Imperial College London

The Multinomial Logit Model Revisited: A Semiparametric Approach in Discrete Choice Analysis

Dr. Baibing Li, Loughborough University

Wednesday, 02 February 2011 - 16:00

Location: Room 610, Skempton (Civil Eng.) Bldg, Imperial College London

Abstract

The multinomial logit model is widely used in transport research. It has long been known that the Gumbel distribution forms the basis of the multinomial logit model. Although the Gumbel distribution is a good approximation in some applications, it is chosen mainly for mathematical convenience. This can be restrictive in many scenarios in practice. We show in this presentation that the assumption of the Gumbel distribution can be substantially relaxed to include a large class of distributions that is stable with respect to the minimum operation. The distributions in the class allow heteroscedastic variances. We then seek a transformation that stabilizes the heteroscedastic variances. We show that this leads to a semiparametric choice model which links travel-related attributes to the choice probabilities via a sensitivity function. This sensitivity function reflects the degree of travellers' sensitivity to the changes in the combined travel cost. Empirical studies were conducted using the developed method.

Biography

Baibing Li is a Reader in Business Statistics & Management Science, School of Business and Economics, Loughborough University. He has previously been a Lecturer in Statistics in School of Mathematics and Statistics at Newcastle University.

The Multinomial Logit Model Revisited:
A Semiparametric Approach
in Discrete Choice Analysis

Baibing Li

School of Business & Economics
Loughborough University

- **Introduction**

- **A distribution class for discrete choice analysis**

- **Semiparametric discrete choice model**

- **Model estimation**

- **Empirical studies**

- **Discussion and conclusions**

- **Why multinomial logit model?**

  - Widely used in transport research

  - Simple and easy to understand in terms of both statistical inference and computation

  - Particularly attractive in many modelling scenarios due to the nature that it is linked to the decision-making process via the maximising (minimising) the utility (travel cost)

- **The underlying assumptions for the logit model**

  In the derivation of the closed-form multinomial logit model, there are three underlying assumptions (McFadden, 1978; Ben-Akiva and Lerman, 1985; Train, 2003; Bhat et al., 2008; Koppelman, 2008), i.e. the random variables of interest are assumed

  - to be independent of each other (assumption I)

  - to have equal variability across cases (assumption II)

  - to follow the Gumbel distribution (assumption III)

- Extensions of the multinomial logit model may be classified into two categories: open-form and closed-form. We mainly focus on the closed-form choice models

- **Existing researches for the closed-form logit model**

  - Relaxation of assumption I to allow dependence or correlation
    - The nested logit model and generalized extreme value (GEV) family (McFadden, 1978)
    - More recent development: paired combinatorial logit (PCL), cross-nested logit (CNL), and generalized nested logit (GNL)

  - Relaxation of assumption II to allow unequality of the variance
    - HMNL: the heteroscedastic multinomial logit model allows the random error variances to be non-identical across individuals/cases (Swait and Adamowicz, 1996)
    - COVNL: the covariance heterogeneous nested logit model was developed on the basis of the nested logit model and it allows heterogeneity across cases in the covariance of nested alternatives (Bhat, 1997)

■ **The research in this study**

- The purpose of this study is to relax assumption III on the underlying distribution: the Gumbel distribution

- Practical motivations:

  Logit model is used in a variety of the problems in transport research. It is hard to believe that a single statistical distribution (the Gumbel) can accommodate such a variety of applications

- Theoretical motivations:

  - Castillo et al. (2008) have proposed using the Weibull distribution as an alternative to the Gumbel distribution

  - Fosgerau and Bierlaire (2009) show that the assumption of the Weibull distribution is associated with the discrete choice model having multiplicative error terms

- **Research question: Are there any other distributions?**

- **Extension from the Gumbel to a general distribution class**
  - **Context**

    Discrete choice analysis can be investigated in various contexts. Consider several travellers who wish to minimize their travel costs

  - **Notation**
    - ◆ C_n denotes the feasible choice set of each individual n
    - ◆ Y_{in} denotes the random travel cost for traveler n when choosing alternative i
    - ◆ We assume the random costs are independent of each other

  - **Theory of individual choice behaviour**

    The probability that any alternative i in C_n is chosen by traveler n is

    $$P_n(i) = \Pr\{Y_{in} < Y_{jn} \text{ for all } j \text{ in } C_n\} = \Pr\{Y_{in} < \min(Y_{jn}) \text{ for } i \neq j\}$$

A new distribution class

Ordinary logit model

- Assumed distribution:

  Gumbel distribution

- Equal variability assumption

  the variance retains constant across all i and n

- Closed under the min-operation

  If Y_{jn} are independent of each other and all follow the Gambel, then $\min\{Y_{jn}\}$ also does

New choice model

- Assumed distribution:

  $F_{in}(t) = \Pr\{Y_{in} < t\} = 1 - [1 - F(t)]^{\alpha_{in}}$

  where the base function $F(t)$ can be any CDF

- Unequal variability assumption

  the variance varies across different cases

- Closed under the min-operation

  If Y_{jn} are independent of each other and all follow a distribution from the above distribution class, then $\min\{Y_{jn}\}$ also does

- **The new class of distributions**

$$F_{in}(t) = \Pr\{\, Y_{in} < t \,\} = 1 - [1 - F(t)]^{\alpha_{in}}$$

- This distribution class includes both the Gumbel and Weibull distributions as its special cases, as well as many others such as

  - ◆ Pareto
  - ◆ Gompertz
  - ◆ Expoenetial
  - ◆ Rayleigh
  - ◆ generalised logistic

- **The parametric approach**

$$F_{in}(t) = \Pr\{ Y_{in} < t \} = 1 - [1 - F(t)]^{\alpha_{in}}$$

- Have knowledge of the random variables a priori
- Specify a base function $F(t)$ in the stage of modelling
- The statistical inference focuses on several unknown parameters

- **A semiparametric approach**
  - Have little knowledge of the distribution of the random variables
  - Do not specify a base function $F(t)$ in the stage of modelling
  - The statistical inference includes both the unknown parameters AND the unknown base function
  - From a practical perspective, the assumption that the random travel costs $F_{in}(t)$ follow any distribution from the distribution family with an unspecified base function $F(t)$ allows researchers great flexibility to accommodate different problems

- **Variance-stabilizing transformation**

  Theorem 1. Suppose that random variables Y_i ($i=1,\ldots,m$) have the following CDFs:

  $$F_i(t)=\Pr\{Y_i < t\}= 1- [1- F(t)]^{\alpha_i} \text{ with } (i=1,\ldots,m),$$

  where $F(t)$ is any chosen CDF. Then there exists a monotonically increasing transformation $h(t)$ such that the transformed random variables have a common variance.

  - The fact that the proposed distribution class allows unequal variances suggests that it is more flexible to accommodate various practical problems
  - The unequal variances may be stabilized via a suitable transformation $h(t)$

- **The mean function**

  Let V_{in} denote the expectation of random travel cost Y_{in}, i.e. $EY_{in} = V_{in}$

  Theorem 2. Suppose that random variables Y_i ($i=1,\ldots,m$) have the following CDFs:

  $$F_i(t)=\Pr\{\,Y_i < t\,\}= 1 - [1 - F(t)]^{\alpha_i} \text{ with } \quad (i=1,\ldots,m),$$

  where $F(t)$ is any chosen CDF. Then there exists a monotonically decreasing function $H(t)>0$ such that expectations $EY_i = V_i$ are linked to the parameter α_i

  $$\alpha_i = H(V_i)$$

  - Special case: $H(t) = 1/t$ for the exponential distribution

■ **Choice probability**

● We suppose that the expectations $EY_{in} = V_{in}$ are linked to a linear function of a q-vector of attributes that influences specific discrete outcomes: $V_{in} = \mathbf{x}_{in}^T \boldsymbol{\beta}$

● Combining the mathematical expectations $V_{in} = \mathbf{x}_{in}^T \boldsymbol{\beta}$ with the mean function $\alpha_{in} = H(V_{in})$ gives $\alpha_{in} = H(\mathbf{x}_{in}^T \boldsymbol{\beta})$

● Note that $\min\{Y_{jn}\}$ follows the same distribution as Y_{in}

● It can be shown that the choice probability is

$P_n(i) = \Pr\{Y_{in} < Y_{jn} \text{ for all } j \text{ in } C_n\} = \Pr\{Y_{in} < \min(Y_{jn}) \text{ for } i \neq j\}$

$= H(\mathbf{x}_{in}^T \boldsymbol{\beta}) / \{\Sigma_j H(\mathbf{x}_{jn}^T \boldsymbol{\beta})\}$

# Semiparametric discrete model

- **Sensitivity function** $S(.)$

  Define $S(.)=\log[H(.)]$ so the range of $S(.)$ is the whole real line:

  $$P_n(i)=\exp[S(\mathbf{x}_{in}^T\boldsymbol{\beta})] / \{\Sigma_j\exp[S(\mathbf{x}_{jn}^T\boldsymbol{\beta})]\}$$

  - $S(.)$ reflects how sensitive a traveler is to the changes in the combined travel cost (including travel time, travel expenses, etc.)

  - When $S(t)=-\theta t$, the model reduces to the logit model and the corresponding underlying distribution is the Gumbel.

  - The above semiparametric choice model extends the logit model by allowing an unspecified functional form $S(.)$ can address issues: (a) nonlinearity; and (b) variance stabilization.

- A linear function $S(t)$ provides a benchmark for comparison

- The dotted line represents the scenario where travelers are more sensitive to one unit increment in travel costs

- The broken line represents the scenario where travelers are more tolerable to the increment in the combined travel cost
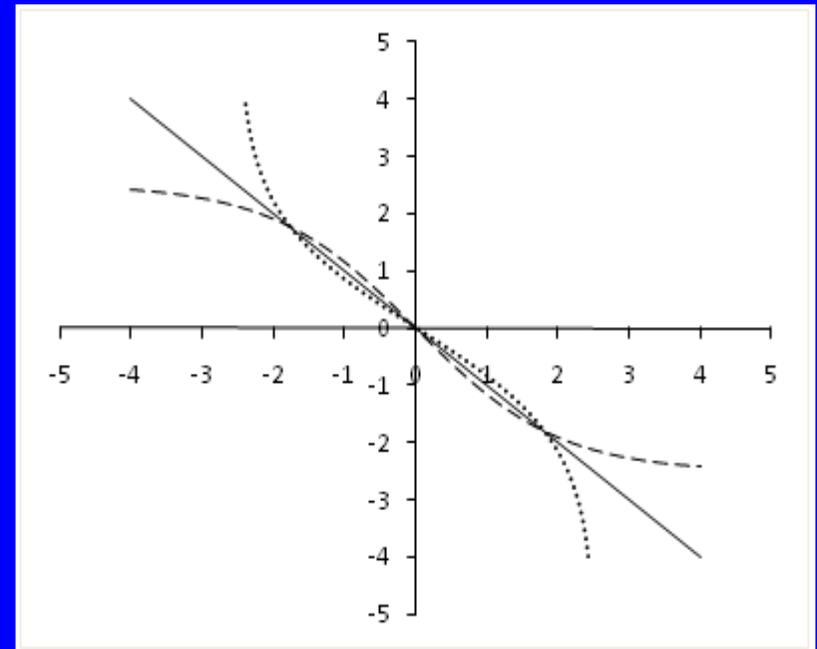
■ **The parametric model**

● If the base function is specified in the stage of modelling, it is required to estimate the coefficients of the attributes, **β**

● The estimation can be done similar to the logit model

■ **The semiparametric model**

● Since the base function is not specified in the stage of modelling, it is required to estimate the coefficients of the attributes **β** and the sensitivity function $S(.)$

- **Identifiability**

- Identifiable up to a level constant and scale constant
  - ◆ Let $S(t) = R(bt)$, then $S(\mathbf{x}^T\boldsymbol{\beta}) = R(\mathbf{x}^T\boldsymbol{\beta}b)$
    $\{S(t), \boldsymbol{\beta}\}$ and $\{R(t), \boldsymbol{\beta}b\}$ fit the given data equally well
  - ◆ Let $S(t) = R(a+t)$, then $S(\mathbf{x}^T\boldsymbol{\beta}) = R(a+\mathbf{x}^T\boldsymbol{\beta})$

- Due to the issue of identifiability, it is required that the linear combination of attributes $\mathbf{x}^T\boldsymbol{\beta}$ does not include an intercept, and that $\boldsymbol{\beta}$ has unit length and one of its entry (say the first one) has a positive sign

- Following Ichimura (1993), some further conditions need to be imposed. In particular $S(.)$ is required not to be constant on the support of $\mathbf{x}^T\boldsymbol{\beta}$. The vector of attributes $\mathbf{x}$ should also admit at least one continuously distributed component.

■ **How to estimate the unknown sensitivity funciton**

● Use B-splines to approximate $S(.)$:

$$S(t) \approx \Sigma_j \, w_j B_j(t),$$

where $B_j(t)$ ($j=1,\ldots,m$) are known basis functions (cubic splines) and w_j are unknown weights to be estimated

● The accuracy of the approximation is guaranteed as m is large

● Since the basis functions $B_j(t)$ ($j=1,\ldots,m$) are known, we only need to estimate weights w_j

■ **Bayesian analysis**

Performing Bayesian analysis to draw statistical inference

● Data:

◆ Let y_{in} be 1 if traveller n chose alternative i and 0 otherwise.

◆ Let **X** and **Y** denote the data matrices comprising $\mathbf{x}_{jn}$ and y_{in}

● Likelihood: $L(\mathbf{Y}; \boldsymbol{\beta}, \mathbf{w}, \mathbf{X}) = \prod_n \prod_i [P_i(n)]^{y_{in}}$

● Prior distribution: non-informative $p(\boldsymbol{\beta}, \mathbf{w})$

● Posterior distribution: $p(\boldsymbol{\beta}, \mathbf{w} \mid \mathbf{Y}, \mathbf{X}) \propto L(\mathbf{Y}; \boldsymbol{\beta}, \mathbf{w}, \mathbf{X}) \, p(\boldsymbol{\beta}, \mathbf{w})$

● Markov chain Monte Carlo (MCMC): simulate draws from the posterior distribution $p(\boldsymbol{\beta}, \mathbf{w} \mid \mathbf{Y}, \mathbf{X})$

- **Data**
  - Fosgerau et al. (2006) carried out a large-scale Danish value-of-time study that involved stated preferences about two train-related alternatives and two bus-related alternatives respectively
  - Travel time for public transport users was broken down into four components: (a) access/egress time (other modes than public transport, including walking, cycling, etc.); (b) in-vehicle time; (c) headway of the first used mode; and (d) interchange waiting time
  - The attributes considered in their study included these four travel time components, plus the number of interchanges and travel expenses. The travellers' time values were inferred from binary alternative routes characterised by these attributes
  - The original stated preferences are panel data. For illustration purposes, we selected only 100 different travellers from each dataset, and then randomly chose one observation for each traveller (hereafter referred to as 'train data' and 'bus data' respectively) in the following analyses

- **Settings in the computation**

  - The splines used in the following analyses included seven cubic basis functions (j=1,…,7) on the support [0, 1]

  - The total number of iterations in the MCMC simulation was set as 10,000. The first 5,000 iterations were considered as burnt-in period and the corresponding draws were discarded. The results are reported below using the remaining 5,000 draws

- **Models used in the analyses**

  Let x_1, …, x_6 represent the six attributes: access-egress time, headway, in-vehicle-time, waiting time, number of interchanges, and travel expenses. Following Fosgerau and Bierlaire (2009), the coefficient of travel expenses was normalized to unit so that other coefficients can be interpreted as willingness-to-pay indicators

  - the ordinary multinomial logit model
    $$S(\mathbf{x}^T\boldsymbol{\beta}) = -\theta\,(\beta_1 x_1 + \ldots + \beta_6 x_6)$$

  - the multiplicative choice model
    $$S(\mathbf{x}^T\boldsymbol{\beta}) = -\theta\,log(\beta_1 x_1 + \ldots + \beta_6 x_6)$$

  - the semi-parametric model
 $$S(\mathbf{x}^T\boldsymbol{\beta}) = S(u + v((\beta_1 x_1 + \ldots + \beta_6 x_6)))$$
    where u and v has two scaling parameters so that $S(.)$ is on $[0, 1]$

Study I: the train data

Multinominal logit model			
log-likelihood= -54.8	ρ^2=0.21 DIC= 120.2		
Attributes	Posterior mean	Posterior standard deviation	95% credible interval
Access time	0.70	0.39	(0.11, 1.63)
Headway	-1.26	7.23	(-14.00, 14.49)
In-vehicle time	0.32	0.15	(0.04, 0.63)
Waiting time	0.78	0.30	(0.27, 1.41)
Interchange	4.41	1.92	(0.68, 8.72)

Multiplicative choice model			
log-likelihood= -57.2	ρ^2=0.17 DIC=122.6		
Attributes	Posterior mean	Posterior standard deviation	95% credible interval
Access time	1.32	0.49	(0.43, 2.38)
Headway	22.78	29.90	(-30.69, 74.40)
In-vehicle time	0.59	0.34	(0.10, 1.50)
Waiting time	0.97	0.12	(0.77, 1.17)
Interchange	6.35	3.11	(0.88, 13.63)

Semiparametric choice model			
log-likelihood= -34.4	ρ^2=0.50 DIC=83.1		
Attributes	Posterior mean	Posterior standard deviation	95% credible interval
Access time	0.36	0.11	(0.16, 0.61)
Headway	-0.23	3.11	(-5.99, 6.27)
In-vehicle time	0.07	0.03	(0.01, 0.13)
Waiting time	0.39	0.11	(0.17, 0.62)
Interchange	0.87	0.39	(0.13, 1.65)

* The value of log-likelihood is -69.3 when all the parameters are set equal to zero.

- The middle part of obtained sensitivity function is not sensitive to the change of the combined travel cost

- Towards to the both extreme ends of the support, it increases (or decreases) rapidly

- Each unit increment in the combined travel cost does not impact on the train users equally
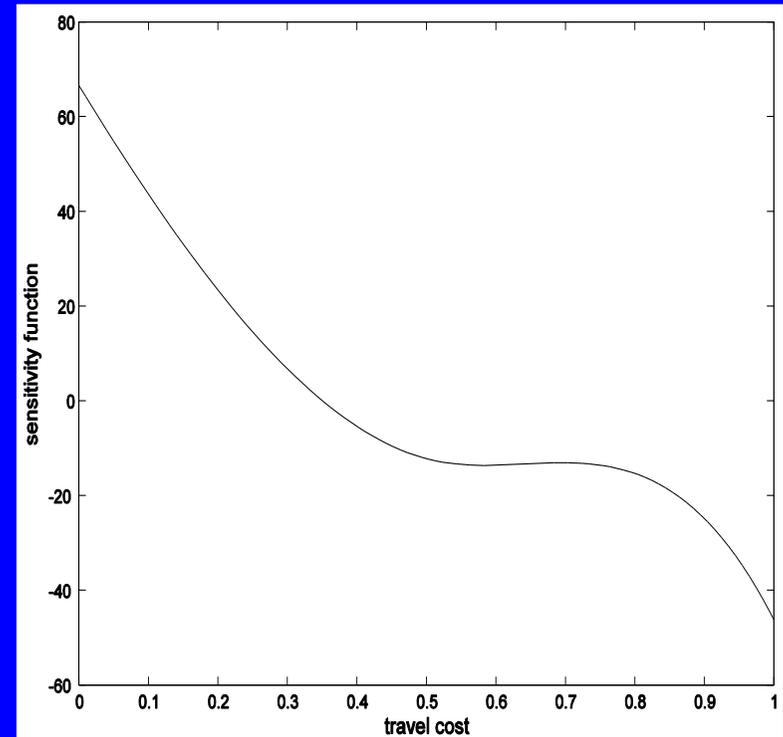
Study II: the bus data

Multinominal logit model			
log-likelihood= -52.2	ρ^2=0.24 DIC=116.4		
Attributes	Posterior mean	Posterior standard deviation	95% credible interval
Access time	0.30	0.17	(0.05, 0.68)
Headway	-4.92	3.53	(-11.54, 1.98)
In-vehicle time	0.01	0.03	(-0.04, 0.07)
Waiting time	0.47	0.20	(0.11, 0.88)
Interchange	1.37	0.51	(0.45, 2.42)

Multiplicative choice model			
log-likelihood= -52.5	ρ^2=0.24 DIC=114.3		
Attributes	Posterior mean	Posterior standard deviation	95% credible interval
Access time	0.37	0.17	(0.05, 0.66)
Headway	-1.47	1.21	(-2.99, 0.97)
In-vehicle time	0.13	0.03	(0.09, 0.18)
Waiting time	0.41	0.12	(0.19, 0.62)
Interchange	0.97	0.53	(0.03, 1.76)

Semiparametric choice model			
log-likelihood= -52.5	ρ^2=0.24 DIC=114.6		
Attributes	Posterior mean	Posterior standard deviation	95% credible interval
Access time	0.38	0.21	(0.04, 0.88)
Headway	-2.84	4.66	(-14.05, 4.47)
In-vehicle time	0.01	0.06	(-0.15, 0.09)
Waiting time	0.42	0.22	(0.04, 0.91)
Interchange	1.20	0.74	(0.11, 3.11)

* The value of log-likelihood is -69.3 when all the parameters are set equal to zero.

- The obtained sensitivity function is quite close to a linear function.

- The semiparametric model produced similar estimates to that of the ordinary multinomial logit model

- Due to its simplicity, it seems that the ordinary multinomial logit model is a sensible choice
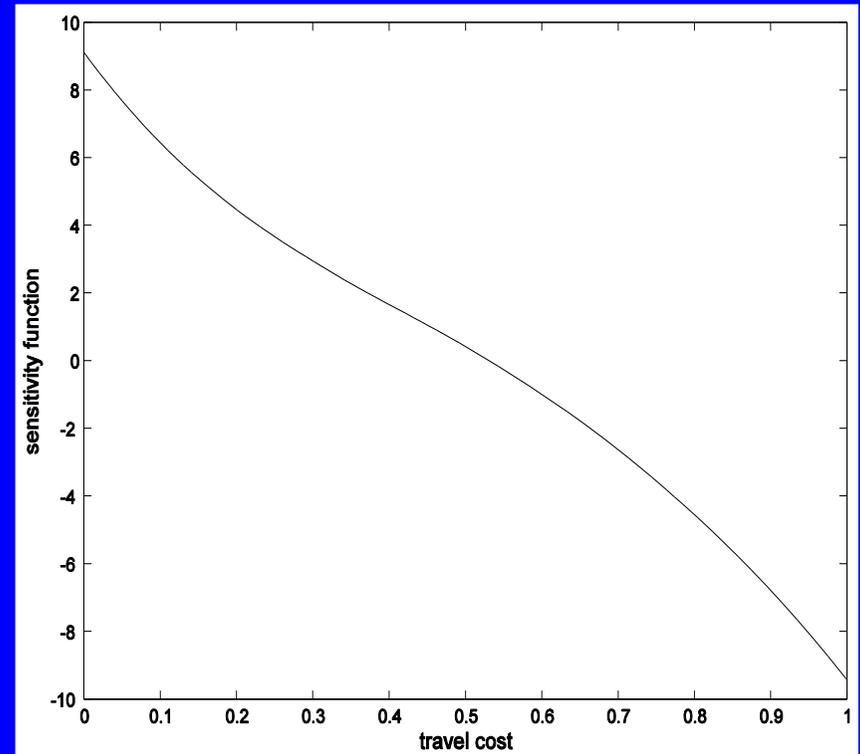
Discussion and conclusions

- **Relaxation of assumption III**
  - The assumption of underlying distributions is extended from the Gumbel to a much wider distribution class
  - It also retains a crucial property in discrete decision analysis, i.e., it is closed under the minimum operation
  - It allows unequal variances across cases

- **Semiparametric choice model and sensitivity function**
  - In the modeling stage the distribution needs not to be specified
  - A semiparametric choice model is derived that links travel-related attributes to the choice probabilities via a sensitivity function
  - When the sensitivity function is nonlinear, travelers' response to the travel cost does not change in a proportionate manner. This has important practical implications for the policy makers

- **The logit model assumptions revisited**
  - Three assumptions for the multinomial logit model:
    - Independence across the cases (assumption I)
    - Equal variability across cases (assumption II)
    - The Gumbel distribution (assumption III)

  - The semiparametric model has substantially relaxed Assumption III and hence Assumption II

  - Assumption I? --- Can the correlation structure be relaxed?

    For stated preferences data, for instance, random effect of individual should be taken into account:

    $$Y_{in} = V_{in} + d_n + e_{in}$$

    where the errors e_{in} are independent but for the same traveller, Y_{in} and Y_{jn} are correlated due to the common random effect d_n

■ **The way to take forward**

- The multinomial logit model is frequently used as a building block in discrete choice analysis to handle more complex scenarios

- In particular, the multinomial logit model can be combined with a random-coefficients structure, leading to the mixed logit (Train, 2003; Bhat et al., 2008)

- **Question**

  For the semiparametric model, can it be combined with a random-coefficients structure to relax Assumption I?

■ **A random coefficient structure**

- Following the mixed logit, we assume that the coefficients vary across travellers in the population with density $q(\boldsymbol{\beta})$ so that the heterogeneity across travellers can be taken into account

- For each traveller, it is assumed that the semiparametric choice probability still holds

$$L_{in}(\boldsymbol{\beta}) = \exp[S(\mathbf{x}_{in}{}^T\boldsymbol{\beta})] \,/\, \{\Sigma_j \exp[S(\mathbf{x}_{jn}{}^T\boldsymbol{\beta})]\}$$

- For each traveller n, since the researcher observes $\mathbf{x}_{jn}$ but not $\boldsymbol{\beta}$, the unconditional choice probability is the integral of over all possible variable of $\boldsymbol{\beta}$:

$$P_n(i) = \int L_{in}(\boldsymbol{\beta})\, q(\boldsymbol{\beta})\, d\boldsymbol{\beta}$$

- This mixed version of the semiparametric model does not exhibit the IIA property and thus is more flexible

- **How the variability is modelled?**

  - The existing mixed logit model

    The ordinary multinomial logit assumes equal variance. Hence all heterogeneity across travellers and across alternatives are modeled solely by $q(\beta)$

  - The mixed semiparametric choice model

    The heterogeneity across alternatives and the heterogeneity across travelers are dealt with separately:

    - ◆ Variability within a traveller: $F(.)$ allows unequal variances across alternatives within a traveller
    - ◆ Variability between travellers: it is modeled by $q(\beta)$
    - ◆ Different sources of variability are modeled separately. It is more straightforward for model specification and interpretation in the mixed semiparametric choice model