V_{1j}-\varepsilon_1}\right]\right\}\left[\prod_{j=M+1}^{K} G(V_{1j} + \varepsilon_1)\right][G(\varepsilon_1)]^m G(V_{11} + \varepsilon_1)e^{-\varepsilon_1}d\varepsilon_1$

$= |J| \int_{-\infty}^{\infty} \prod_{j=1}^{K} G(V_{1j} + \varepsilon_1)\left[\prod_{j=2}^{M} e^{-V_{1j}-\varepsilon_1}\right][G(\varepsilon_1)]^m e^{-\varepsilon_1}d\varepsilon_1$

$= |J| \int_{-\infty}^{\infty}\left[\prod_{j=1}^{K} exp(-e^{-V_{1j}-\varepsilon_1})\right]\left(\prod_{j=2}^{M} e^{-V_{1j}}\right)(e^{-\varepsilon_1})^{M-1}e^{-\varepsilon_1} \cdot [G(\varepsilon_1)]^m d\,\varepsilon_1$

$= |J|\left(\prod_{j=2}^{M} e^{-V_{1j}}\right) \int_{-\infty}^{+\infty} exp\left(-\sum_{j=1}^{K} e^{-V_{1j}-\varepsilon_1}\right)(e^{-\varepsilon_1})^{M-1} exp\left[-e^{-\varepsilon_1+ln(m)}\right]e^{-\varepsilon_1}d\varepsilon_1$

Define the integral part in the formula above as "Int":

$Int = \int_{-\infty}^{+\infty} exp\left(-\sum_{j=1}^{K} e^{-V_{1j}-\varepsilon_1}\right)(e^{-\varepsilon_1})^{M-1} exp\left[-e^{-\varepsilon_1+ln(m)}\right]e^{-\varepsilon_1}d\varepsilon_1$

$= -\int_{-\infty}^{+\infty} exp\left(-\sum_{j=1}^{K} e^{-V_{1j}-\varepsilon_1}\right)(e^{-\varepsilon_1})^{M-1} exp\left[-e^{-\varepsilon_1+ln(m)}\right]de^{-\varepsilon_1}$

Let $w = e^{-\varepsilon_1}$. Since ε_1 ranges from $-\infty$ to $+\infty$, w should range from $+\infty$ to 0. Then,

$Int = -\int_{+\infty}^{0} exp\left(-w\sum_{j=1}^{K} e^{-V_{1j}}\right)w^{M-1} exp(-mw)dw$

$= -\int_{+\infty}^{0} exp\left[-w\left(\sum_{j=1}^{K} e^{-V_{1j}} + m\right)\right]w^{M-1}dw.$

Let $a = \sum_{j=1}^{K} e^{-V_{1j}} + m$ and $Int = -\int_{+\infty}^{0} \exp(-w \cdot a)w^{M-1}dw$. Let $b = -aw$, then $w = -\frac{b}{a}$ and $dw = -\frac{db}{a}$. Since "w" ranges from $+\infty$ to 0 and a > 0, "b" should range from $-\infty$ to 0. Thus, $Int = -\int_{-\infty}^{0} \mathrm{e}^b \left(-\frac{b}{a}\right)^{M-1}\left(-\frac{1}{a}\right)db = \frac{(-1)^{M-1}}{a^M}\int_{-\infty}^{0} \mathrm{e}^b \cdot b^{M-1}\,db.$

As per Bhat (2005), $\int_{-\infty}^{0} \mathrm{e}^{\mathrm{b}} \cdot \mathrm{b}^{\mathrm{M}-1}d\mathrm{b} = (-1)^{\mathrm{M}-1} \cdot (\mathrm{M}-1)!$. Then, $Int = \frac{(-1)^{M-1}}{a^M}(-1)^{\mathrm{M}-1} \cdot (\mathrm{M}-$

$1)! = \frac{(M-1)!}{a^M} = \frac{(M-1)!}{\left(m+\sum_{j=1}^{K} e^{-V_{1j}}\right)^M}$.

Then, $q_1(m) = |J|\left(\prod_{j=2}^{M} e^{-V_{1j}}\right)\frac{(M-1)!}{\left(m+\sum_{j=1}^{K} e^{-V_{1j}}\right)^M} = |J|\left(\prod_{j=2}^{M} e^{-V_1+V_j}\right)\frac{(M-1)!}{\left(m+\sum_{j=1}^{K} e^{-V_1+V_j}\right)^M}$

$= |J|(M-1)!\frac{\left(\prod_{j=1}^{M} e^{V_j}\right)}{\left(m\cdot e^{V_1}+\sum_{j=1}^{K} e^{V_j}\right)^M}$.

As per page 704 in Bhat (2005), $|J| = \left(\prod_{j=1}^{M} c_j\right)\left(\sum_{j=1}^{M}\frac{1}{c_j}\right)$, where $c_j = \frac{1-\alpha_j}{t_j+\gamma_j}$. Thus, $q_1(m) =$

$\left(\prod_{j=1}^{M} c_j\right)\left(\sum_{j=1}^{M}\frac{1}{c_j}\right)\left[\frac{\prod_{j=1}^{M} e^{V_j}}{\left(m\cdot e^{V_1}+\sum_{j=1}^{K} e^{V_j}\right)^M}\right](M-1)!$.

For the observation where the resource allocation $t_1 = 0$, one may pick out another alternative (say "2") to which the resource is allocated and therefore $t_2 > 0$. According to KT conditions,

$V_2 + \varepsilon_2 = V_j + \varepsilon_j$, j = 3, ..., (M+1);

$V_2 + \varepsilon_2 > V_1 + \varepsilon_1$;

$V_2 + \varepsilon_2 > V_j + \varepsilon_j$, j = (M+2), (M+3), ..., K.

The likelihood value can be computed as:

$P = |J| \cdot \int_{-\infty}^{\infty}\left[\prod_{j=3}^{M+1} g(V_{2j}+\varepsilon_2)\right]F(V_{21}+\varepsilon_2)\left[\prod_{j=M+2}^{K} G(V_{2j}+\varepsilon_2)\right]g(\varepsilon_2)d\varepsilon_2$.

Here, the CDF of the SNP distribution, as in Equation (10), is used for F(x). Then, $P = |J| \cdot$

$\int_{-\infty}^{\infty}\left[\prod_{j=3}^{M+1} g(V_{2j}+\varepsilon_2)\right]\sum_{m=0}^{2}\left\{\frac{\xi_m[G(V_{21}+\varepsilon_2)]^{m+1}}{m+1}\right\}\left[\prod_{j=M+2}^{K} G(V_{2j}+\varepsilon_2)\right]g(\varepsilon_2)d\varepsilon_2$

$= \sum_{m=0}^{2}\xi_m\frac{|J|}{m+1}\int_{-\infty}^{\infty}\left[\prod_{j=3}^{M+1} g(V_{2j}+\varepsilon_2)\right][G(V_{21}+\varepsilon_2)]^{m+1}\left[\prod_{j=M+2}^{K} G(V_{2j}+\varepsilon_2)\right]g(\varepsilon_2)d\varepsilon_2$.

Let $q_2(m) = \frac{|J|}{m+1}\int_{-\infty}^{\infty}\left[\prod_{j=3}^{M+1} g(V_{2j}+\varepsilon_2)\right][G(V_{21}+\varepsilon_2)]^{m+1}\left[\prod_{j=M+2}^{K} G(V_{2j}+\varepsilon_2)\right]g(\varepsilon_2)d\varepsilon_2$ and

$P = \sum_{m=0}^{2}\xi_m q_2(m)$. Then, let $q_2(m) = \frac{|J|}{m+1}Int1$, where

$Int1 = \int_{-\infty}^{\infty}\left[\prod_{j=3}^{M+1} g(V_{2j}+\varepsilon_2)\right][G(V_{21}+\varepsilon_2)]^{m+1}\left[\prod_{j=M+2}^{K} G(V_{2j}+\varepsilon_2)\right]g(\varepsilon_2)d\varepsilon_2$.

$= \int_{-\infty}^{\infty}\left[\prod_{j=3}^{M+1} G(V_{2j}+\varepsilon_2)\right]\left[\prod_{j=3}^{M+1} exp(-V_{2j}-\varepsilon_2)\right][G(V_{21}+\varepsilon_2)]^{m+1}\left[\prod_{j=M+2}^{K} G(V_{2j}+\varepsilon_2)\right]G(V_{22}+$

$\varepsilon_2)\exp(-\varepsilon_2)d\varepsilon_2$

$= \int_{-\infty}^{\infty}\left[\prod_{j=2}^{K} G(V_{2j}+\varepsilon_2)\right]\left[\prod_{j=3}^{M+1} exp(-V_{2j}-\varepsilon_2)\right][G(V_{21}+\varepsilon_2)]^{m+1}\exp(-\varepsilon_2)d\varepsilon_2$

$= \int_{-\infty}^{\infty}\left[\prod_{j=2}^{K}\exp(-e^{-V_{2j}-\varepsilon_2})\right]\left[\prod_{j=3}^{M+1} exp(-V_{2j})\right](e^{-\varepsilon_2})^{M-1}\cdot[G(V_{21}+\varepsilon_2)]^{m+1}\cdot\exp(-\varepsilon_2)d\varepsilon_2$

$= \left[\prod_{j=3}^{M+1} exp(-V_{2j})\right]\int_{-\infty}^{\infty}\left[\prod_{j=2}^{K}\exp(-e^{-V_{2j}-\varepsilon_2})\right](e^{-\varepsilon_2})^{M-1}\cdot[G(V_{21}+\varepsilon_2)]^{m}[G(V_{21}+\varepsilon_2)]^{1}\cdot$

$\exp(-\varepsilon_2)d\varepsilon_2$

$= \left[\prod_{j=3}^{M+1} exp(-V_{2j})\right]\int_{-\infty}^{\infty}\left[\prod_{j=2}^{K}\exp(-e^{-V_{2j}-\varepsilon_2})\right](e^{-\varepsilon_2})^{M-1}\cdot[G(V_{21}+\varepsilon_2)]^{m}\exp(-e^{-V_{21}-\varepsilon_2})\cdot$

$\exp(-\varepsilon_2)d\varepsilon_2$

$= \left[\prod_{j=3}^{M+1} exp\left(-V_{2j}\right)\right] \int_{-\infty}^{\infty} \left[\prod_{j=1}^{K} \exp(-e^{-V_{2j}-\varepsilon_2})\right] (e^{-\varepsilon_2})^{M-1} \cdot [G(V_{21}+\varepsilon_2)]^m \cdot \mathrm{e}^{-\varepsilon_2} d\varepsilon_2$

$= \left[\prod_{j=3}^{M+1} exp\left(-V_{2j}\right)\right] Int2$, where

$Int2 = \int_{-\infty}^{\infty} \left[\prod_{j=1}^{K} \exp(-e^{-V_{2j}-\varepsilon_2})\right] (e^{-\varepsilon_2})^{M-1} \cdot [G(V_{21}+\varepsilon_2)]^m \cdot \mathrm{e}^{-\varepsilon_2} d\varepsilon_2$

$= \int_{-\infty}^{\infty} \left[\prod_{j=1}^{K} \exp(-e^{-V_{2j}-\varepsilon_2})\right] (e^{-\varepsilon_2})^{M-1} \cdot \exp\left[-e^{-\varepsilon_2-V_{21}+\ln(m)}\right] \cdot \mathrm{e}^{-\varepsilon_2} d\varepsilon_2$

$= \int_{-\infty}^{\infty} \exp\left[-e^{-\varepsilon_2} \sum_{j=1}^{K} e^{-V_{2j}}\right] (e^{-\varepsilon_2})^{M-1} \cdot \exp\left[-e^{-\varepsilon_2} e^{-V_{21}+\ln(m)}\right] \cdot \mathrm{e}^{-\varepsilon_2} d\varepsilon_2$

$= -\int_{-\infty}^{\infty} \exp\left[-e^{-\varepsilon_2} \sum_{j=1}^{K} e^{-V_{2j}}\right] (e^{-\varepsilon_2})^{M-1} \cdot \exp(-e^{-\varepsilon_2} e^{-V_{21}+\ln(m)}) d\mathrm{e}^{-\varepsilon_2}$

Let $\mathrm{w} = \mathrm{e}^{-\varepsilon_2}$. Since ε_2 ranges from $-\infty$ to $+\infty$, w should range from $+\infty$ to 0. Thus,

$Int2 = -\int_{+\infty}^{0} \exp\left[-w \sum_{j=1}^{K} e^{-V_{2j}}\right] w^{M-1} \cdot \exp\left[-w \cdot e^{-V_{21}+\ln(m)}\right] d\mathrm{w}$

$= -\int_{+\infty}^{0} \exp\left\{-w\left[e^{-V_{21}+\ln(m)} + \sum_{j=1}^{K} e^{-V_{2j}}\right]\right\} w^{M-1} d\mathrm{w}.$

Let $a = e^{-V_{21}+\ln(m)} + \sum_{j=1}^{K} e^{-V_{2j}}$ and then $Int2 = -\int_{+\infty}^{0} \mathrm{e}^{-aw} w^{M-1} d\mathrm{w}$. Let $b = -aw$, then $w = -\frac{b}{a}$, $dw = -\frac{db}{a}$. Let $Int2 = -\int_{-\infty}^{0} \mathrm{e}^{b} \left(-\frac{b}{a}\right)^{M-1} \left(-\frac{1}{a}\right) db = \frac{(-1)^{M-1}}{a^M} \int_{-\infty}^{0} \mathrm{e}^{b} \cdot b^{M-1} db$. As per Bhat (2005), $\int_{-\infty}^{0} \mathrm{e}^{\mathrm{b}} \cdot \mathrm{b}^{\mathrm{M}-1} d\mathrm{b} = (-1)^{\mathrm{M}-1} \cdot (\mathrm{M}-1)!$. Then, $Int2 = \frac{(-1)^{M-1}}{a^M} (-1)^{\mathrm{M}-1} \cdot (\mathrm{M}-1)! = \frac{(M-1)!}{a^M}$.

Thus, $q_2(m) = \frac{|\mathrm{J}|}{\mathrm{m}+1} \left[\prod_{j=3}^{M+1} exp\left(-V_{2j}\right)\right] \frac{(M-1)!}{a^M}$

$= \frac{|\mathrm{J}|}{\mathrm{m}+1} \left[\prod_{j=3}^{M+1} exp\left(-V_{2j}\right)\right] \frac{(M-1)!}{\left(e^{-V_{21}+\ln(m)} + \sum_{j=1}^{K} e^{-V_{2j}}\right)^M} = \frac{|\mathrm{J}|(M-1)!}{(\mathrm{m}+1)} \frac{\left[\prod_{j=3}^{M+1} e^{-V_{2j}}\right]}{\left(me^{-V_{21}} + \sum_{j=1}^{K} e^{-V_{2j}}\right)^M}$

$= \frac{|\mathrm{J}|(M-1)!}{(\mathrm{m}+1)} \frac{\left[\prod_{j=3}^{M+1} e^{V_j}\right] \left(e^{-V_2}\right)^{M-1}}{\left(me^{V_1} + \sum_{j=1}^{K} e^{V_j}\right)^M (e^{-V_2})^M} = \frac{|\mathrm{J}|(M-1)!}{(\mathrm{m}+1)} \frac{\prod_{j=2}^{M+1} e^{V_j}}{\left(me^{V_1} + \sum_{j=1}^{K} e^{V_j}\right)^M}$

$= \left(\prod_{\mathrm{j}=1}^{\mathrm{M}} \mathrm{c_j}\right) \left(\sum_{\mathrm{j}=1}^{\mathrm{M}} \frac{1}{\mathrm{c_j}}\right) \frac{(M-1)!}{(\mathrm{m}+1)} \frac{\prod_{j=2}^{M+1} e^{V_j}}{\left(me^{V_1} + \sum_{j=1}^{K} e^{V_j}\right)^M}.$

Since $\mathrm{t_1} = 0$, e^{V_1} will not be multiplied with the numerator term. However, the numerator term is still the product of exponentials of utilities for "M" alternatives being allocated resource. In summary, the SNP likelihood function can be expressed as $P = \sum_{\mathrm{m}=0}^{2} \xi_{\mathrm{m}}\, \mathrm{q(m)}$, where

$$\mathrm{q(m)} = \begin{cases} \left(\prod_{\mathrm{j}=1}^{\mathrm{M}} \mathrm{c_j}\right) \left(\sum_{\mathrm{j}=1}^{\mathrm{M}} \frac{1}{\mathrm{c_j}}\right) \left[\frac{\prod_{j=1}^{M} e^{V_j}}{\left(m \cdot e^{V_1} + \sum_{j=1}^{K} e^{V_j}\right)^M}\right] (M-1)!, \ if\ \mathrm{t_1} > 0, \\ \left(\prod_{\mathrm{j}=1}^{\mathrm{M}} \mathrm{c_j}\right) \left(\sum_{\mathrm{j}=1}^{\mathrm{M}} \frac{1}{\mathrm{c_j}}\right) \left[\frac{\prod_{j=2}^{M+1} e^{V_j}}{(\mathrm{m}+1)\left(me^{V_1} + \sum_{j=1}^{K} e^{V_j}\right)^M}\right] (M-1)!, \ if\ \mathrm{t_1} = 0 \end{cases}.$$