Discrete Choice Model for Public Transport Development in Kuala Lumpur

Abdullah Nurdden[1,*], Riza Atiq O.K. Rahmat[1] and Amiruddin Ismail[1]

[1]*Department of Civil and Structural Engineering, Faculty of Engineering University Kebangsaan, Malaysia 43600 UKM Bangi, Selangor Darul Ehsan, Malaysia,*

[*]*Corresponding author: kamba@vlsi.eng.ukm.my, Tel 0060389216209;*

Abstract

Discrete choice models have been intensively used to analyze and predict the behavior of people in transportation modes. In this paper A binary logit model and multinomial logit model have been developed to study a group of policies aimed at discouraging the use of private transportation and this study sought to identify the factors preventing own transport users from shifting to public transport in order to formulate the policies to achieve this. A survey was carried out on users of private and public (both bus and urban train transport) using the stated preference (SP) and revealed preference (RP) techniques (n =1000). Six variables found likely to encourage the use of public transport were age, income, car ownership, gender, reduced travel time and cost. For the commuter to switch to public transport he would have to be incentivated to do so.

KEY WORDS: Discrete choice model, mode choice, logit model, public transport attitudes

Why Discrete Choice Models was Chosen for this Study

Discrete choice models have played an important role in transportation modeling for the last 25 years. They are used to provide a detailed representation of the complex aspects of the transportation demand, based on strong theoretical justifications. Moreover, several packages and tools are available to help practioners using these models for real applications, making discrete choice models more and more popular.

A discrete choice model predicts a decision made by an individual (choice of mode, choice of route, etc.) as a function of any number of variables, including factors that describe policy change. The model can be used to estimate the total number of people who change their behavior in response to an action. The model can also be used to derive elasticities, i.e., the percent change in private car users to a given change in any particular variable.

A discrete choice model is a mathematical function, which predicts an individual's choice based on the utility or relative attractiveness (bike or drive) (Ben-Akiva and Lerman, 1985; Horowitz *et al.,* 1986). The logit function is a common mathematical form used in discrete choice modeling. Early applications of discrete choice models involved research on consumer choices in which the "costs," "prices," or other characteristics of the choice were the major explanatory variables. A classical example considers alternative modes of transportation, such car, bus, train, which are assumed to differ with regard to transit time and cost (Hensher, 1986). The mode can be applied across a population to estimate the total number of people who change their behavior in response to an action.

Multinomial Logit Model

The logistic probability unit, or the logit model, was first introduced in the context of binary choice where the logistic distribution is used. Its generalization to more than two alternatives is referred to as the multinomial logit model. The multinomial logit model is derived from the assumption that the error terms of the utility functions are independent and identically distributed (or Type I extreme value). That is, ϵin for all i, n is distributed as:

$$F\ (\varepsilon) = \exp\ [-e - \mu\ (\varepsilon - \mu)], \mu > 0 \qquad [1]$$

$$(\varepsilon) = \mu\ e^{-\mu(\varepsilon-\eta)}\ \mathbf{exp\ [-e}^{-\mu(\varepsilon-\eta)}\mathbf{]} \qquad [2]$$

Where η is a location parameter and μ is a strictly positive scale parameter. The mean of this distribution is

$$\eta + \gamma / \mu \qquad [3]$$

Where

$$\gamma = {}_{k\to\infty}\ \mathbf{LIM} \sum_{i=1}^{k} \frac{1}{i}\ \mathbf{- ln(k)} \cong 0.5722 \qquad [4]$$

Is the Euler constant? The variance of the distribution is

$$л2/6\mu 2. \qquad [5]$$

The probability that a given individual n chooses alternative i within the choice set Cn is given by

$$P(i \setminus Cn) = \frac{e\mu vin}{\sum_{j\in cn} e\mu vjn} \qquad [6]$$

An important property of the multinomial logit model is independence from irrelevant alternatives (IIA). This property can be stated as follows: the ration of the probabilities of any two alternatives is independent of the choice set. That is, for any choice sets CI and C2 such that CI ⊆ Cn and C2 ⊆ Cn and for any alternatives i and j in both C1 and C2, we have

$$\frac{P(i \setminus Ci)}{P(j \setminus Ci)} = \frac{P(i \setminus C2)}{P(j \setminus C2)} \qquad [7]$$

An equivalent definition if the IIA property is: the ratio of the choice probabilities of any two alternatives is unaffected by the systematic utilities of any other alternatives. The IIA property of multinomial logit models is a limitation for some practical applications. This limitation is often illustrated by the red bus/blue paradox in the model choice context. We use here instead the following path choice example.

Model Structure

The multinomial model (MNL) is derived through application of utility maximization concepts to a set of alternatives from which one, the alternative with maximum utility, is chosen. The model assumes the utility of an alternative *i* to and individual *q*, *Ui q,* includes a deterministic component, *Vi q,* and an additive random component, *εiq;* that is

$$Uin = Vin + \epsilon in \qquad [8]$$

The deterministic component of the utility function, which is commonly specified as x linear in parameters, includes variables, which represents the attributes of the alternative, the decision context and the characteristics of the traveler or decision maker. The linearity of the utility function can be overcome by prior transformation of variables, quadratic form, spline function (line segment approximations) or estimation with special purpose software.

Assuming that random component, which parameter errors in the model's ability to represent all of the elements which influence the utility of an alternative to an individual, is independently and identically distributed across cases and alternatives leads to the multinomial logit model.

$$\text{Piq} \quad = \quad \frac{biViq}{\sum_{I=1}^{J} e^{biViq}} \qquad \text{[9]}$$

Where:

Piq is the probability that alternative i is chosen by individual q

e is the exponential function

Viq is the deterministic component of the utility of alternative i for individual n

J is the number of the alternatives

bi is the coefficients

This model is called the multinomial logit model and has been widely applied in transportation planning. The utility functions in the above formulation are individually described below

Viq (CAR) = b0+b1 b2*travel time+ b2*travel cost b3*SE [10]

Viq (BUS) = b0 +b1*travel time +b2*travel cost +b3*SE [11]

Viq (train) = b0+b1*travel time +b2*travel cost +b3*SE [12]

Where:

Viq (car) is the utility for car choice

Viq (bus) is the utility for bus choice

Viq (train) is the utility for the train choice

Tiq is travel time

Tc is the travel cost

b1,b2,b3 are coefficients

SE is the socioeconomic characteristic such as age, income, etc

An issue that is related to the desegregation issue is whether travel time (or one or more of its components) should be represented as a generic or mode specific variables. Travel time (or one of its components) is generic if its represented by the same variables in all modes. It is mode specific if it is represented by different variables in different modes. Use of mode specific variable to represent of attribute admits the possibility that traveler's evaluate that attribute differently for different modes. Use of generic variables excludes this possibility. In-vehicle travel time is a travel time components that sometimes is represented by x mode-specific variables in mode choice models. Travel cost also can represent mode specific or generic as x travel time.

Suppose that the model choice between car (c), bus (b), and train (t) for travel work, it is desired that travel time should be mode specific variables and travel cost should be generic the utility function can be specified as follows

VC = b1+b2IVTTc +b3OVTTc + b4C c +b5A +b6SE [13]

V b = b1+b2IVTTb +b3OVTTb + b4Cb +b5A+b6SE [14]

$$V t = b1+b2\ IVTTt + b3OVTTt + b4Cb + b5A + b6SE \quad [15]$$

Where:

IVTT denotes in-vehicle travel time,

OVTT denotes out-of-vehicle travel time,

C denotes travel cost,

A denotes the number of auto mobile owned by the traveler's household,

SE denotes socio-economic characteristics

Probabilities of modal choice are

$$\Pr(car) = \frac{\exp(Vcar)}{\exp(Vcar) + \exp((Vbus) + \exp(Vtrain))} \quad [16]$$

Probability of choosing an individual mode of car

$$\Pr(bus) = \frac{\exp(Vbus)}{\exp(Vbus) + \exp((Vcar) + \exp(Vtrain))} \quad [17]$$

Probability of choosing an individual mode of bus

$$\Pr(train) = \frac{\exp(Vtrain)}{\exp(Vtrain) + \exp((Vbus) + \exp(Vtcar))} \quad [18]$$

Probability of choosing an individual mode of train

Utility Theory

Virtually all-operational models for predicting individuals' choices are based on a behavioral can be stated called "utility maximization." This principle and its relation to choice can be stated in words very simply.

According to the utility maximization principle, there is a mathematical function, called a utility function, whose numerical value depends on attributes of the available options and the individual. The utility function has the property that its value for one option exceeds its value for another if and only if the individual prefers the first option to the second. The utility maximization principle can be stated mathematically as follows. Let C donate the set of options available to an individual {e.g., car, bus, and train in the case of mode choice}. C is called the choice set. For each option *i* in C, let X a*i* denote the attributes of *i* for the individual in question. Foe example, if *i* corresponds to drive car, Xai denotes the travel time, travel cost, and the relevant attributes of the driver car mode for the individual in question. Let S denote the attributes of the individual that car relevant to preferences among the options in C {e.g., income, automobiles owned, etc}. Then, according to the utility maximization principle, there is x function *U* {the utility function} of the attributes of options and individuals that describe individuals' preferences. *U* has the property that for any two options *i* and *j* in C

$$U(Xi,S) > U(Xj, S) \quad [19]$$

Implies that the individual prefers alternative *i* to alternative *j* and will choose *i* if given a choice between *i* and *j* . Given a choice among many options, alternative *i* in **C** is chosen if

$$U(Xi,S) > U(Xj, S) \quad [20]$$

For all alternatives *j* {other than *i* } in **C.**

The utility theory presents strong limitations for practical applications. Indeed, the complexity of human behavior suggests that x choice model should explicitly capture some level of uncertainty; the utility theory fails to do so.

The exact source of uncertainty is an open question. Some models assume that the decision rules are intrinsically. Others consider that the decision rules are deterministic, and motivate the uncertainty from the impossibility of the analyst to observe and capture all dimensions of the problem, due to its high complexity.

Model Results

The mode choice probabilities were categorized by various levels of travel time and travel cost. Mode choice probabilities ranged from 60% likelihood of car use with current train total travel time and current weekly travel costs (70 minutes and RM=25) to 35% likelihood of car use with a reduction in weekly train total travel cost and travel time (10 minute, 10 = 9). At the same time, the probability oftrain ridership increased from 40% with current train total travel time and weekly travel cost of (70 minute, RM 25) to 80% of likelihood with a RM10 and 10 minute reduction in weekly bus total travel cost and travel time. A 50:50 split may be achieved when the travel cost and time are set at RM30 per week and 25 minutes per trip for train travel.

Table 1: Estimation results for multinomial mode choice model (n =1000)

Mode of transport (a)		B	Std. Error	Wald	df	Sig.	Exp (B)
Train	Intercept	8.450	1.074	61.879	1	.000	
	Car ownership	-.607	.193	9.847	1	.002	.545
	Income	.000	.000	3.994	1	.046	1.000
	Travel time	.141	.017	67.567	1	.000	1.152
	Travel cost	-.245	.022	123.037	1	.000	.783
	Age	-.050	.025	4.047	1	.044	.951
	Gender =1	-.883	.436	4.096	1	.043	.413
	Gender= 2	0(b)	.	.	0	.	.
Bus	Intercept	9.840	1.239	63.069	1	.000	
	Car ownership	-.847	.221	14.650	1	.000	.429
	Income	-.001	.000	11.760	1	.001	.999
	Travel time	.227	.019	145.290	1	.000	1.255
	Travel cost	-.437	.035	157.991	1	.000	.646
	Travel cost	-.065	.030	4.763	1	.029	.937
	Gender =1	-1.609	.505	10.154	1	.001	.200
	Gender= 2	0(b)	.	.	0	.	.

a the reference category is: car.

b this parameter is set to zero because it is redundant.

Conclusion

One of the most important uses of mode choice models is the ability to predict the effects of policy measures on consumers. In order to promote greater use of public transport, this study examined the effect on private car use if total bus travel time and travel costs were reduced. This finding was done by solving the Multinomial logit equation for probability using several options of travel time and cost scenarios. The results suggest that travel time and travel cost are characteristics that determine why private car use is a favored modal choice.

References

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

Hensher, D.A. 1986. Valuation of commuter travel time savings: an alternative producer. In I G. Heggie (ed), Model Choice and the Value of time. Oxford: Clarendom Press.

Horowitz, J.L., Koppelman, F.S., and Lerman, S.R. 1986. A self-instructing course in disaggregate mode choice modeling. Technology Sharing Program, US Department of Transportation, Washington, D.C.