Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2016) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

INTERNATIONAL SYMPOSIUM ON TRAFFIC AND TRANSPORTATION THEORY

ISTTT22

On the Development of a Semi-Nonparametric Generalized Multinomial Logit Model for Travel-related Choices

Xin Ye [a,*], Ram M. Pendyala[b]

[a]Key Laboratory of Road and Traffic Engineering of Ministry of Education,
School of Transportation Engineering,
Tongji University, 4800 Cao'an Road, Shanghai, China, 201804
[b]School of Sustainable Engineering and the Built Environment,
Arizona State University,
660 S. College Avenue, Tempe, AZ 85281

Abstract

A semi-nonparametric generalized multinomial logit model is formulated based on orthonormal Legendre polynomials to extend the standard Gumbel distribution. The resulting semi-nonparametric function can represent a probability density function for a large family of multimodal distributions. The model has a closed-form log-likelihood function that facilitates model estimation. The proposed method is applied to model commute mode choice among four alternatives (auto, transit, bicycle and walk) using data from Argau, Switzerland. Comparisons between the multinomial logit model and the proposed semi-nonparametric model show that violations of the standard Gumbel distribution assumption lead to considerable inconsistency in parameter estimates and model inferences.

Peer-review under responsibility of the scientific committee of the 22nd International Symposium on Transportation and Traffic Theory.

Keywords: Semi-nonparametric Approach, Gumbel Distribution, Orthonormal Legendre Polynomial, Multinomial Choice Model, Mode Choice Model

* Corresponding author. Tel.: +86-21-59946091; fax: +86-21-59946091.
E-mail address: xye@tongji.edu.cn

Peer-review under responsibility of the scientific committee of the 22nd International Symposium on Transportation and Traffic Theory.

1. Introduction

The Gumbel distribution (also referred to as the Type-I extreme value distribution) plays a central role in discrete choice models for travel demand analysis (Ben-Akiva and Lerman, 1985). This can be attributed to two major reasons. First, the Gumbel distribution closely resembles the normal distribution, which is often the preferred distribution to characterize the random disturbance term in an econometric model that accounts for the effect of unobserved factors. Second, when the Gumbel distribution is assumed for random components of utility functions, a closed-form likelihood function is obtained in the context of the application of the microeconomic utility maximization principle. With a closed-form likelihood function, maximum likelihood estimation (MLE) methods can be applied with ease to estimate model coefficients consistently and efficiently. Due to these appealing features of the Gumbel distribution, the Multinomial Logit (MNL) model is widely applied in practice and preferred over its counterpart that is based on the assumption of a normally distributed random error component (i.e., Multinomial Probit or MNP model).

However, according to the theory of maximum likelihood estimation, the consistency and efficiency of maximum likelihood estimators depend on the validity of the distributional assumption made on the random error term. Recently, the authors developed a practical method to test the validity of the distributional assumption on the random disturbance term in an MNL model and obtained significant statistical evidence to reject the Gumbel distribution assumption in a very commonly encountered empirical setting (Ye et al., 2016) dealing with long distance travel mode choice. That finding motivates this particular study which aims to develop and present the formulation for a Semi-nonparametric Generalized Multinomial Logit Model (SGMNL) for travel-related choices. The objective of this study is to generalize the MNL model by relaxing the assumption of a Gumbel distribution using a semi-nonparametric approach, and then demonstrate the efficacy of the approach by applying the generalized model to an empirical setting of travel mode choice. It should be noted that this generalization essentially differs from other extensions of the MNL that have yielded the Nested Logit, Cross-nested Logit, Heteroskedastic Logit or Multinomial Probit models (Train, 2009). Those models are generalized extensions that still employ the unimodal Gumbel or normal marginal distributions, whereas the proposed semi-nonparametric model presented in this paper allows the marginal error distribution to have multiple modes. Thus, the proposed model provides the ability to examine potential bias in model coefficients, marginal effects and elasticities in a discrete choice model that may arise when a unimodal distribution like the standard Gumbel distribution is violated in random components of utility functions.

The remainder of the paper is organized as follows. In Section 2, the literature on semi-nonparametric choice models is reviewed. In Section 3, the orthonormal Legendre polynomial is introduced and then applied to extend the standard Gumbel distribution, thus enabling the development and formulation of the Semi-nonparametric Generalized Multinomial Logit Model (SGMNL). In Section 4, data used for the empirical study is described, and then empirical estimation results are presented and discussed. Finally, conclusions and directions for future research are presented in the last section.

2. Literature Review

As early as the time when McFadden initially proposed the MNL model (McFadden, 1974), econometricians (e.g., Manski, 1975) have been questioning the validity of the distributional assumption on the error term in random utility functions. Concerns about the effects of violations of distributional assumptions on the random error components have 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 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, on the other hand, is developed 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 more widely applied in practice than the semiparametric approach (e.g., Chen and Randall, 1997; Creel and Loomis, 1997; Crooker and Herriges, 2007).

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

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

In Equation (1), $\varphi(\varepsilon)$ represents the PDF of the standard normal distribution and is referred to as "a priori distribution". The denominator ensures that $\int_{-\infty}^{+\infty} f(\varepsilon) d\varepsilon = 1$. Equation (1) can be extended as follows:

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

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

For the probability value above, one may apply the recursion formulae to derive the indefinite integral of $\int \varepsilon^{i+j}\varphi(\varepsilon) d\varepsilon$. The above SNP choice model is limited to a binary choice situation due to its computational complexity in the context of a multinomial choice situation.

3. Modeling Methodology

3.1. Extending the Standard Gumbel Distribution with the Orthonormal Legendre Polynomial

Bierens (2008) proposed a new polynomial, called the orthonormal Legendre polynomial, for estimating distributions on the unit interval in a semi-nonparametric way. In the transportation literature, this approach has been used to test normal and log-normal distributions of random coefficients in mixed logit models (Fosgerau and Bierlaire, 2007). As per Fosgerau and Bierlaire (2007) and Bierens (2008), the orthonormal Legendre polynomial can be recursively defined as:

$$L_0 = 1,\ L_1 = \sqrt{3}(2x - 1), \tag{4}$$

$$L_n = \alpha_n (2x - 1) L_{n-1} + \beta_n L_{n-2},\ n \geqslant 2 \tag{5}$$

In Equation (5), $\alpha_n = \frac{\sqrt{4n^2-1}}{n}$, $\beta_n = -\frac{(n-1)\sqrt{2n+1}}{n\sqrt{2n-3}}$. The advantage of using this polynomial is that it ensures

$$\int_0^1 L_m(x) L_n(x) dx = \begin{cases} 0 & \text{if } m \neq n \\ 1 & \text{if } m = n \end{cases}. \tag{6}$$

According to Gallant and Nychka (1987), the priori distribution in the semi-nonparametric approach can be a distribution other than the standard normal distribution. In this paper, the orthonormal Legendre polynomial is used to construct a semi-nonparametric probability density function that extends the standard Gumbel distribution as follows:

$$f(x) = \frac{\left\{1 + \sum_{k=1}^{K} \delta_k L_k[G(x)]\right\}^2}{1 + \sum_{k=1}^{K} \delta_k^{\,2}} g(x), \tag{7}$$

where $g(x) = \exp(-e^{-x}) \cdot \exp(-x)$, $G(x) = \exp(-e^{-x})$, δ_k are scalar parameters and K represents the total number of polynomials. Using Equation (6), it can be shown that $\int_{-\infty}^{+\infty} f(x) = 1$. As f(x) is positive, it qualifies to be a probability density function.

Figure 1 compares the semi-nonparametric probability densities when the number of polynomials is 1 ($K = 1$) and the parameter δ_1 takes a value of -2, 0, 1 or 2. When δ_1 is 0, the distribution reduces to a standard Gumbel distribution, as shown in the red curve. When δ_1 takes a value of -2, 1 or 2, the distributions are bimodal, although the secondary peak in the distribution is rather flat when δ_1 is equal to -2 or 1.

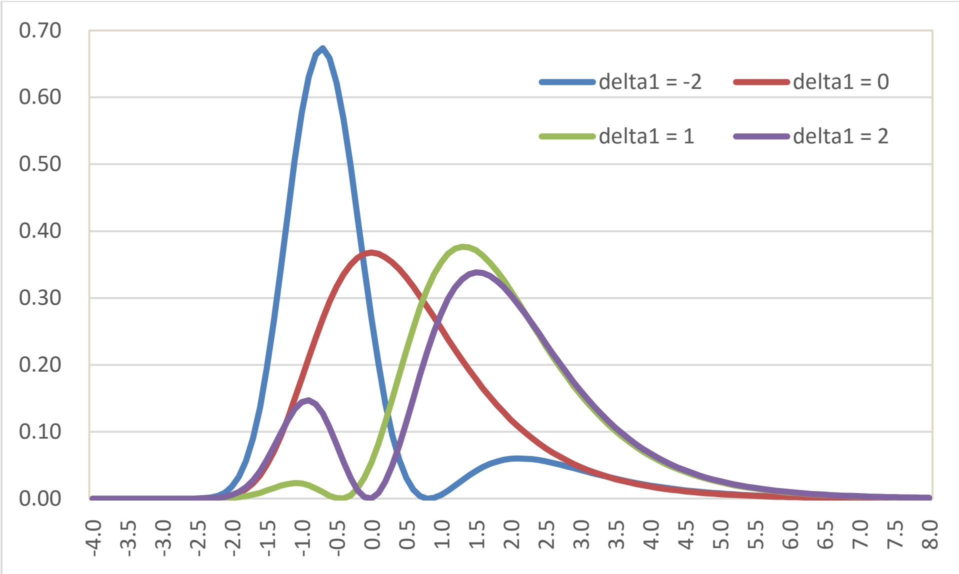

Fig. 1. Comparisons of Semi-nonparametric Probability Densities When K = 1

Figure 2 compares the semi-nonparametric probability densities when the number of polynomials is 2 ($K = 2$) and two scalar parameters δ_1 and δ_2 are involved. With two polynomials in the function where the highest power term of "$G(x)$" increases to 2, the SNP function represented in Equation (7) can generate a more flexible probability density distribution. It can be seen that, when δ_1 is 2 and δ_2 is -2, the distribution exhibits two modes with almost equal probability densities. When δ_1 is 0 and δ_2 is 2, the distribution shows three modes. It may further be expected that, when the number of polynomials (K) or the highest power term of "$G(x)$" increases, the SNP function with a flexible form can effectively represent the probability density function for a large family of distributions with multiple modes.

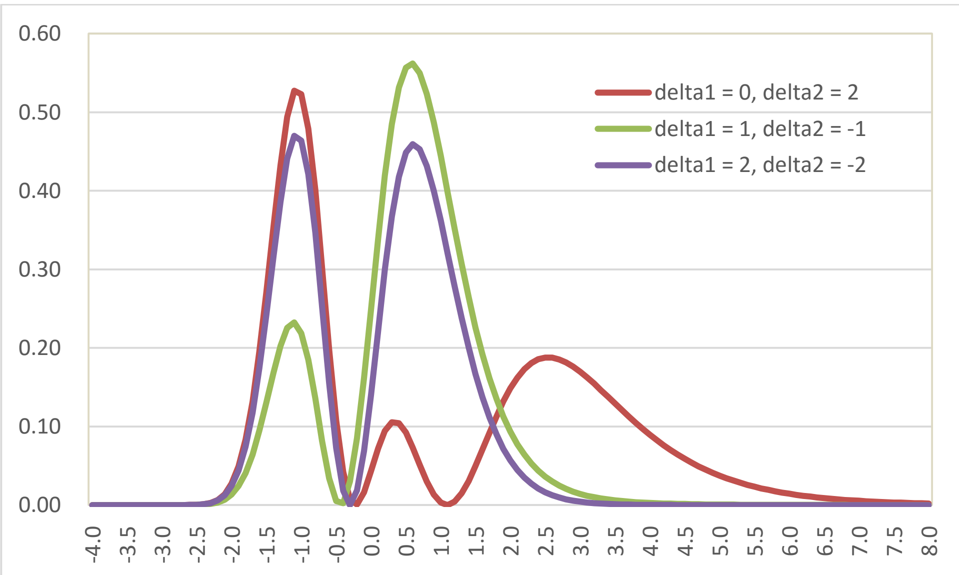

Fig. 2. Comparisons of Semi-nonparametric Probability Densities When K = 2

3.2. Simplifying the Semi-nonparametric (SNP) Probability Density Function (PDF)

Following Gallant and Nychka (1987), it is possible to employ the SNP PDF function in Equation (7) to construct random components in utility functions so that multiple modes may be accommodated in their distributions. Before the choice probability can be derived, the SNP PDF needs to be simplified first. Using Equations (4) and (5), it is possible to write the polynomial in a general form as:

$$\mathrm{L_n} = \sum_{k=0}^{n} c_{n,k} x^k \,, \tag{8}$$

where $c_{n,k}$ is a constant coefficient for the term "x^k" in the n^th^ polynomial. When k > n, $c_{n,k} = 0$. Let $a = 2\sqrt{3}$ and $b = -\sqrt{3}$. Then, $L_0 = 1$ and $L_1 = ax + b$. When $n \geqslant 2$, as per Equation (5), $L_n = \alpha_n(2x-1)L_{n-1} + \beta_n L_{n-2} = \alpha_n(2x-1)\sum_{k=0}^{n-1} c_{n-1,k}x^k + \beta_n \sum_{k=0}^{n-2} c_{n-2,k}x^k$

$= 2\alpha_n \sum_{k=0}^{n-1} c_{n-1,k}x^{k+1} - \alpha_n \sum_{k=0}^{n-1} c_{n-1,k}x^k + \beta_n \sum_{k=0}^{n-2} c_{n-2,k}x^k$.

Since $c_{n-2,n-1} = 0$, $L_n = 2\alpha_n \sum_{k=0}^{n-1} c_{n-1,k}x^{k+1} - \alpha_n \sum_{k=0}^{n-1} c_{n-1,k}x^k + \beta_n \sum_{k=0}^{n-1} c_{n-2,k}x^k$

$= \left(-\alpha_n c_{n-1,0} + \beta_n c_{n-2,0}\right)x^0 + \sum_{k=1}^{n-1}\left[\alpha_n\left(2c_{n-1,k-1} - c_{n-1,k}\right) + \beta_n c_{n-2,k}\right]x^k + 2\alpha_n c_{n-1,n-1}x^n$.

Then, it is possible to write:

$$L_n(x) = c_{n,0}x^0 + \sum_{k=1}^{n-1} c_{n,k}x^k + c_{n,n}x^n \,. \tag{9}$$

In the equation above,
$$\begin{cases} c_{n,0} = -\alpha_n c_{n-1,0} + \beta_n c_{n-2,0}; \\ c_{n,k} = \alpha_n\left(2c_{n-1,k-1} - c_{n-1,k}\right) + \beta_n c_{n-2,k}, 0 < k < n; \\ c_{n,n} = 2\alpha_n c_{n-1,n-1}. \end{cases} \tag{10}$$

When $n = 0$ or 1, define $c_{0,0} = 1$, $c_{1,0} = b$, and $c_{1,1} = a$. For any integer "n" ($n \geqslant 2$), the recursion equations (10) can be applied to compute the coefficients $c_{i,j}$ and all of the $c_{i,j}$ values form a lower triangular matrix, called the "c" matrix in this paper. Table 1 provides an example of such a "c" matrix when "n" reaches 6. With the "c" matrix, the general form of the orthonormal Legendre polynomial (given the "n" value) may be obtained. For example, when $n = 4$, the fourth row vector of coefficients in the "c" matrix can be extracted to write the polynomial as $L_4(x) = 3x^0 - 60x^1 + 270x^2 - 420x^3 + 210x^4$.

Table 1. An Example of "c" Matrix

n \ k	0	1	2	3	4	5	6
0	1.00	0.00	0.00	0.00	0.00	0.00	0.00
1	-1.73	3.46	0.00	0.00	0.00	0.00	0.00
2	2.24	-13.42	13.42	0.00	0.00	0.00	0.00
3	-2.65	31.75	-79.37	52.92	0.00	0.00	0.00
4	3.00	-60.00	270.00	-420.00	210.00	0.00	0.00
5	-3.32	99.50	-696.49	1857.31	-2089.47	835.79	0.00
6	3.61	-151.43	1514.33	-6057.33	11357.49	-9994.59	3331.53

After the "c" matrix is generated, δ_0 needs to be defined as 0 and the numerator term in the SNP probability density function in Equation (7) can be rewritten as:

$$\left\{1 + \sum_{k=1}^{K} \delta_k L_k[G(x)]\right\}^2 = \left\{\sum_{k=0}^{K} \delta_k L_k[G(x)]\right\}^2 = \left[\sum_{k=0}^{K} \delta_k \sum_{i=0}^{K} c_{k,i} G(x)^i\right]^2 = \left[\sum_{i=0}^{K}\left(\sum_{k=0}^{K} \delta_k c_{k,i}\right) G(x)^i\right]^2.$$

Define a "d" vector, where each element $d_i = \sum_{k=0}^{K} \delta_k c_{k,i}$. Since $c_{k,i} = 0$ when k < i,

$$d_i = \sum_{k=i}^{K} \delta_k c_{k,i}. \tag{11}$$

Thus, $\left\{1 + \sum_{k=1}^{K} \delta_k L_k[G(x)]\right\}^2 = \left[\sum_{i=0}^{K} d_i\, G(x)^i\right]^2 = \sum_{i=0}^{K} \sum_{j=0}^{K} d_i d_j G(x)^{i+j}$. The SNP probability density function in Equation (7) may then be rewritten as:

$$\mathrm{f(x)} = \frac{\{1+\sum_{k=1}^{K}\delta_k L_k[G(x)]\}^2}{1+\sum_{k=1}^{K}{\delta_k}^2}\cdot g(x) = \sum_{i=0}^{K}\sum_{j=0}^{K}\left[\left(\frac{d_i d_j}{\sum_{k=0}^{K}{\delta_k}^2}\right)G(x)^{i+j}\right]g(x) = \{\sum_{m=0}^{M}\xi_m[G(x)]^m\}g(x).$$

In the formula above, $M = 2K$ and $\xi_m = \begin{cases} \frac{\sum_{i=0}^{m} d_i d_{m-i}}{\sum_{k=0}^{K}{\delta_k}^2}, & \text{if } m \le K; \\ \frac{\sum_{i=m-K}^{K} d_i d_{m-i}}{\sum_{k=0}^{K}{\delta_k}^2}, & \text{if } K < m \le 2K. \end{cases}$ (12)

Essentially, the SNP PDF in Equation (7) has been simplified as:

$$f(x) = \{\sum_{m=0}^{M}\xi_m[G(x)]^m\}g(x), \tag{13}$$

where ξ_m is a function with respect to parameters δ_k and $M\ (= 2K)$ is the highest power term of "$G(x)$" in the formula. The relationship between ξ_m and δ_k is described by Equations (11) and (12). The cumulative distribution function (CDF) of the extended probability density function may be formulated as:

$$F(x) = \int_{-\infty}^{x}\{\sum_{m=0}^{M}\xi_m[G(\varepsilon)]^m\}g(\varepsilon)d\varepsilon = \sum_{m=0}^{M}\left\{\frac{\xi_m\cdot[G(x)]^{m+1}}{m+1}\right\}. \tag{14}$$

3.3. Derivation of Choice Probabilities and Likelihood Function

Suppose there are "J" alternatives in the choice set and their random utility functions are $U_1, U_2, \ldots, U_J$. Let the utility U_j be expressed as the sum of the systematic component V_j and the random component ε_j (i.e., $U_j = V_j + \varepsilon_j$). Assume that ε_j independently follows the extended distribution and its semi-nonparametric PDF and CDF are given as:

$$f_j(x) = \left\{\sum_{m_j=0}^{M_j}\xi_{j,m_j}[G(x)]^{m_j}\right\}g(x); \tag{15}$$

$$F_j(x) = \sum_{m_j=0}^{M_j}\left\{\frac{\xi_{j,m_j}[G(x)]^{m_j+1}}{m_j+1}\right\}. \tag{16}$$

The subscript "j" is added to allow ε_j in various random utilities to have different SNP distributions. In addition, three Lemmas, whose proofs are furnished in the Appendix, are used in the subsequent derivation of choice probabilities. Based on the utility maximization principle,

$P(y = 1) = P(U_1 > U_2, U_1 > U_3, \ldots, U_1 > U_J) = P(V_1 + \varepsilon_1 > V_2 + \varepsilon_2, V_1 + \varepsilon_1 > V_3 + \varepsilon_3, \ldots, V_1 + \varepsilon_1 > V_J + \varepsilon_J)$,

where "y" is a categorical choice variable indicating the specific alternative that is chosen. Then,

$P(y = 1) = P(\varepsilon_2 < V_{12} + \varepsilon_1, \varepsilon_3 < V_{13} + \varepsilon_1, \ldots, \varepsilon_J < V_{1J} + \varepsilon_1)$, where $V_{ij} = V_i - V_j$.

$$P(y = 1) = \int_{-\infty}^{+\infty} F_2(V_{12} + \varepsilon_1)F_3(V_{13} + \varepsilon_1), \ldots, F_J(V_{1J} + \varepsilon_1)f_1(\varepsilon_1)d\varepsilon_1$$

$$= \int_{-\infty}^{+\infty}\sum_{m_2=0}^{M_2}\left\{\frac{\xi_{2,m_2}[G(V_{12}+\varepsilon_1)]^{m_2+1}}{m_2+1}\right\}\ldots\sum_{m_J=0}^{M_J}\left\{\frac{\xi_{J,m_J}[G(V_{1J}+\varepsilon_1)]^{m_J+1}}{m_J+1}\right\}\{\sum_{m_1=0}^{M_1}\xi_{1,m_1}[G(\varepsilon_1)]^{m_1}\}g(\varepsilon_1)d\varepsilon_1.$$

According to Lemma 1 in the Appendix, $[G(\varepsilon)]^m = G[\varepsilon - \ln(m)]$, where m > 0. Thus,

$$P(y = 1) = \sum_{m_1=0}^{M_1}\sum_{m_2=0}^{M_2}\ldots\sum_{m_J=0}^{M_J}\frac{(m_1+1)\prod_{i=1}^{J}\xi_{i,m_j}}{\prod_{j=1}^{J}(m_j+1)}\int_{-\infty}^{+\infty}\left\{\prod_{j=2}^{J}G\left[V_{1j} + \varepsilon_1 - \ln\left(m_j + 1\right)\right]\right\}G[\varepsilon_1 - \ln(m_1)]g(\varepsilon_1)d\varepsilon_1 \quad .$$

The integral part in the formula is defined as "Int", i.e.,

$$\mathrm{Int} = \int_{-\infty}^{+\infty}\left\{\prod_{j=2}^{J}G\left[V_{1j} + \varepsilon_1 - \ln\left(m_j + 1\right)\right]\right\}G[\varepsilon_1 - \ln(m_1)]g(\varepsilon_1)d\varepsilon_1.$$

According to Lemma 2 in the Appendix, $\left\{\prod_{j=2}^{J}G\left[V_{1j} + \varepsilon_1 - \ln\left(m_j + 1\right)\right]\right\}G[\varepsilon_1 - \ln(m_1)] = G(\varepsilon_1 + c)$, where $c = -\ln\left[e^{\ln(m_2+1)-V_{12}} + e^{\ln(m_3+1)-V_{13}} + \cdots e^{\ln(m_J+1)-V_{1J}} + e^{\ln(m_1)}\right]$. Then,

$\mathrm{Int} = \int_{-\infty}^{+\infty}G(\varepsilon_1 + c)g(\varepsilon_1)d\varepsilon_1$. According to Lemma 3 in the Appendix,

$$\mathrm{Int} = \int_{-\infty}^{+\infty}G(\varepsilon_1 + c)\,g(\varepsilon_1)d\varepsilon_1 = \frac{1}{1+e^{-c}} = \frac{1}{1+m_1+e^{\ln(m_2+1)-V_{12}}+e^{\ln(m_3+1)-V_{13}}+\cdots+e^{\ln(m_J+1)-V_{1J}}} =$$

$$\frac{e^{V_1}}{(1+m_1)e^{V_1}+(1+m_2)e^{V_2}+\cdots+(1+m_J)e^{V_J}} = \frac{e^{V_1}}{\sum_{j=1}^{J}(1+m_j)e^{V_j}}.$$

By substituting "Int" into the choice probability expression, an elegant closed-form equation for the choice probability may be obtained:

$$P(y=1)=\sum_{m_1=0}^{M_1}\sum_{m_2=0}^{M_2}\cdots\sum_{m_J=0}^{M_J}\left\{\left[\frac{\prod_{i=1}^{J}\xi_{i,m_j}}{\prod_{j=1}^{J}(m_j+1)}\right]\cdot\left[\frac{(m_1+1)e^{V_1}}{\sum_{j=1}^{J}(m_j+1)e^{V_j}}\right]\right\}. \quad (17)$$

The derivation above is shown for the case when $y = 1$, but can be generalized to the situation where $y = k$. Without loss of generality,

$$P(y=k)=\sum_{m_1=0}^{M_1}\sum_{m_2=0}^{M_2}\cdots\sum_{m_J=0}^{M_J}\left\{\left[\frac{\prod_{i=1}^{J}\xi_{i,m_j}}{\prod_{j=1}^{J}(m_j+1)}\right]\cdot\left[\frac{(m_k+1)e^{V_k}}{\sum_{j=1}^{J}(m_j+1)e^{V_j}}\right]\right\}. \quad (18)$$

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

$$LL=\sum_{i=1}^{N}\sum_{k=1}^{J}I(y_i=k)\cdot\ln[P(y_i=k)], \quad (19)$$

where $I(\,)$ is an indicator function; the subscript "i" is the index for an observed choice in the sample and "N" is the sample size. The log-likelihood function can be maximized to estimate model coefficients in the systematic component V_j as well as parameters in the vector δ_j that have been incorporated into ξ_{i,m_j}. When all $M_j = 0$, $P(y=k)=\frac{e^{V_k}}{\sum_{j=1}^{J}e^{V_j}}$ and the model reduces to the familiar MNL model. Thus, the proposed model may be considered a generalized multinomial logit model based on a semi-nonparametric approach.

4. Data and Empirical Estimation Results

4.1. Data and Modeling Procedure

Data for the empirical study is extracted from the 2000 Swiss Microcensus travel survey. A sample consisting of 2,756 commuting trips reported by residents of Aargau Canton in Switzerland is used in this study to estimate models for commute mode choice. Four major commute modes are considered and defined as auto, transit, bicycle and walk. The sample market shares for these four alternatives show that the Aargau Canton of Switzerland depicts a multimodal transportation environment, where 57.62% of commuting trips are made by private auto and the remaining 42.38% of commuting trips are made by transit or non-motorized travel modes. In particular, the transit mode share is 15.86%, the bicycle mode share is 8.31%, and the walk mode share is 18.21%. The mode shares offer a sufficient number of observations in each travel mode, thus supporting the estimation of a mode choice model with multiple alternatives. In addition, multimodal network skim (level of service) data and commuters' demographic and socioeconomic attributes are incorporated in the model specification.

The modeling effort started with the estimation of a simple MNL mode of mode choice. Model estimation results are presented in the first part of Table 2. Both level of service (LOS) attributes and commuters' demographic and socioeconomic attributes are included as explanatory variables in the utility functions. Travel times, including auto in-vehicle time, transit in-vehicle time, and bicycle and walk times, exhibit significantly negative coefficients in the respective utility functions. Transit service frequency takes a significantly positive coefficient, indicating that a high service frequency would increase propensity of commuters to use transit. Model coefficients associated with demographic and socioeconomic attributes show that female commuters are less likely to use auto and bicycle modes. Low-income commuters are more likely to use transit or bicycle modes, while high-income commuters are less likely to use the transit mode. Commuters with lower education level are less likely to use auto than those with high education level. Older commuters are less likely to use public transit. All of the estimation results are behaviorally intuitive and consistent with expectations. The model's log-likelihood value at convergence is -2495.646, corresponding to an adjusted likelihood ratio index of 0.1923 for the overall goodness-of-fit measure of the model.

Next, the proposed SGMNL (semi-nonparametric generalized multinomial logit) model is estimated to relax the standard Gumbel distribution for random components in modal utility functions. First, consider the specification in which K_j is set at 1, where "K" is the number of polynomials in Equation (7) and "j" is an index for travel mode (i.e., j = 1, 2, 3 or 4). When $K_1 = 1$, it is found that the log-likelihood value improves from -2495.646 to -2488.037. As the current model nests the original MNL model, the likelihood ratio chi-square test may be applied to show that the improvement is statistically significant [i.e., (2495.646-2488.037)×2 = 15.22 > 3.84, the critical chi-square value for one degree of freedom at a 95% confidence level]. This result strongly rejects the standard Gumbel distributional assumption for the random component in the auto utility function.

Table 2. Model Estimation Results of MNL and SGMNL-11

	MNL		SGMNL-11	
Explanatory Variable	Est. Coef.	t-stat	Est. Coef.	t-stat
Auto Utility				
Constant	-0.0919	-1.030	0.9242	11.085
Auto in-vehicle time (min)	-0.0766	-11.410	-0.0698	-11.333
Commuter is female	-0.6618	-6.912	-0.5583	-6.568
Education level is less than or equal to middle school	-0.6461	-4.658	-0.4882	-4.226
Transit Utility				
Constant	-2.3730	-10.481	-2.1819	-10.167
Transit in-vehicle time (min)	-0.0380	-5.915	-0.0311	-5.071
Transit service frequency per hour	0.0548	10.221	0.0531	10.288
Commuter's household monthly income is less than CHF 4,000	0.5536	2.432	0.4915	2.391
Commuter's household monthly income is more than CHF 10,000	-0.3342	-2.243	-0.3181	-2.325
Commuter's age (years)	-0.0120	-2.818	-0.0119	-3.060
Bicycle Utility				
Constant	-1.1107	-8.332	-1.0927	-8.227
Bicycle travel time (min)	-0.0756	-13.070	-0.0678	-11.569
Commuter is female	-0.4383	-2.805	-0.4290	-2.768
Commuter's household monthly income is less than CHF 4,000	0.7798	3.399	0.6945	3.223
Walk Utility				
Walk travel time (min)	-0.0381	-24.515	-0.0350	-22.002
Delta Values				
$\delta_{1,1}$	--	--	-1.1776	-1.699
Model Statistics				
LL(β)	-2495.646		-2488.037	
Chi^2-test	--		15.22	
Adj. $\rho^2(c)^*$	0.1923		0.1945	

* The log-likelihood value with constants only: LL(c) = -3104.836.

The model estimation results are presented in the second half of Table 2 and denoted as "SGMNL-11". In this model, the signs of explanatory variable coefficients do not change from those obtained in the standard MNL model, but the magnitudes of coefficients in the auto utility function are found to differ. As expected, the alternative specific constant in the auto utility function changes substantially from -0.0919 to 0.9242 because the expectation of the new SNP distribution is very different from the expectation (i.e., the Euler constant ≈ 0.577) of the standard Gumbel distribution, and the alternative specific constant reflects this difference. An interesting finding is that the significance level of the single coefficient $\delta_{1,1}$ (as indicated by the t-statistic) is not as strong as that implied by the Chi-square test for the overall model fit. However, it should be noted that the likelihood ratio test should be applied to determine whether a semi-nonparametric choice model form is more appropriate because the significance level of multiple coefficients, and their contribution to overall goodness-of-fit, needs to be tested in most occasions.

Table 3. Model Estimation Results of SGMNL-21 and SGMNL-22

	SGMNL-21		SGMNL-22	
Explanatory Variable	Est. Coef.	t-test	Est. Coef.	t-test
Auto Utility				
Constant	0.8312	10.113	0.8584	7.938
Auto in-vehicle time (min)	-0.0386	-10.217	-0.0455	-5.779
Commuter is female	-0.3915	-7.239	-0.4254	-6.307
Commuter's education level is less than or equal to middle school	-0.4060	-5.020	-0.4319	-4.716
Transit Utility				
Constant	-0.4047	-2.842	-1.3658	-8.807
Transit in-vehicle time (min)	-0.0203	-5.202	-0.0235	-4.562
Transit service frequency per hour	0.0330	10.444	0.0388	6.262
Commuter's household monthly income is less than CHF 4,000	0.2537	2.224	0.2644	1.983
Commuter's household monthly income is more than CHF 10,000	-0.1543	-2.094	-0.1836	-2.126
Commuter's age (years)	-0.0055	-2.603	-0.0060	-2.394
Bicycle Utility				
Constant	-1.1433	-8.721	-1.1312	-8.539
Bicycle travel time (min)	-0.0571	-10.509	-0.0592	-10.172
Commuter is female	-0.3041	-2.071	-0.3309	-2.170
Commuter's household monthly income is less than CHF 4,000	0.6960	3.276	0.6925	3.260
Walk Utility				
Walk travel time (min)	-0.0312	-21.910	-0.0319	-20.304
Delta Values				
$\delta_{1,1}$	-1.0236	-0.845	-0.9842	-0.778
$\delta_{2,1}$	-0.8471	-6.461	1.0613	1.625
$\delta_{2,2}$	--	--	-1.9138	-2.370
Model Statistics				
LL(β)	-2472.741		-2469.455	
Chi^2-test	30.59		6.57	
Adj. ρ^2(c)	0.1991		0.1998	

After one significant coefficient $\delta_{1,1}$ is found for the first utility function, K_2 in the second utility function is then set at 1 and $\delta_{2,1}$ is estimated. The model estimation results, denoted as "SGMNL-21", are presented in the first half of Table 3. It can be seen that, after $\delta_{2,1}$ is introduced in the model specification, $\delta_{1,1}$ becomes insignificant but $\delta_{2,1}$ becomes highly significant as indicated by the t-statistics. The likelihood ratio test indicates that the model "SGMNL-21" with additional coefficient $\delta_{2,1}$ is significantly better than the model "SGMNL-11" which does not include parameter $\delta_{2,1}$ [(2488.037-2472.741) × 2 ≈ 30.59 > 3.84]. The likelihood ratio test also shows that "SGMNL-21" is significantly better than the regular MNL model specification [(2495.646-2472.741) × 2 ≈ 45.81 > 5.99, the critical chi-square value corresponding to two degrees of freedom at a 95% confidence level]. Given that both "SGMNL-11" and “SGMNL-21” performed significantly better than the regular MNL model, both $\delta_{1,1}$ and $\delta_{2,1}$ should be retained in the SNP model. A comparison of coefficient magnitudes shows considerable differences among “SGMNL-21”, “SGMNL-11”, and “MNL”, particularly for the transit utility functions. This is consistent with expectations as the introduction of $\delta_{1,1}$ and $\delta_{2,1}$ will change the expectation and standard deviation of random components; both alternative specific constants and coefficients of explanatory variables change accordingly.

When $\delta_{3,1}$ or $\delta_{4,1}$ for bicycle and walk modes are introduced, no significant improvement is observed. In the interest of brevity, those estimation results are not presented here. The modeling effort now moves to the second stage, where the “K” value is increased to 2 and the coefficients $\delta_{1,2}$, $\delta_{2,2}$, $\delta_{3,2}$ and $\delta_{4,2}$ are further specified into the model one by one. In this stage, it is found that only the introduction of $\delta_{2,2}$ in the transit utility function significantly improves the overall model fit (chi-square test value = 6.57 > 3.84) while all other δ values do not. A final model estimation effort is performed, in which the “K” value is increased to 3 and parameter $\delta_{2,3}$ is introduced in the model. The maximum likelihood estimation procedure fails to converge, indicating that the sample of 2,756 choice observations may not be sufficient to support model estimation where the “K” value is increased to 3. Thus, the final best model is considered to be that which adopts a “K” value of 2 and introduces parameter $\delta_{2,2}$, in addition to parameters $\delta_{1,1}$ and $\delta_{2,1}$ introduced in “SGMNL-21”. This final model is designated “SGMNL-22”. If its model coefficients are compared with those in “SGMNL-21”, there is no substantial difference observed, except for the alternative specific constant and the coefficient associated with the “high-income” dummy variable in the transit utility function. As this is considered the final model, all subsequent comparisons are conducted between the MNL model and the final “SGMNL-22” model.

4.2. Plotting Probability Density Distributions of Random Components in the “SGMNL-22” Model

Figure 3 depicts the probability density distributions of random components in the “SGMNL-22” model. Equations (11) and (12) are used to convert the estimated δ values to ξ values and then Equation (13) is used to compute the probability densities based on ξ values. The green curve represents the standard Gumbel distribution for random components (i.e., e3 and e4 in the figure) in bicycle and walk mode utility functions. The blue curve represents the random component in the auto utility function. The coefficient $\delta_{1,1}$ not only reduces the variance of the distribution of the random component but also shifts its mode towards the negative side by about 0.6 units. This helps explain why the alternative specific constant in the auto utility of the “SGMNL-22” model is substantially more positive than that in the MNL model. The positive alternative specific constant offsets the negative expectation of the new random component. The lower variance of the distribution for the auto utility may be due to the existence of fewer unspecified or unobserved random factors associated with auto mode choice than with other mode choices. The distribution of the random component in the transit utility function (i.e., e2) is quite interesting in the context of this study. With the inclusion of parameters $\delta_{2,1}$ and $\delta_{2,2}$ in the model (both of which are significant), “e2” depicts a bimodal distribution as shown by the red curve. The major mode on the right side is located near 0.6 and the minor one on the left side is near -1.2 on the coordinate axis. Based on this finding, it may be conjectured that there are two groups of commuters mixed in the sample. One group of commuters has a positive attitude and inclination towards using transit and constitute the distribution near the major mode. Meanwhile, a smaller group of commuters has negative attitude towards transit and comprises the distribution near the minor mode. Although the exact source of the bimodal distribution is uncertain, the proposed SNP modeling method depicts the existence of such a phenomenon and exposes the potential problem in conventional MNL choice models that are based on unimodal distributional assumptions.

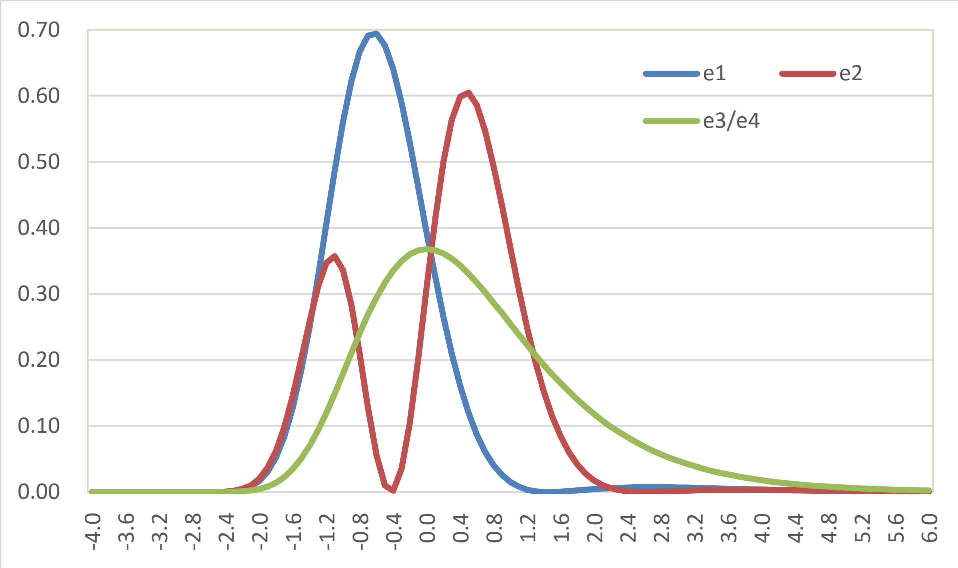

Fig. 3. Probability Density Distributions of Random Components in the "SGMNL-22" model

4.3. A Comparison of Aggregate Marginal Effects and Elasticities

Model coefficients in choice models usually do not directly reflect the impact of an explanatory variable on choice probabilities, particularly when the standard deviations of random components are scaled up or down, as in the transit or auto utility in the SGMNL model estimated in this study. To better understand differences in model sensitivity between MNL and SGMNL models, marginal effects and elasticities are computed and compared. In this subsection, aggregate marginal effects (AME) and aggregate elasticities (AE) with respect to level of service (LOS) variables are computed based on the following two equations:

$$AME = \frac{\partial(\sum_{i=1}^{N} P_i/N)}{\partial z_i} \approx \frac{\sum_{i=1}^{N}[P(x_i,z_i+\Delta)-P(x_i,z_i)]}{N\cdot\Delta}; \quad (20)$$

$$AE = \frac{\partial(\sum_{i=1}^{N} P_i/N)}{(\sum_{i=1}^{N} P_i/N)\cdot\partial z_i/z_i} \approx \frac{\sum_{i=1}^{N}[P(x_i,z_i+\Delta\cdot z_i)-P(x_i,z_i)]}{\Delta\cdot\sum_{i=1}^{N} P(x_i,z_i)}. \quad (21)$$

In the above equations, "P" represents the choice probability expression of the MNL or SGMNL model. "x_i" represents a vector of explanatory variables except the one (i.e., z_i) whose marginal effect or elasticity is being computed. "Δ" takes a value of 0.01 in this study as it is found that such a small interval provides sufficiently accurate estimates for "AME" and "AE" in both MNL and SGMNL models. Table 4 presents a comparison of computed "AME" and "AE" values between MNL and SGMNL-22 models. Considerable differences in "AME" and "AE" are found between these two models, which validates the notion that maximum likelihood estimators are inconsistent when distributional assumptions are violated.

4.4. A Comparison of Disaggregate Marginal Effects and Elasticities

"AME" or "AE" provide sample sensitivity to explanatory variables at the aggregate level and show how a level of service (LOS) variable, for example, affects market shares of alternatives based on the assumption that the sample is randomly drawn and can therefore represent the population shares well. However, the aggregate analysis masks an important difference between MNL and SGMNL models, in that the MNL model has the IIA (Independence of Irrelevant Alternatives) property but the SGMNL model does not have this property. In order to illustrate this important difference between the models, disaggregate marginal effects and elasticities are computed and compared between the two models for a specific individual commuter, who is a 40 year old male, with medium-level income and education level above middle school. The multimodal transportation level of service variables for this individual's commute are as follows: auto in-vehicle time is 5 minutes; transit in-vehicle time is 8 minutes; transit service frequency is 6 times per hour; bicycle travel time is 12 minutes; and walk travel time is 35 minutes. Given these input variables for this specific commuter, both MNL and SGMNL-22 models are applied to compute choice probabilities of alternative travel modes, as shown in Table 5. A substantial difference is found in the choice probability of transit mode. The computations show that the MNL model returns a transit choice probability that is higher than that provided by the SGMNL-22 model by 41.8%, presumably because the model does not capture and reflect the bimodal distribution of the random component in the transit utility function. Table 5 also presents a comparison of predicted means of market shares (i.e., $\sum_{i=1}^{N} \widehat{P}_i / N$) over the entire sample. It is known that the MNL model can replicate the observed sample shares perfectly with the alternative specific constants in utility functions (Ben-Akiva and Lerman, 1985). The SGMNL model does not have this feature, but the greatest difference occurs in the transit share; the relative difference is found to be only 1.5%, which is quite reasonable and acceptable.

Table 4. Comparisons of Aggregate Marginal Effects (AME) and Elasticities (AE)

Level-of-Service Variable	Auto	Transit	Bicycle	Walk
Aggregate Marginal Effects				
Model	SGMNL-22			
Auto in-vehicle time	-0.0140	0.0077	0.0021	0.0042
Transit in-vehicle time	0.0040	-0.0047	0.0002	0.0005
Transit service frequency per hour	-0.0066	0.0078	-0.0004	-0.0008
Bicycle travel time	0.0027	0.0006	-0.0044	0.0011
Walk travel time	0.0030	0.0006	0.0006	-0.0042
Model	MNL			
Auto in-vehicle time	-0.0154	0.0067	0.0030	0.0058
Transit in-vehicle time	0.0033	-0.0043	0.0004	0.0006
Transit service frequency per hour	-0.0048	0.0062	-0.0005	-0.0009
Bicycle travel time	0.0029	0.0007	-0.0055	0.0018
Walk travel time	0.0029	0.0006	0.0009	-0.0044
Aggregate Elasticities				
Model	SGMNL-22			
Auto in-vehicle time	-0.310	0.885	0.173	0.115
Transit in-vehicle time	0.126	-0.482	0.027	0.017
Transit service frequency per hour	-0.110	0.496	-0.072	-0.060
Bicycle travel time	0.077	0.080	-0.789	0.045
Walk travel time	0.164	0.163	0.160	-0.731
Model	MNL			
Auto in-vehicle time	-0.322	0.840	0.265	0.168
Transit in-vehicle time	0.111	-0.452	0.040	0.024
Transit service frequency per hour	-0.082	0.431	-0.091	-0.074
Bicycle travel time	0.089	0.096	-0.976	0.081
Walk travel time	0.169	0.173	0.266	-0.806

Table 5. Comparisons of Market Shares and Individual Choice Probabilities

Statistics	Auto	Transit	Bicycle	Walk
Observed Sample Share	0.5762	0.1586	0.0831	0.1821
Predicted Sample Share from SGMNL-22	0.5738	0.1610	0.0829	0.1823
Predicted Sample Share from MNL	0.5762	0.1586	0.0831	0.1821
Predicted Individual Choice Probabilities from SGMNL-22 for Specific Commuter	0.5877	0.0388	0.1219	0.2516
Predicted Individual Choice Probabilities from MNL for Specific Commuter	0.5771	0.0550	0.1233	0.2446

The IIA property, which is a key feature of the MNL model, is also manifested in the form of equal cross-elasticities

(Bhat, 1995). Methods similar to those expressed in Equations (20) and (21) are applied to compute disaggregate marginal effects and elasticities with respect to LOS variables. The only difference is that the equations are applied to the specific individual commuter, as opposed to all of the commuters in the sample. Results of the computations are presented in Table 6.

Table 6. Comparisons of Disaggregate Marginal Effects and Elasticities

Level-of-Service Variable	Auto	Transit	Bicycle	Walk
Disaggregate Marginal Effects				
Model	SGMNL-22			
Auto in-vehicle time	-0.0145	0.0030	0.0038	0.0077
Transit in-vehicle time	0.0016	-0.0020	0.0002	0.0003
Transit service frequency per hour	-0.0026	0.0033	-0.0003	-0.0005
Bicycle travel time	0.0049	0.0004	-0.0066	0.0013
Walk travel time	0.0054	0.0004	0.0007	-0.0066
Model	MNL			
Auto in-vehicle time	-0.0187	0.0024	0.0055	0.0108
Transit in-vehicle time	0.0012	-0.0020	0.0003	0.0005
Transit service frequency per hour	-0.0017	0.0028	-0.0004	-0.0007
Bicycle travel time	0.0054	0.0005	-0.0082	0.0023
Walk travel time	0.0054	0.0005	0.0011	-0.0070
Disaggregate Elasticities				
Model	SGMNL-22			
Auto in-vehicle time	-0.124	0.390	0.154	0.154
Transit in-vehicle time	0.021	-0.417	0.010	0.010
Transit service frequency per hour	-0.026	0.519	-0.012	-0.012
Bicycle travel time	0.099	0.119	-0.646	0.063
Walk travel time	0.322	0.385	0.203	-0.910
Model	MNL			
Auto in-vehicle time	-0.162	0.221	0.221	0.221
Transit in-vehicle time	0.017	-0.287	0.017	0.017
Transit service frequency per hour	-0.018	0.311	-0.018	-0.018
Bicycle travel time	0.112	0.112	-0.793	0.112
Walk travel time	0.325	0.325	0.325	-1.004

It can be seen that cross-elasticities are equal in the MNL model, which reflects its IIA property. However, with unequal variances in auto and transit utilities in the SGMNL model, cross-elasticities for auto and transit choice probabilities are not equal, thus demonstrating that the SGMNL model does not possess the IIA property. However, because the random components in bicycle and walk utilities have equal variance, cross-elasticities for these two alternatives are still equal and therefore the IIA property holds for the bicycle and walk modes even in the case of the SGMNL model. This is similar to the situation where two alternatives belong to the same nest in a nested logit model.

4.5. A Comparison of Changes in Transit Choice Probability in Response to a Service Frequency Improvement

To further illustrate the differences in policy implications obtained from SGMNL and MNL models, changes in transit choice probability in response to a service frequency improvement are compared for the specific individual commuter described previously. The result of this comparison is presented in Figure 4. Relative to the SGMNL model, the MNL model overestimates the transit choice probability when the service frequency is low (<18 per hour) but underestimates it when the service frequency is high (⩾18 per hour). A service frequency of 18 transit vehicles per hour is quite high, reflecting a headway just over three minutes. In most transit service configurations, service frequencies are likely to be less than that number. This implies that the MNL model is likely to overestimate the transit choice probability when it is applied in most real-world contexts where frequency of transit service is less than 18 vehicles per hour. When the service frequency is very low (⩽4 per hour), the relative difference between the predicted transit choice probabilities computed from the MNL and SGMNL models can be more than 50%.

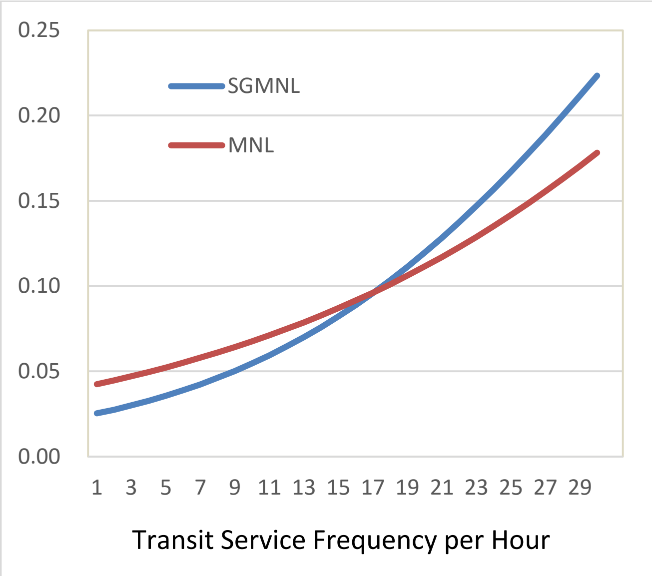

Fig. 4. Comparison of Changes in Transit Choice Probability for a Specific Commuter in Response to an Improvement in Service Frequency

5. Conclusions

In this paper, a semi-nonparametric generalized multinomial logit (SGMNL) model is formulated and developed by applying orthonormal Legendre polynomials to extend the standard Gumbel distribution that lies at the core of multinomial logit models applied in practice. The semi-nonparametric function with flexible forms can represent a probability density function for a large family of multimodal distributions. Unlike the existing semi-nonparametric modeling method which is applied to binary choice situations in the econometric literature, the proposed method allows for modeling multinomial choices, which are typically encountered in travel-related choice behavior analysis and travel demand modeling. The advantage of the proposed method is that the formulation results in a closed-form likelihood function and standard maximum likelihood estimation methods can be applied for parameter estimation. Thus, the model estimation procedure is computationally efficient and free from simulation-based complexity or simulation errors.

The proposed modeling method is applied to an empirical setting of commute travel mode choice among four alternatives (auto, transit, bicycle and walk), based on travel survey and network skim (level of service) data from the Canton of Argau in Switzerland. It is found that the distribution of the random component in the auto utility function is similar to a Gumbel distribution, but has substantially smaller variance. More notably, the random component in the transit utility function follows a bimodal distribution, which indicates a significant departure from and violation of the assumption of a Gumbel distribution for the transit utility random error term. Unequal variances accommodated in the formulation allow the semi-nonparametric model to not be restricted by features of the IIA property that are inherent to the multinomial logit model. The semi-nonparametric model specifications are found to offer superior goodness-of-fit when compared with the MNL model. The violation of the standard Gumbel distribution assumption in the multinomial logit model leads to inconsistent coefficient estimates, marginal effects, elasticities and choice probabilities. In the empirical context examined in this study, the transit choice probability is found to be overestimated in the multinomial logit model when the service frequency takes a value within a normal range. The findings reported in the paper are based on comparisons between the semi-nonparametric model and multinomial logit model.

A few limitations of the proposed method and future research directions are worthy of note. At first, it may be challenging to directly apply the proposed method to model choice behaviors in the context of a large choice set. The likelihood function, depicted in Equation (18), involves multiple levels of summations and the number of levels is dependent on the number of alternatives in the choice set. Thus, the computational complexity will geometrically increase with an increase in the number of alternatives in the choice set. Future research may be focused on reducing computational complexity in the context of large choice sets. Second, the proposed model is developed based on the assumption that random components in utility functions are mutually independent. However, this assumption may not hold in empirical settings. In future research, there may be the potential to introduce correlations in joint semi-nonparametric distributions and develop nested or cross-nested versions of the proposed semi-nonparametric multinomial choice model. Third, it is uncertain whether the empirical result seen in this study, i.e., the random component in the transit utility follows a bimodal distribution, applies in different geographical and modal contexts. In future research, travel survey data from other areas should be used to address this question.

Acknowledgements

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.E., S. Lerman, 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, Massachusetts, U.S.

Bhat, C.R., 1995. A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice. Transportation Research Part B: Methodological 29(6), 471-483.

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

Chen, H.Z., A. Randall, 1997. Semi-nonparametric Estimation of Binary Response Models with an Application to Natural Resource Valuation. Journal of Econometrics 76(1-2), 323-340.

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

Crooker, J.R., 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), 451-480.

Fosgerau, M., M. Bierlaire, 2007. A Practical Test for the Choice of Mixing Distribution in Discrete Choice Models. Transportation Research Part B: Methodological 41, 784–794.

Gallant, A.R., D. Nychka, 1987. Semi-nonparametric Maximum Likelihood Estimation. Econometrica 55, 363–390.

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

Lee, L.F., 1995. Semiparametric Maximum Likelihood Estimation of Polychotomous and Sequential Choice Models. Journal of Econometrics 65, 381-428.

Manski, C.F., 1975. Maximum Score Estimation of the Stochastic Utility Model of Choice. Journal of Econometrics 3(3), 205–288.

McFadden, D.L., 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka, ed., *Frontiers in Econometrics*. New York: Academic Press, 105–142.

Train, K.E., 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press.

Ye, X., V.M. Garikapati, D. You, R.M. Pendyala, 2016. A Practical Method to Test Validity of the Standard Gumbel Distribution in Logit-based Multinomial Choice Models of Human Travel Behavior. Working Paper in the School of Transportation Engineering, Tongji University, Shanghai, China.

Appendix

Lemma 1: $[G(\varepsilon)]^m = G[\varepsilon - \ln(m)]$, where $m > 0$.
Proof: $[G(\varepsilon)]^m = \exp(-e^{-\varepsilon})^m = \exp(-m \cdot e^{-\varepsilon}) = \exp\{-e^{-[\varepsilon - \ln(m)]}\} = G[\varepsilon - \ln(m)]$.

Lemma 2: $G(a+\varepsilon)G(b+\varepsilon) = G(c+\varepsilon)$, where $c = -\ln(e^{-a} + e^{-b})$. More generally, $\prod_{j=1}^{J} G(a_j + \varepsilon) = G(c+\varepsilon)$, where $c = -\ln\left(\sum_{j=1}^{J} e^{-a_j}\right)$.
Proof: $G(a+\varepsilon)G(b+\varepsilon) = \exp(-e^{-a-\varepsilon})\exp(-e^{-b-\varepsilon}) = \exp(-e^{-a-\varepsilon} - e^{-b-\varepsilon})$
$= \exp\left[-e^{-\varepsilon}\left(e^{-a} + e^{-b}\right)\right] = \exp\left[-e^{-\varepsilon + \ln\left(e^{-a} + e^{-b}\right)}\right] = G(c+\varepsilon)$, where $c = -\ln(e^{-a} + e^{-b})$.
More generally, $\prod_{j=1}^{J} G(a_j + \varepsilon) = \prod_{j=1}^{J} \exp(-e^{-a_j - \varepsilon}) = \exp(\sum_{j=1}^{J} -e^{-a_j - \varepsilon})$
$= \exp\left[-e^{-\varepsilon} \sum_{j=1}^{J} e^{-a_j}\right] = \exp\left[-e^{-\varepsilon + \ln\left(\sum_{j=1}^{J} e^{-a_j}\right)}\right] = G(c+\varepsilon)$, where $c = -\ln\left(\sum_{j=1}^{J} e^{-a_j}\right)$.

Lemma 3: $\int_{-\infty}^{+\infty} G(x+c)\, g(x) dx = \frac{1}{1+e^{-c}}$.
Proof: $\int_{-\infty}^{+\infty} G(x+c)\, g(x) dx = \int_{-\infty}^{+\infty} \exp(-e^{-x-c}) \exp(-e^{-x}) e^{-x} dx$
$= -\int_{-\infty}^{+\infty} \exp(-e^{-c} e^{-x}) \exp(-e^{-x}) d(e^{-x})$.
Let $y = e^{-x}$, then $\int_{-\infty}^{+\infty} G(x+c)\, g(x) dx = -\int_{+\infty}^{0} \exp(-y \cdot e^{-c}) \exp(-y) d(y)$
$= -\int_{+\infty}^{0} \exp[-y \cdot (1 + e^{-c})]\, d(y)$.
Let $k = (1 + e^{-c})$, then $\int_{-\infty}^{+\infty} G(x+c)\, g(x) dx = -\int_{+\infty}^{0} e^{-ky}\, d(y) = \frac{1}{k} \int_{+\infty}^{0} e^{-ky}\, d(-ky)$.
Let $q = -ky$, then $\int_{-\infty}^{+\infty} G(x+c)\, g(x) dx = \frac{1}{k} \int_{-\infty}^{0} e^{q}\, dq = \frac{1}{k} [e^{q}]_{-\infty}^{0} = \frac{1}{k} = \frac{1}{1+e^{-c}}$.