

Comparative study on mode split discrete choice models

Xianlong Chen · Xiaoqian Liu · Fazhi Li

Received: 6 September 2013 / Revised: 18 November 2013 / Accepted: 21 November 2013 / Published online: 20 December 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract Discrete choice model acts as one of the most important tools for studies involving mode split in the context of transport demand forecast. As different types of discrete choice models display their merits and restrictions diversely, how to properly select the specific type among discrete choice models for realistic application still remains to be a tough problem. In this article, five typical discrete choice models for transport mode split are, respectively, discussed, which includes multinomial logit model, nested logit model (NL), heteroscedastic extreme value model, multinomial probit model and mixed multinomial logit model (MMNL). The theoretical basis and application attributes of these five models are especially analysed with great attention, and they are also applied to a realistic intercity case of mode split forecast, which results indicating that NL model does well in accommodating similarity and heterogeneity across alternatives, while MMNL model serves as the most effective method for mode choice prediction since it shows the highest reliability with the least significant prediction errors and even outperforms the other four models in solving the heterogeneity and

similarity problems. This study indicates that conclusions derived from a single discrete choice model are not reliable, and it is better to choose the proper model based on its characteristics.

Keywords Discrete choice model · Mode split · NL · MMNL · HEV · MNP

1 Introduction

A good understanding on the travellers' mode choice behaviours serves as one of the prerequisites for passenger transport policy-making. Being important tools of travellers' mode choice behaviour studies, discrete choice models are widely used both in theory and practice within transportation planning field.

In the context of transportation, travellers tend to choose the transport mode which fits them best, and in the meanwhile, transportation means tend to 'choose' the most capable travellers as well. As a result, the decision making process of travellers' mode choice is influenced by the attributes of transportation means as well as the internal factors (individual attributes) of the travellers themselves, such as economic capability, personal preference, etc. Since several factors affect the description and the prediction accuracy of the mode choice behaviours, selecting a suitable discrete choice model with good interpretation ability appears to be very critical.

The discrete choice model used in early times is the multinomial logit model (MNL). The rigid assumption that the utility random terms of alternative parts satisfy independent identical distribution (IID) conditions makes MNL simple in calculation as well as provides MNL with independence of irrelevant alternatives (IIA) property, which

X. Chen (✉)
Guangzhou Transport Planning and Research Institute, 11th
Floor, No. 80 of Jixiang Road, Guangzhou 510030, China
e-mail: harleych@163.com

X. Liu
Transportation Planning and Design Division, Architectural
Design & Research Institute of Tongji University (Group) Co.,
Ltd., Shanghai 200092, China

F. Li
Guangzhou Transport Planning and Research Institute,
Guangzhou 510030, China

weakens its ability of recurring the actual choice behaviours [1].

The IIA property of MNL model stems from the rigid assumption that the utility random terms of alternatives parts have totally independent structures. To relax the MNL model’s IIA property, and in the meanwhile, keep its calculation convenience, the researchers gradually relax the restrictions on the assumption of utility random terms structures and successively explore and develop several MNL-based models, which are more capable of recurring decision-makers’ choice behaviours, such as nested logit model (NL) [2], generalised extreme value model [3, 4], heteroscedastic extreme value model (HEV) [5], mixed multinomial logit model (MMNL) [6], etc. Many scholars [7–10] have analysed the travel mode of urban commuters. Schmidt and Strauss [11] and Boskin [12] analysed occupational choice among multiple alternatives. Rossi and Allenby [13] studied consumer brand choices in a repeated choice (panel data) model. Train [14] studied the choice of electricity supplier by a sample of California electricity customers. Hensher et al. [15] analysed choices of automobile models by a sample of consumers that offered a hypothetical menu of features. In each of these cases, there is a single decision among two or more alternatives. Faced with so many discrete choice models, how to choose an appropriate one to simulate travel behavior is still a rather difficult problem. In this article, we focused on transport mode choice behaviour modelling and made a comparison between five typical discrete choice models and discussed the rules for choosing the optimal discrete choice model.

2 MNL model and its application restrictions

Stemming from psychology and economics, discrete choice theory has become a mainstream since 1980s. Most discrete choice theory studies are grounded on the utility functions, which are expressed as

$$U_{nj} = V_{nj} + \varepsilon_{nj} = \gamma'_n Z_{nj} + \varepsilon_{nj} = \beta'_n x_{nj} + \alpha'_n y_{nj} + \varepsilon_{nj}, \quad (1)$$

where U_{nj} denotes the utility that the decision-maker n associates with alternative j ; V_{nj} denotes measurable utility; ε_{nj} is the error term (immeasurable utility), γ'_n is the parameter vector of decision-maker n , Z_{nj} is the observed variable; x_{nj} is the individual attributes vector of decision-maker n , y_{nj} is attributes vector of alternative j ; $n \in N$, N denotes the amount of decision-makers; $j \in J$, J denotes the amount of alternatives.

In the modelling process of individual choice models, assume that the consumers are rational choosers; therefore, the probability that individual n selects programme i is

$$P_{ni} = P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}, \quad \forall j), \quad i, j \in J, \quad (2)$$

$$= P(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj}, \quad \forall j \neq i).$$

Equation (2) assumes that the error term ε_{nj} satisfies independent and identically distributed (IID) assumption and subjects to type I extreme value distribution:

$$f(\varepsilon_{nj}) = \lambda e^{-\lambda(\varepsilon_{nj}-\eta_{nj})} \exp\left[-e^{-\lambda(\varepsilon_{nj}-\eta_{nj})}\right], \quad (3)$$

where η_{nj} and λ denote the location parameter and dispersion parameter, respectively. The variance of this distribution is $\pi^2/6\lambda^2$. Let $\lambda = 1$ and $\eta_{nj} = 0$, the MNL choice probability model takes the form of

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}}. \quad (4)$$

In terms of measurable utility, it is usually defined as linear-in-parameter specification, i.e. $V_{ni} = \beta_n x_{ni} + \alpha y_{ni}$. Thus, the probability that individual n chooses alternative i can be expressed as:

$$P_{ni} = \frac{e^{\beta_n x_{ni} + \alpha y_{ni}}}{\sum_{j=1}^J e^{\beta_n x_{nj} + \alpha y_{nj}}}. \quad (5)$$

The following equation is produced by a simple derivation from Eq. (5) and shows the IIA property of multinomial logit model:

$$\frac{P_{ni}}{P_{nj}} = \frac{\exp(\beta_n x_{ni} + \alpha y_{ni})}{\exp(\beta_n x_{nj} + \alpha y_{nj})}, \quad i \neq j, \quad i, j \in J. \quad (6)$$

Equation (6) means that among all the alternative sets, the ratio of choice probabilities of any two of the decision-maker’s alternatives only associates with the utilities of these two alternatives but has nothing to do with the utilities of any other alternatives.

On the other hand, if we assume that the decision individual n makes is affected by his personal attributes x_{ni} , then MNL model in Eq. (5) can be rewritten as:

$$P_{ni} = \frac{e^{\alpha y_{ni}}}{\sum_{j=1}^J e^{\alpha y_{nj}}}. \quad (7)$$

Equation (7) is the conditional logit model. The feature of this model lies in that all the decisions are merely dependent on the attributes of the chosen alternatives (y_j) but irrelevant to decision-makers’ attributes (X_n). No matter how many sets of alternatives exist, only one group of parameters needs to be estimated because of the assumption that the influences on individual utility of every choice set are identical. If there are many alternatives, then the conditional logit model can be

served as a better choice for modelling and is also simpler to be compared with other models.

In practice, multinomial logit model (MNL) has to satisfy IIA property, which means that, all the alternatives are independent with each other and the ratio of choice probabilities only associates with the utilities of the given alternatives and is irrelevant to the utilities of any other alternatives. IIA property is generated from IID assumption of error term (The error term satisfies independent and identical type I extreme value distribution). IID constrains cannot be guaranteed if heterogeneity and similarity problems remain, which may result in the wrong statistic inference.

Except for IIA property restrictions, MNL model has two shortcomings in application. One is its incapability of handling with random preference discrepancies, the other is its incapability of finding out the correlative factors with panel data. Two merits of MNL are that it has a close-formed structure, and its parameters can be easily estimated.

Some scholars hold that IIA property is totally reasonable in terms of model theory. McFadden and Domencich [16] found that although the IIA restrictions result in a value loss of those studies on MNL, the deviations caused by IIA are owing to study objects, not the theory itself. They believed that IIA property is tenable in homogeneous populations. Ben-Akiva and Lerman [17] further pointed out that although IIA does not fit the whole populations, it does exist in homogeneous populations, such that the reliability of IIA property relies on whether the populations significantly show their heterogeneity. MNL model has the best performance to explain discrete choice behavior if the populations' heterogeneity is not significant.

3 Improvement and development of discrete choice models

Heterogeneity and similarity problems are directly related to the assumption of error terms in model. Observed samples and alternatives will cause the error terms. Therefore, we can consider heterogeneity and similarity problems from the perspective of these two factors. Viewing the heterogeneity issues from the perspective of samples means that the decision-maker holds different viewpoints towards specific transport modes in his mode choice behaviour, which can be called divergent tastes in individuals, or individual heterogeneity. Individual heterogeneity mainly comes from preference heterogeneity and response heterogeneity. The former includes the observed and unobserved effects that individual socioeconomic characteristics put on transport mode choice, and the latter refers to individual evaluation discrepancy on level of service across transport modes, which brings observed and unobserved effects as well.

Similarity across alternatives refers to the situation that similarity issues arise because of spatial or time autocorrelation during the survey process of samples (e.g. repeatedly investigations on the same respondent or samples across sampling objects are self-correlated because of adjacent zone effects, etc.) [5]. Heterogeneity and similarity are prone to biased parameter evaluation or even overestimation on the effects of some specific factors.

When it comes to examining heterogeneity and similarity attributes from the perspective of alternatives, we need to consider whether alternatives share IIA property. If the alternatives appear to be dependent or heterogeneous, there may be similarity and heterogeneity problems among alternatives, which are called alternatives similarity and alternatives heterogeneity, respectively.

In the recent three decades, many models were developed to alleviate the heterogeneity and similarity problems, including NL model [17], HEV model [5], mixed multinomial logit (MMNL) model [10, 15], and multinomial probit (MNP) model [6].

3.1 NL model

NL model introduces the concept of nest layers, in which similar alternatives are put in the same nest layer. Assuming that, the error terms across alternatives in the same nest layer are independently and identically type I extreme value distributed, and the error terms across alternatives which belong to different nest layers are different. Here, we take two-layer nest structure as an example. Suppose that, there are M nests in the model and J_m alternatives in the m th nest layer. Alternative i is one of the alternatives in the m th nest layer, and thus, P_j^i is the probability that decision-maker chooses alternative i :

$$P_j^i = P_m^j \cdot P_j(i|m), \quad (8)$$

where P_m^j represents the marginal probability that decision-maker chooses nest layer m ,

$$P_m^j = \frac{\exp(\mu_m \cdot I_m)}{\sum_{k=1}^M \exp(\mu_k \cdot I_k)}, P_j(i|m);$$

denotes the conditional probability that alternative i of nest layer m is chosen,

$$P_j(i|m) = \frac{\exp(V_i/\mu_m)}{\exp(I_m)};$$

I_m is inclusive value, which means comprehensive utility of nest layer m , $I_m = \ln \sum_{j \in J_m} \exp(V_j/\mu_m)$; μ_m is the parameter of inclusive value, which explains the similarity degree of the alternatives in nests. The estimated inclusive value parameter must subject to $0 \leq \mu_m \leq 1$, i.e. the principle of

utility maximisation. When $\mu_m = 1$, NL model is simplified as MNL model. The more μ_m approaches to zero, the higher the correlation degree among alternatives is.

In order to make all the alternatives independent with each other, NL model imposes all the correlated alternatives on the same independent nest layer and makes use of inclusive value to represent the common utility of these alternatives, and then builds models with other independent alternatives. The NL model is good at solving the similarity problems among alternatives. However, its disadvantages are also evident. First, it has to be assigned a fixed nest layer structure; Second, it is not able to accommodate the situation that all the error terms correlate with each other at the same time; third, decision procedures should be supposed to satisfy the continuity condition; Forth, each alternative is restricted to appear in only one nest.

3.2 HEV model

HEV model is put forward by Bhat [18]. This model allows that alternative ε_{nj} satisfies independent non-uniform type I extreme value distribution, which means that each alternative has its own variance, and the variances may be same or not, but the covariance of different alternatives is zero. The probability that individual n chooses alternative i is

$$P_i = \int_{w=-\infty}^{w=+\infty} \prod_{j \in C, j \neq i} \Psi \left[\frac{V_i - V_j + \theta_i w}{\theta_j} \right] \psi(w) dw, \tag{9}$$

where $\Psi(\cdot)$ and $\psi(\cdot)$ are cumulative distribution function and probability density function of type I extreme value distribution, respectively; C is the choice set; $\text{var}(\varepsilon_i) = \pi^2 \cdot \theta_i^2 / 6$, $w = \varepsilon_{ni} / \theta_i$; θ_i represents heterogeneity parameter of alternatives, and it reflects the degree of uncertain factors, namely the weight of uncertain factors. Different alternatives have different effects on the whole utility. Increasing θ_i will decrease the unit variation that observed utility brings to choice probability.

HEV model allows variance discrepancy (the variance can be identical to each other or not) among error terms across alternatives by introducing scale factor into the expression of error terms, and the covariance across different alternatives is zero. HEV model is only able to handle with the heterogeneity problems among alternatives. During the model application process, large deviation may happen if similarity problems among alternatives simultaneously exist.

3.3 MNP model

Daganzo [6] proposed that the MNP model can be derived if assuming that random error terms follow normal

distribution in Eq. (2). MNP model allows the situation that not all the random error terms are independent and identical with each other. It is the most generalised model as it fully reflects the realistic choice behaviours. The MNP model can be expressed as:

$$P_{ni} = \int I(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj}, \quad \forall j \neq i) \phi(\varepsilon_n) d\varepsilon_n, \tag{10}$$

where $I(\cdot)$ is the index function, $I(\cdot) = 1$ means that the decision-maker has the one with max utility chosen, otherwise is zero; $\phi(\varepsilon_n)$ obeys multivariate normal distribution,

$$\phi(\varepsilon_n) = \frac{\exp\left(-\frac{1}{2} \varepsilon_n' \Omega^{-1} \varepsilon_n\right)}{(2\pi)^{\frac{J}{2}} |\Omega|^{\frac{1}{2}}}, \quad \varepsilon_n' = (\varepsilon_{n1}, \varepsilon_{n2}, \dots, \varepsilon_{nJ}),$$

expectation $E(\varepsilon_n) = \mathbf{0}$, and Ω denote covariance matrix.

Equation (10) has very complex integral components. When there are more than four alternatives within the choice set, it is difficult to estimate parameters.

MNP model is free of the MNL model's three restrictions. It is capable of handling with heteroscedasticity problems, defining error structures of any types as well as dealing with error terms related to time series by using panel data. The only limitation of MNP model is that all the error terms of utility functions must normally distribute. In most cases, assuming that the random terms satisfy normal distribution seems to be proper, but in some cases, this assumption may lead to unconventional prediction results. The most well-known example is about the price variable coefficient, the density distribution of which ought to only appear in the side of distribution greater than zero. Besides, MNP model appears to be much complicated in finding its parameters.

3.4 MMNL model

MMNL model based upon the assumption that decision-makers show different preferences. It assumes that marginal utility obeys Gumbel distribution, and the probability of MMNL has to be obtained by integrating the parameters of MNL model. The probability that decision-maker n chooses alternative i is:

$$P_{ni} = \int L_{ni}(\gamma) g(\gamma) d\gamma = \int L_{ni}(\gamma) g(\gamma|\theta) d\gamma, \tag{11}$$

where $L_{ni}(\gamma)$ is the multinomial logit choice probability along with specific parameter vectors,

$$L_{ni}(\gamma) = \frac{\exp[\mathbf{V}_{ni}(\gamma)]}{\sum_{j=1}^J \exp[\mathbf{V}_{nj}(\gamma)]}. \tag{12}$$

$g(\gamma) = g(\gamma|\theta)$ represents probability density function, θ denotes deep parameter vector, which include mean value,

variance or covariance, etc., and $V_{ni}(\gamma)$ is measurable utility. If the utility is linearly combined, i.e. $V_{ni}(\gamma) = \gamma'Z_{ni}$, then the choice probability of MMNL can be expressed as below:

$$P_{ni} = \int \frac{\exp(\gamma'Z_{ni})}{\sum_{j=1}^J \exp(\gamma'Z_{nj})} g(\gamma) d\gamma. \tag{13}$$

The choice probability of MMNL model relies on the distribution form of γ . $g(\gamma)$ is normally distributed or log normally distributed [19].

In the utility function of MMNL model, except for observed non-random terms and error terms, unobserved random terms are also involved. The correlation, heterogeneity and individual preference heterogeneity of alternatives need to be considered with these random terms. MMNL is able to deal with heterogeneity and similarity problems simultaneously. Thus, the assumption of MMNL model is most practical and performs best in interpreting preference behaviours.

If the utility of MMNL model is set as linear combination, then MNL turns out to be the special case of MMNL model. The merits of MMNL model lie in that preference discrepancy among individuals is allowed, the correlation among different trips of the same consumer can be described, and it can approach to the estimated results of any other random utility models. The demerit of MMNL model is its complex computing process.

Note that, the parameter estimation methods of MNL, NL, HEV, MMNL and MNP models are not totally the same. Generally, the parameters of MNL, NL and HEV models can be identified and obtained by the maximum likelihood estimation method while the unknown parameters of MMNL model and MNP model can only be estimated by the maximum simulated likelihood method.

By analysing discrete choice model on error terms' assumptions from two dimensions, which are independent and identical, five introduced models can be classified as shown in Table 1. Each model has its own merits and demerits, and in application, desirable results can be obtained if they are well combined.

4 Illustrations

The source data used for the comparison of these five discrete choice models were drawn from the questionnaire survey on transport mode (car, train, bus and air) choice behaviours from 210 commuters between Sydney and Melbourne [12]. The main variables include:

- TTME Terminal time, The TTME for car is zero (min)
- INVT In-vehicle time (min)
- GC Generalised cost

HINC Household income

The utility function to be estimated is constructed as

$$U_{nj} = \alpha_{air}d_{i,air} + \alpha_{train}d_{i,train} + \alpha_{bus}d_{i,bus} + \gamma_G GC_{ij} + \gamma_T TTME_{ij} + \gamma_H d_{i,air} HINC_i + \varepsilon_{ij}, \tag{14}$$

where for each j , ε_{ij} has the same independent, type 1 extreme value distribution,

$$F_\varepsilon(\varepsilon_{ij}) = \exp(-\exp(-\varepsilon_{ij})),$$

which has standard deviation $\pi^2/6$. $d_{i,m}$ is the binary variable which indicates if individual i made choice m , $m = \text{air, train, bus, car}$. α_m is an estimate parameter for mode m .

We take the car mode as the basic alternative to construct MNL model and estimate parameters by maximum likelihood estimation method. The parameter values of universal set and restricted set are shown in Table 2, where restricted set means that the set excludes the air mode.

The calculation results show that as Hausman test value $HM = 33.3363$ is greater than $\chi^2_{0.05} = 9.488$, which indicates that IIA assumption of the MNL model is not proper, and there exist heterogeneity and similarity problems. The train, bus and car modes can all be used as the standard basic group, except for the air mode that will result in non-identified parameters,

The tree-like NL model is shown in Fig. 1.

After testing the estimated results of NL model we have $10.945 > \chi^2_{0.05} = 5.99$, which rejects IIA assumption. It is shown that the NL model outperforms the MNL model in terms of interpreting choice behaviours.

Table 1 Discrete choice models' classification

Error term assumption	Independent	Dependent (Similarity)
Identical	MNL	NL
Non-identical (Heterogeneity)	HEV	MMNL, MNP

Table 2 Results for MNL calibration

Estimated parameters	Universal set	Restricted set
α_{air}	5.207	–
α_{train}	3.869	4.464
α_{bus}	3.163	3.105
γ_G	–0.015	–0.0639
γ_T	–0.096	–0.0699
γ_H	0.013	–
Log likelihood at $\gamma = 0$	–291.1218	
Log likelihood at convergence	–199.1284	

It is possible that heterogeneity exists across alternatives. Here, we try to introduce scale parameters into the error terms across alternatives to make error terms unequal and alternatives heterogeneous. At least, one of the alternative scale parameter has to be fixed in HEV model. For the convenience of model comparison, car mode is set to be the basic alternative, and its scale parameter is assumed to be 1. The estimated scale parameter values for other three modes are: $\alpha_{air} = 0.2485$, $\alpha_{train} = 0.2595$, $\alpha_{bus} = 0.6065$. In Bhat’s empirical study [5], HEV model has better interpretation ability over NL model and MNL model. The example below (Table 3) indicates that HEV model does out-perform MNL model on interpretation of choice behaviours but this doesn’t mean it is better than NL model.

The parameter value of car mode is set as zero in the parameter estimation of MNP model, and the results are shown in Table 3. The MNP model does not enhance the interpretation ability of choice behaviours. This is because, some error terms of utility function do not distribute normally.

The MMNL model was built on the basis of the MNL model under the universal set mentioned above. The parameters estimated by maximum simulated likelihood estimation method are listed in Table 4 (classified as independent random parameters and correlated random

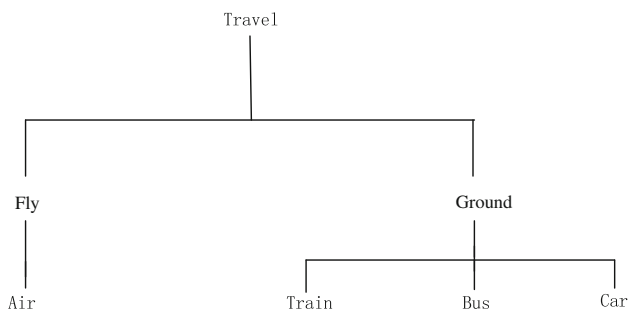


Fig. 1 Nested logit model structure

Table 3 Calibration results for NL, HEV and MNP models

Estimated parameters	NL Model	HEV Model	MNP Model
α_{air}	6.062	7.833	1.358
α_{train}	4.096	6.866	4.298
α_{bus}	5.065	7.172	3.609
γ_G	-0.032	-0.052	-0.035
γ_T	-0.112	-0.197	-0.077
γ_H	0.015	0.040	0.059
Log likelihood at convergence	-193.6561	-195.6605	-196.9244

Table 4 Calibration results for the MMNL model

Estimated parameters	Independent random parameters		Correlated random parameters	
	Estimate	t-stat	Estimate	t-stat
α_{air}	9.41	5.83	10.8	3.8
α_{train}	9.55	5.75	10.7	3.6
α_{bus}	8.59	5.69	9.7	3.7
γ_G	-0.03	-3.35	-4.02	1.9
γ_T	-0.21	-5.96	-13.4	3.9
γ_H	0.059	2.50	5.5	2.0
Chol(GC, GC)			3.00	1.3
Chol(TTME, TTME)	0.13	3.85	3.86	0.4
Chol(GC, TTME)			7.70	2.0
Log likelihood at convergence	-178.810		-176.816	

Chol denotes the parameter value derived by Cholesky decomposition of variance-covariance matrix

parameters). It is shown that the MMNL model has the best performance in interpretation among all the models.

5 Conclusions

- (1) MMNL model can be the first option when the parameter distribution is available because it performs best in interpretation.
- (2) MNP model has its natural defect that all its error terms of utility functions should be normally distributed, which leads to a poor interpretation performance in practical application. Of all the test models, MNP shows the poorest performance in interpretation.
- (3) The prediction accuracy of NL model depends on the given behaviour structure of decision-makers. If the decision-making procedures are unknown, then it will turn out to be very difficult to construct choice structure, and it has great influences on final model results if the decision structure is built with considerable mistakes.
- (4) Illustration analysis indicates that the HEV model has a better interpretation ability in behavioral choice than the MNL model, but worse than NL model.

Acknowledgments This study was supported by the Science & Technology pillar project (No. 0556) of Guangzhou, China in 2011 for the study on railway transportation passenger flow forecast in the context of intercity and city integration based on low carbon.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Fuss MA, McFadden D (1978) *Production economics: a dual approach to theory and applications*. North-Holland Pub. Co, Amsterdam
2. McFadden D, Cowing TG (1984) *Microeconomic modeling and policy analysis: studies in residential energy demand*. Academic Press, Orlando
3. McFadden D (1976) *The theory and practice of disaggregate demand forecasting for various modes of urban transportation*. Dept. of Transportation, Office of the Secretary, Washington
4. Hensher DA, Swait JD, Louviere JJ (2000) *Stated choice methods: analysis and applications*. Cambridge University Press, Cambridge
5. Bhat CR (1998) Accommodating flexible substitution patterns in multi-dimensional choice modeling: formulation and application to travel mode and departure time choice. *Transp Res B* 32(7):455–466. doi:[10.1016/S0191-2615\(98\)00011-3](https://doi.org/10.1016/S0191-2615(98)00011-3)
6. Daganzo CF (1979) *Multinomial probit: the theory and its application to demand forecasting*. Academic Press, New York
7. Hensher D (1986) Sequential and full information maximum likelihood estimation of a nested logit model. *Rev Econ Stat* 68(4):657–667
8. Hensher D (1991) Efficient estimation of hierarchical logit mode choice models. *J Jpn Soc Civil Eng* 425(IV-14):117–128
9. McFadden D (1974) The measurement of urban travel demand. *J Public Econ* 3:303–328
10. Greene WH (2012) *NLOGIT Version 5 reference guide*. Econometric Software, New York
11. Schmidt P, Strauss R (1975) The prediction of occupation using multinomial logit models. *Int Econ Rev* 16:471–486
12. Boskin M (1974) A conditional logit model of occupational choice. *J Polit Econ* 82:389–398
13. Rossi P, Allenby G (2003) Bayesian statistics and marketing. *Mark Sci* 22:304–328
14. Train K (2009) *Discrete choice methods with simulation*, 2nd edn. Cambridge University Press, New York
15. Hensher DA, Rose JM, Greene WH (2006) *Applied choice analysis*. Cambridge University Press, Cambridge
16. McFadden D, Domencich TA (1975) *Urban travel demand: a behavioral analysis: a Charles River Associates research study*. North-Holland Pub. Co., Amsterdam
17. Ben-Akiva ME, Lerman SR (1985) *Discrete choice analysis: theory and application to travel demand*. MIT, Cambridge
18. Bhat CR (2001) Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. *Transport Res B* 35(7):677–693. doi:[10.1016/S0191-2615\(00\)00014-X](https://doi.org/10.1016/S0191-2615(00)00014-X)
19. Hensher DA, Greene WH (2003) The mixed logit model: the state of practice. *Transport* 30(2):133–176