A Stated preference Residential Location Choice Model in Indian context

Kumar, Molugaram[1], Krishna Rao, K.V [2]
[1] University College of Engineering, Osmania University, Hyderabad, India
[2] Department of Civil Engineering, IIT Bombay, Mumbai, India

1 Introduction

Travel demand in general depends on land use activity and their spread. Land use means the utilization of land for different activities like residential, commercial, industrial, educational etc., and travel is the link between these activities. The major goal of urban transportation systems is to connect people with various activities. The change in the location of these activities will change travel behavior. There are evidences that land use patterns do have influences on travel patterns, such as trip, mode split, and trip generation. Combining location choice into transportation/land use analysis will make land use patterns endogenous rather than exogenous in the transportation/land use link, and it also opens the door to the land use impacts of transportation, which is still an under-studied area. Decisions relating to location and travel choice have increasingly been modeled by using the discrete choice theory developed based on the concept of utility maximization. Discrete choice models have a long history of application in the economic, transportation, marketing, and geography fields, among other areas. Models of location choice are important tools for analysing urban economic policy, urban housing policy, transportation policy, and urban social spatial structure. Attempts made by many researchers to develop disaggregate location and travel choice models in the context of developed countries have proved the successful application of these models. In the context of developing countries, however, the applicability of the disaggregate choice models for location and travel has not been explored fully.

In the above context, a study has been taken up by the authors to explore the applicability of discrete choice models in arriving at realistic decision framework for the various alternative choices involved in location and travel aspects in the mega cities of the developing countries. It is assumed that each household must choose a home location in a neighbourhood, a workplace location, non-work activity locations along with travel patterns for all household members. These decisions are interrelated, and can be considered together as one joint household (HH) decision.

This paper focuses on; a) particular kind of Revealed preference (RP), Stated preference (SP) data and associated design analysis approach namely the experimental choice approach (pioneered by Hensher and Louviere, 1982) for understanding residential location choice behaviour of travellers; (b) appropriate model structure when using more than one dataset; (c). influence of RP and SP data in joint model estimation d) comparisons of simultaneous and sequential estimation methods in the context of developing countries for residential location choice behaviour.

2 Disaggregate Residential Location Choice: A Review

The integrated analysis of land-use and transportation interactions has gained renewed interest and importance with the passage of the Inter-modal Surface Transportation Efficiency Act (ISTEA) and the Transportation Equity Act for the 21st Century (TEA-21). In this context, one of the most important household decisions is that of residential location, especially because residential land-use occupies about two-thirds of all urban land and home-based trips account for a large proportion of all travel (Harris, 1996). The household residential location decision not only determines the association between the household

and the rest of the urban environment, but also influences the households' budgets for activity travel participation. To be sure, there is a substantial and rich body of literature related to household residential choice. One stream of research on residential location modeling is based on a discrete choice formulation. Sermons and Koppelman (2001) identify at least two appealing characteristics of such a formulation for residential location analysis. First, the discrete choice approach is based on microeconomic random utility theory and models the residential location choice decisions as a tradeoff among various locational attributes such as commute time, housing costs, neighborhood and accessibility to participation in activities. Second, the discrete choice approach allows the sensitivity to locational attributes to vary across socio-demographic segments of the population through the inclusion of interaction variables of locational characteristics with demographic characteristics of households. The early applications of the discrete choice formulation to residential location analysis include the works of McFadden (1978), Lerman (1975), Onaka and Clark (1983), Weisbrod et al. (1980), Quagley (1985) and Gabriel and Rosenthal (1989). More recent applications include Timmermans et al. (1992), Hunt et al. (1994), Waddell (1993, 1996), Abraham and Hunt (1997), Ben-Akiva and Bowman (1998), Sermons (2000), and Sermons and Koppelman (2001). Some of the above studies have focused only on residential location choice (for example, McFadden, 1978; Gabriel and Rosenthal, 1989; Weisbrod et al., 1980; Hunt et al., 1994; and Sermons and Koppelman, 2001), while others have focused on residential choice as one element of a larger mobility-travel decision making framework (for example, Lerman, 1975; Quagley, 1985; Waddell, 1993, 1996; Abraham and Hunt, 1997; and Ben-Akiva and Bowman, 1998). Similarly some studies have focused on location choice for specific demographic groups (such as single worker and Caucasian households), while others have been more inclusive.

The current study may be distinguished from earlier studies in the sense that it is applied in the context of a developing country. Therefore, this study explores the applicability of discrete choice models for residential location choice model in the context of developing countries. The next section discusses the model structure which is used in the current study.

3 Model Structure

Most discrete choice models are based on the random utility maximization (RUM) hypothesis. Within the class of RUM-based models, the multinomial logit (MNL) model has been the most widely used structure for residential location choice. The random components of the utilities of the different alternatives in the MNL model are assumed to be independent and identically distributed (IID) with a type I extreme value (Gumbel) distribution (Johnson and Kotz, 1970). The logit model is a mathematical formulation that represents how individuals make tradeoffs among the attributes of alternatives when choosing one alternative out of a set of available alternatives (McFadden, 1974). The choice situation that is considered in this research is a well-known multinomial logit model (MNL) of the following form:

$$P_n(i) = \frac{\exp(V_{in})}{\sum_j \exp(V_{jn})} \quad (1)$$

where, $P_n(i)$ = probability of household n choosing residential location i.
V_{in} = a deterministic component of utility a function of exogenous variables, and it can be written as

$$V_{in} = \alpha_i + \beta X_{in} \quad (2)$$

Where, α_i = constant specific to the alternative *i*, β = vector of parameters to be estimated, and X_{in} = vector of attributes for the individual *n* and the alternative *i*.

In the case of using RP data only the random error (ε) term is associated with the independendent variables and is assumed to be same for estimation and prediction cases. But in case of using SP data the utility computed is pseudo utility and the random error (η) term associated is a difference between random error terms of RP and SP data, and hence cannot be used as such for prediction purposes. Bradley and Daly (1992) have suggested a scaling approach, which correlates the variance of error term of different observations. The difference between the RP and SP errors can be represented as a function of their variances, such that:

$$\mu^2 = \text{var}(\varepsilon_{iq}) / \text{var}(\eta_{iq}) \quad (3)$$

where μ is the scale factor, scaling the error in SP with respect to the error in RP. Based on the above theoretical framework the utility functions in case of combination of RP and SP data can be written for an alternative 'i' ∈ A (Ben-Akiva and Morikawa, 1990) as

$$U_{iq}^{RP} = \alpha X_{iq}^{RP} + \beta Y_{iq}^{RP} + \varepsilon_{iq} \quad (4)$$

$$\mu U_{qi}^{SP} = \mu(\alpha X_{qi}^{SP} + \gamma Z_{qi}^{SP} + \eta_{qi}) \quad (5)$$

where, α, β and γ are parameters to be estimated; X^{RP} and X^{SP} are vectors of common attributes to both type of data; and Y^{RP} and Z^{SP} are the vectors of attributes specific to RP or SP data, respectively. The stochastic errors ε and η are independently distributed Gumbel with zero mean the choice probabilities are defined on the basis of their utility functions on a logit type structure. The maximization of the joint likelihood function is a non linear problem because μ multiplies some of the parameters to be estimated. To solve this problem two techniques have been used with good results: the simultaneous estimation method developed by Bradly and Daly (1991) and the sequential estimation method proposed by Ben-Akiva and Morikawa (1990).

3.1 The simultaneous estimation method

From equations (4) and (5) it may be seen that the scale factor *μ* is an essential link in the estimation of coefficients *α, β* and *γ*. It is also important to remark that the appearance of the product *μα* is a key element as it transforms the mixed data estimation problem in to a non-linear problem. This may be resolved using ALOGIT (Daly, 1992) with artificial tree structure. This method, developed by Bradley and Daly (1991), consists of constructing an artificial tree which has twice as many alternatives as there are in reality. Half of these are labeled RP alternatives and the other half as SP alternatives. For the RP alternatives, SP choices are unavailable, and therefore, these are estimated as standard logit model. The tree has as many elementary alternatives as there are in RP and SP sets combined. The RP alternatives emerge directly from the root, whereas, the SP alternatives are placed in nest, emerging from the root. The RP alternatives are modeled using the nest structure and SP alternatives are modeled using the tree structure. In case of SP alternatives, each nest comprises of only one alternative. The main utility of the dummy-alternative can be computed as suggested by Daly (1987) and is given by

$$V^{COMP} = \mu \log \sum \exp(V^{SP}) \quad (6)$$

as there is only one alternative in the nest the expected maximum utility (EMU) of the nest becomes equal to the utility of the alternative itself and can be given as

$$V^{SP} = \alpha X^{SP} + \phi Z^{SP} \quad (7)$$

Therefore, the utility of the nest will become

$$V^{SP} = \mu(\alpha X^{SP} + \phi Z^{SP}) \quad (8)$$

which is exactly the same required and presented in the equation (5). The scale factor should take the same value for all the SP alternatives. Also as the individuals are not modeled as choosing from among the RP and SP alternatives simultaneously, the assumption of scale factor not exceeding unity, does not apply. If the scale factor is higher than unity, then it implies that SP data has less noise than the RP data and opposite is true if scale factor is less than unity (Ortuzar and Willumsen, 1994).

3.2. The sequential estimation method

The procedure is as follows (Ben-Akiva and Morikawa, 1990): (a) Estimate the SP model according to utility functions given in equation (9) in order to obtain the estimators of $\mu\theta$ and$\mu\phi$. Then, define a new variable:

$$\hat{V}_i^{RP} = \mu \theta \mathbf{X}_i^{RP} \quad (9)$$

(b) estimate the following RP model with the new variable included, in order to estimate the parameters λ and α:

$$U_i^{RP} = \lambda \hat{\mathbf{V}}_i^{RP} + \alpha \mathbf{Y}_i^{RP} + \varepsilon_i \quad (10)$$

where $\lambda = 1/\mu$.

(c) multiply *X* and *Z* of *SP* data by μ to obtain a modified *SP* data set. Pool the *RP* data and the modified *SP* data and then estimate the two models jointly.

3.3 Measurement of goodness of fit

After the calibration, any model is tested for the accuracy with which the model approximates the observed data using goodness-of-fit measures. In addition, it is to be checked whether the explanatory variables used in the model are statistically significant or not. The significance of the parameter estimates are found using t-distribution. Goodness-of-fit statistics like ρ-square and chi-square are used to test the accuracy with which the model approximates the observed data. These goodness-of-fit measures are discussed below.

*Significance of variables***:** The *t*-statistic is used to examine the significance of the explanatory variables i.e. whether the magnitude of a parameter estimate is different from zero or not using *t*-distribution.

$$t = \theta_{NK} / \sigma_{NK} \quad (11)$$

where θ_N is a normally distributed vector and the *t*-statistic is distributed with *(N- K)* degrees of freedom, such that *N* is the number of data sets and *K* is the number of variables in the model. If one rejects the null hypothesis, one concludes that the variable is statistically significant in explaining the mode choice.

ρ^2 -Square Statistic:

$$\rho_0^2 = 1 - \frac{L(\hat{\theta})}{L(0)} \quad (12)$$

Where *L(0*) relates to log-likelihood estimate with only dummy variables, $L(\hat{\theta})$ relates to final value of log-likelihood. The likelihood ratio index with constants is given as:

$$\rho_c^2 = 1 - \frac{L(\hat{\theta})}{L(c)} \quad (13)$$

where L(c) relates to log-likelihood estimate with only constants. For the same estimation data set the ρ^2 of model will always increase or at least stay the same whenever new variables are added to the utility functions. For this reason, the adjusted likelihood ratio index or modified likelihood ratio index (rho-squared bar) is used.

$$\bar{\rho}^2 = 1 - \frac{L(\hat{\theta}) - K}{L(0)} \quad (14)$$

where, *K* is the number of variables in model. The literature suggests that rho-squared tests provide a theoretical and sound index of goodness-of-fit (Ortuzar and Willumsen, 1994)

4 Data-set and Methodology

The data set for the study was mainly derived from the data of work-place based revealed and stated preference interviews conducted for residential location choice study at various places in Thane City of Mumbai Metropolitan Region (MMR), Maharashtra. The area of Thane city is 128 Sq. Km. Thane is connected with various parts of the country by rail and road. The population of Thane city has increased from 0.8 million in 1991 to 1.6 million in 2001. This means that Thane's population has doubled in a period of one decade. The city has been experiencing phenomenal growth during the last few decades. There are 3481 small-scale industries in Thane area as per Thane Small Scale Industries Association (TSSIA, 1995). Thane city has been delineated by dividing the are into 115 Traffic Analysis Zones (TAZ) and 11 sectors as per Thane Mass Rapid Transit System (MRTS) study (2001). The office based interview (face to face) survey data contained socioeconomic, travel information and stated preference information. The total interviews conducted for 1998 members at different locations in Thane City. This resulted in 1750 valid samples, which were used in the calibration of RP and SP residential location choice models of this study. Nine alternatives (sectors) were considered for RP residential location choice model based on the data available in the sectors. Two sectors were not considered as alternatives as there is no population in these sectors at present.

5 Design of RP Questionnaire

The revealed preference questionnaire was designed by considering various household socio-economic factors, travel factors and residential location facility factors. The various attributes were designed based on the literature and experts suggestions. The neighborhood related aspects i.e., school availability etc., were considered in the *RP* residential location choice study. The provision is made in the questionnaire for the collection of a variety of information such as socioeconomic status, household characteristics and residential location aspects etc,.

The neighborhood index was designed based on facilities available at the location and near by the location. The information on existing facilities was collected during the survey. For each facility type, four options or scale was given against the question (facility). The scale indicates; *1* very bad, *2* bad, *3* good and *4* is excellent, these were decided based

on some criteria depending on probable facility range available in general. The sum of maximum scale of highest facility (excellent) for all questions (12 Different facilities) is *48* and at the same time sum of minimum (if *1* for all facility) rating are *12*. If a respondent resides in a location of excellent environment with all facilities indicated, his/her NBHI is 48/48=1. Similarly, if a respondent resides in a very bad environment of all facility type, his/her NBHI is 12/48=0.25. Therefore, the worst facility (NBHI) is 0.25 and the highest facility of NBHI is 1. The same NBHI scale is reconstructed to 1 to 10 scale which is used at calibration stage. For each observation all the facility scales were added and compared to the maximum rating scale of value 48. After converting all the information in to NBHI, sector wise aggregation was done to get a single NBHI value which represents the NBHI value of a particular sector. These NBHI values were used in the model as a generic variable. The following mathematical formula was used for calculation of NBHI value.

$$\mathbf{nbhi}_{i} = \sum F_{ij} / \sum F_{j\,max} \qquad (15)$$

Where, $nbhi_{i}$ = Neighborhood index of i^{th} respondent.
$Fij = j^{th}$ facility rating or scale of $_i$th respondent
$F_{j\,max}$ = facility rating or choice.

$$\mathbf{NBHI}_{i} = \sum \text{nbhi}_{i} / \sum N. \qquad (16)$$

where, $NBHI_{i}$= Neighborhood index of i^{th} sector.
N = Total number of samples in the sector/zone.

6 Design of Stated Preference Experiment

Most stated preference techniques are characterized by the use of experimental designs to construct hypothetical alternatives presented to respondents. The attributes that were used in this residential location choice *SP* experiment were travel time, travel cost, rental value, family income, accommodation type, public transport availability and neighborhood index. The neighborhood index takes care of residential location facilities. Based on the opinion of the experts and literature (Hunt et al., 1994) it was found that the above attributes would play a major role in taking decisions relating to residential location choice. It has been found that only a relatively small number of attributes should be presented in stated preference experiments (Bates, 1988). The experimental design presented in this study was known as a fractional factorial design. This is because every possible attribute level is not possible to use. Initially the experiment was designed by taking more attributes with more levels which lead to full factorial design resulting in more than 600 options. It was very difficult to reduce the options up to the considerable number even after following the fractional factorial design process.

According to Pearmain and Swanson (1990) even if a particular design allows a lot of attributes to be presented, it is advisable to limit them to 6 or 7 attributes. With these guidelines the numbers of attributes have been reduced to 6 and the levels of attributes have been considered based on the requirement and recommendations from literature. The influence of the transportation system on residential location choice in terms of travel time and travel cost were considered in the design. Accordingly, consideration was limited to a subset of what appeared from the literature review to be some of the most important attributes influencing residential location choice, including the attributes related to transportation system. Finally, the experiment was designed by taking two attributes (travel time/travel cost and family income) at 3 levels, 4 attributes (rental value, accommodation type, public transport availability and neighbourhood index) at 2 levels. These are detailed as follows:

- The attribute travel time was considered at 3 levels (*15 minutes, 30minutes and 60 minutes*) based on general trend of travel observed in Thane city.

- The attribute travel cost was considered at 3 levels (*5 rupees, 10 rupees and 15 rupees*) based on existing fare structure.

- The experiment, however, was simplified by taking the two attributes, travel time and travel cost, together as one attribute due to their dependency.

- The attribute rental value was considered at 2 levels (*5 rupees/sq-ft and 10 rupees sq-ft*) based on some preliminary investigation and available information in the study area.

- The attribute projected income was considered at 3 levels (*1.5 times of existing Household income, 2 times of existing household income, 2.5 times of existing Household income*).

- The attribute accommodation type was considered at 2 levels (*1 Bedroom Hall Kitchen – BHK- and 2 BHK*) based on some investigation in the study area.

- The attribute public transport availability (*PTA*) was considered at 2 levels (*beyond walk and walk*) based on standards.

- The attribute neighborhood index was considered at 2 levels (*level 1 and level 2*); the specification of neighborhood index was shown in Table 1**.**

Table 1. Neighbourhood Index Specification

NBHI: Level 1	NBHI: Level 2
1. **Schools:** Bus is required	1. **Schools:** Within walk
2. **Average Environs** • Air/Noise pollution is less • Average quality of Neighbourhood • Average safe area/location • Average crime rate • Average prestigious location	2. **Good Environs** • Air/Noise pollution is very low • Good quality of Neighbourhood • Very safe area/location • Crime rate is very low • High prestigious location
3. **Average utilities** • Consumable goods available far from residence • Water available limited • Power available limited • Fire/police station is far away from Residence	3. **Good utilities** • Consumable goods available very near • Water available 24 hours • Power available 24 hours • Fire/police station available

In the experimental design, some options were dominating among other options but as per the literature at least one best/worst option should remain in the choice set, so that their logical or illogical positioning by each respondent provides some indication of the reliabilility of the responses. Descriptions of the hypothetical alternatives considered in the stated preference experiments performed for this research were developed by selecting one out of a set of possible values for each option. A full factorial design acquiesce 144 options. The experiment, however, was simplified by taking fractional factorial design

(*FFD*) exercise (Kocur et al., 1982). The fractional factorial design capitulate 72 options. To keep the total number of possible options at a manageable level, only a few pragmatic values were specifies for each attribute. The result was a set of 50 hypothetical options finalized after physical verification. A 10 cm x12.5 cm card was prepared showing the bundle of values for the attributes for each of these options. The basic structure of stated preference experiment used in the present study was shown in Figure **1.** The information on socioeconomic characteristics and travel for work trip was collected during the survey and the required information has been transferred to the *SP* experiment sheet.

7 Administration of RP and SP Experiment

In May 2004, 1998 ranking/ choice experiments were conducted with individuals selected randomly at various work/industrial/business/shopping areas in Thane city of MMR.

Existing Attributes		**Residential location choice Attributes**	
Travel time/travel cost	*Stated*	Travel time/travel cost	*3 levels*
Rental value	*Stated*	Rental value	*2 levels*
Present monthly HH income	*Stated*	Present monthly HH income	*3 levels*
Accommodation type	*Stated*	Accommodation type	*2 levels*
Public transport availability	*Stated*	Public transport availability	*2 levels*
Neighborhood Index	*Stated*	Neighborhood Index	*2 levels*

Ranking				
1	2	3	4	Choice

Figure 1. Structure of Stated Preference Experiment

A Team of about 10 enumerators (M.Tech students of IIT Bombay) were thoroughly trained for a week for administering the experiment on the respondents. A face-to-face work-place based pilot survey was conducted in Thane city, before taking up main survey, in order to arrive at a suitable survey instrument design for this study. The number of people contacted in the pilot survey was 125. The number of people satisfying the laid down criteria were 48 and those who expressed to participate in the *SP* interview were 36. Out of this number, six people were discontinued in half way. The minimum and maximum time consumed for each interview was 10 and 20 minutes respectively.

Based on the experience gained from pilot survey, few modifications were made to different parts of the questionnaire instruments for increasing the efficiency of survey. The main survey was administered with 16 thoroughly trained enumerators at work places, business centers and industrial areas of Thane city.

In majority of cases the interviews were conducted by taking prior appointments from concerned authorities. The enumerators would first explain the objective of the study with the help of leaflet to the respondent and then collect his/her personal and trip information by filling the appropriate forms. The attributes travel time, travel cost, rental value, household income, accommodation type public transport availability and neighborhood index by the existing information are then transferred to the appropriate place in the *SP* questionnaire. Each experiment was a voluntary interview in which the respondent was

approached and asked to choose four hypothetical residential alternatives (cards) and arrange them in order of preference from the best to worst, taking into accounts the needs and wants of the respondent's present household location. A set of four typical hypothetical ranking cards chosen by the respondent was shown in Figure **2**.

Card No. 11

Travel Time /Travel Cost	Rental Value for 500 sqft (1 BHK)	Projected Monthly income (Rs)	Accomm -odation Type	Public Transport Availability	Neighbourhood index (NBHI)
15 Min / Rs.5	*Rs. 2500 (Rs.5/sqft)*	*1.5 Times of Existing HH Income*	*2 BHK*	*Beyond Walk*	*1. School within walk 2. Good environs 3. Good Utilities*

Card No. 25

Travel Time /Travel Cost	Rental Value for 500 sqft (1 BHK)	Projected Monthly income (Rs)	Accomm -odation Type	Public Transport Availability	Neighbourhood index (NBHI)
30Min/ Rs.10	*Rs. 2500 (Rs.5/sqft)*	*2.5 Times of Existing HH Income*	*1 BHK*	*Walk*	*1. Schools: bus is required 2.Average environs 3.Utilities Average*

Card No. 41

Travel Time /Travel Cost	Rental Value for 500 sqft (1 BHK)	Projected Monthly income (Rs)	Accomm -odation Type	Public Transport Availability	Neighbourhood index (NBHI)
60Min/Rs. 15	*Rs. 5000 (Rs.10/sqft)*	*2.5 Times of Existing HH Income*	*2 BHK*	*Walk*	*1. Schools: bus is required 2.Average environs 3.Utilities Average*

Card No. 19

Travel Time /Travel Cost	Rental Value for 500 sqft (1 BHK)	Projected Monthly income (Rs)	Accomm -odation Type	Public Transport Availability	Neighbourhood index (NBHI)
30Min/ Rs.10	*Rs. 2500 (Rs.5/sqft)*	*2 Times of Existing HH Income*	*1 BHK*	*Beyond Walk*	*1. School within walk 2. Good environs 3. Good Utilities*

Figure 2: Typical Ranking Exercise with four cards (numbers shown**)**

In each case these four alternatives were selected randomly from the full set of 50 alternatives in the "deck" of cards to maintain the orthogonality of the variables (Louviere et al., 1981). The total interviews were conducted for 1998 members at different locations in Thane City. This resulted in 1750 valid samples after cleaning of the data. These samples were used for the calibration of *SP* residential location choice models in this study. The number of collected samples was satisfied as per the laid down criteria of 75 samples per segment (Pearmain and Swanson 1990; Bradley and Kroes 1990; and Swanson et al. 1992). Each Income group or level was considered as segment in this

study which was shown in Table 2. The four hypothetical alternatives were considered for *SP* residential location choice model at calibration stage.

Table 2: Household Income Wise Valid *SP* Samples

S.No.	Household income (Rs)	No. of Samples
1	≤ 5000	196
2	5001-10000	528
3	10001-20000	547
4	20001-30000	300
5	30001-40000	131
6	>40000	48
	Total	1750

8 Survey Results and Analysis

The existing information on socio-economic characteristics was collected along with *SP* survey. This data, after checked thoroughly and applying logical checks, were stored in a data base. The residential location facility was indicated in terms of Neighbourhood index (NBHI).The number of samples obtained for each facility indicated against the facility scale was shown in Table 3.

Table 3. Samples obtained for each existing facility type on rating scale

Sl. No	Facility Name	Samples obtained on facility Scale			
		1	2	3	4
1	School Availability	*2*	*75*	*726*	*947*
2	College Availability	*11*	*462*	*981*	*297*
3	Consumer Goods Shop Availability	*2*	*58*	*565*	*1125*
4	Durable Shopping centre availability	*11*	*275*	*661*	*803*
5	Water Availability	*76*	*157*	*661*	*896*
6	Power Availability	*47*	*47*	*516*	*1140*
7	Fire / Police Station Availability	*83*	*425*	*1054*	*188*
8	Air / Noise Pollution	*17*	*322*	*901*	*510*
9	Quality of Neighbourhood	*29*	*291*	*923*	*507*
10	Safety	*20*	*301*	*775*	*654*
11	Crime rate	*1*	*41*	*984*	*724*
12	Prestige of the location	*15*	*335*	*849*	*551*
	Total	*314*	*2788*	*9556*	*8342*

It was observed from the survey that 40% of employees were in the age group between 30 and 40. It was also observed that 76% people have own houses, 60% of people have their income between Rs. 5000 ands Rs. 20000. Relating to the completeness of information in the *SP* survey sheets, 100 percent complete information was obtained in 70% samples, 90 percent information was obtained in 20% samples, 75% of information was obtained in 6% of samples, and 50% of information was obtained in the remaining 4%

samples. Relating to the erroneous entries, 72% samples without any wrong entries, 13% samples with 15 percent wrong entries, 5 % samples with 30 percent wrong entries and the remaining 12% samples with more than 50 percent wrong entries were observed in the samples.

9 Specification and Calibration of RP Model

All the possible variables were used in the R*P* residential location choice model including neighborhood index. The nine sectors were considered as the alternatives for *RP* residential location choice model. Two sectors (10 and 11) were not considered due to negligible development in these sectors at present. A simple *MNL* model (equation 1) was used for proper specification of *RP* residential location choice.

As a starting point all the variables were used along with the mode-specific constants, system variables in defining the utility of different residential location alternatives. The attention was given to the use of different mode-specific and generic variables in utility function for different residential location choices. The variables were eliminated if found non-significant or found to have unreasonable indications. The variables so eliminated were then used in the utility function of other alternatives and again the same checks were made. On reaching the most optimal model specification, the absence of non-significant variables did not alter the value of coefficient estimates of remaining variables. With these variables, several specifications were tried before selecting the final model. All the generic variables were showing proper sign and were found to be significant in the final model. The estimated co-efficients and t-values of final model are shown in Table 4. The travel time variable has negative sign and is found to be significant at 95% confidence level based on its t-value. The attribute rental value was worked out based on present rental values in the sectors. The income is a mode-specific variable obtained from the survey. The rental value/HH income was showing logical sign and the 't' statistic is satisfied at 90% confidence level. The variable NBHI, constructed as explained earlier (preceding section), was used as generic variable. This variable has proper sign and is significant at 95% confidence level.

Table 4. Co-efficient Estimates and Statistics of Revealed Preference model

Variable	Co-efficient	t- statistic	remarks
TT	-0.08509	-16.1	Generic
RV/INC	-0.01063	-1.5*	Generic
NBHI	0.2086	2.8	Generic
Structural parameters			
ρ^2	0.1065	-	-
L(0)	- 3836.35	-	-
L(θ)	- 3427.86	-	-
No. of observations	1750	-	-

*significant at 90% confidence level.

10 Specification and Calibration of SP model

All the variables used in the *SP* residential location choice were hypothetical, based on experimental design. Four hypothetical alternatives were considered for *SP* residential location choice model. 1750 valid samples were considered for calibration of choice models. Simple *MNL* model (equation 1) was used to estimate parameters by proper

specification of *SP* residential location choice.

The same procedure (explained in proceeding section) was used for calibration of SP residential location choice model. The estimated co-efficients and t-values of final model were shown in Table 5. The travel time variable was showing negative sign and was found significant at 95% confidence level. The rental value/HH income was showing logical sign and ‘was found significant at 95% confidence level. The NBHI was showing proper sign and ‘based on its t-statistic was found to be significant at 95% confidence level. The variable accommodation type (ACTP) was showing logical sign and significant at 95% confidence level. The variable public transportation availability (PTAV) was showing proper sign and ‘its t-value was satisfied at 95% confidence level. Most of the chosen SP variables were performing well in the calibration of SP model. The structural parameter ρ^2 value was found to be 0.3827, which indicates good fit compared to RP model ρ^2 value. Based on these values (RP and SP), it is very difficult to judge the reliability of data type. Therefore it was decided to mix the RP and SP data by joint modelling to mitigate the weakness of RP and SP data type. The joint RP and SP models will be discussed in the following section.

Table 5. Calibration results of Stated Preference model

Variable	Co-efficient	t- statistic	remarks
TT	-0.0407	16.5	Generic variable
RV/INC	-0.0517	19.2	Generic variable
NBHI	0.5812	29.2	Generic variable
ACTP	0.1966	2.9	Generic variable
PTAV	-0.8475	11.5	Generic variable
Structural parameters			
ρ^2	0.3827	-	-
L(0)	-2426.015	-	-
L(θ)	-1497.624	-	-
No. of observations	1750	-	-

11 Joint Estimation of RP and SP Data

11.1 Data structure

Many studies proved that a fundamental element in the construction of good model is to have good quality data. Two types of data sets were used in the mixed estimation. The first one is *RP* data and the second one is *SP* data. The *RP* data (1750 observations) and SP data (1750) were obtained from RP&SP residential location choice survey which was explained in section 5and 6 of preceding pages.

11.2 Artificial Structure for Mixed *RP*/*SP* Data Estimation

As discussed in previous section the number of samples used in the joint estimation was 3500. The joint model was developed using *RP* and *SP* data by simultaneous and sequential estimation method. The following sections will discuss the calibration of joint model with simultaneous and sequential estimations.

11.3 Calibration with Simultaneous Estimation

The artificial tree structure (Bradley and Daly, 1997) is conditional on the data and alternatives available. The method consists of constructing an artificial tree which has twice

as many alternatives as there are in reality. In this study 9 and 4 residential location alternatives were considered in the development of *RP* and *SP* joint models respectively. Therefore, nine RP residential location choice alternatives i.e., sector 1, sector 2, …..sector 9 were considered. Similarly four alternatives are taken for *SP* residential location choice. Half of these are labelled *RP* alternatives, and another half is *SP* alternatives. The utility functions are U^{RP} and U^{SP} as per the equations (4 and 5). Figure 3 shows the artificial nested-tree structure comprising of *RP* and *SP* alternatives. The *RP* alternatives are placed just below the root of the tree; however the *SP* alternatives are each placed in single alternative nest. In this case for an *RP* observation, the *SP* alternatives set unavailable and the choice was modeled as standard logit model; for an *SP* observation, the *RP* alternatives are set unavailable and the choice was modeled by a nested (tree) logit structure. ALOGIT (1992) was used for the estimation of parameters. The variables are entered in the utility function of different alternatives as generic fashion. The variables are examined before arriving at the final list of significant variables. The significance of variables is examined at 95 percent confidence level.

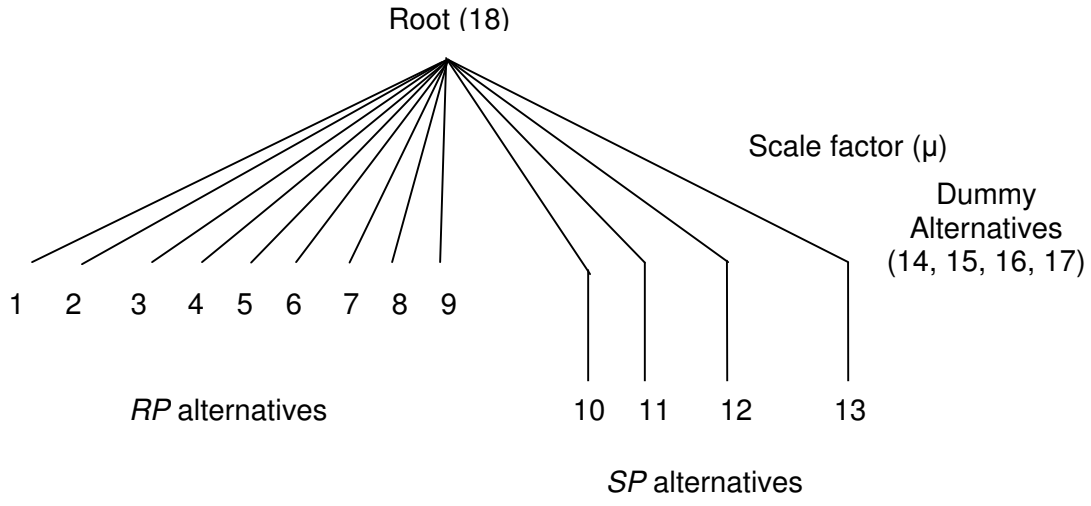

Figure 3. Artificial tree structure for mixed RP and SP data

The statistical measures like rho- square value and likelihood ratio tests were used for examining the goodness of fit-of-the model. The model with the selected variables was found to be significantly analyzing the behaviour of the households.

11.4 Calibration with Sequential Estimation

This method, developed by Ben-Akiva and Morikawa (1990), has the advantage of allowing the use of ordinary logit or probit estimation software. In this method the *SP* model is estimated as regular procedure and find out the common variables from *RP* and *SP* model. The common variables of *SP* coefficients are to be multiplied with *RP* common variables. Now the new common variables are treated as generic or common variables to all the *RP* alternatives. We estimate the *RP* model considering the above common variables to get the scale factor μ. The scale factor value is to be multiplied with *SP* data, then the *RP* data is combined to estimate the joint model of *RP*/*SP*. The different statistical measures like rho-square value and likelihood–ratio tests are used for examining the goodness of fit-of-the model. In this method the model developed using *RP* and *SP* data together is observed to be

having good statistical fit as per the rho-square value and also matching with the values estimated by simultaneous method.

11.5 Calibration Results of *RP*&*SP* Joint Estimation

Table 4 & 5 shows the individual models of revealed preference (*RP)* and stated preference (*SP*). All these models are characterized for having correct signs and good statistical significance. One variable in *RP* model is not satisfactory at 95% confidence level. The same RP and SP data were used for joint estimation by simultaneous and sequential methods as discussed in preceding sections.

Table 6 shows the parameters obtained for mixed models using both the approaches (simultaneous estimation and sequential estimation methods). It was observed that all parameters have the correct sign and that most of them are statistically significant (t-statics are shown in brackets) at the 95% confidence level. It was also observed that the estimated scale co-efficient μ is lower than one and highly significant, confirming the hypothesis that the SP data has slightly more noise than the RP data. It is also found that the *μ* value (0.702 & 0.773) is close to *1*, which is indicating that both the *RP* and *SP* data sets have approximately the same noise.

Table 6. Parameters of mixed models estimated by simultaneous and sequential methods

Attribute	Simultaneous Method	Sequential Method	Remarks
TT	-0.0613 (15.9)	-0.0575 (21.1)	generic
RV/INC	-0.0621 (14.8)	-0.0488 (20.3)	generic
NBHI	0.8787 (20.3)	0.8082 (35.9)	generic
ACTP	0.2195 (2.3)	0.2341 (2.7)	generic
PTAV	-1.132 (10.4)	-1.08 (11.8)	generic
Structural Parameters			
ρ^2	0.2178	0.2196	-
μ	0.702 (16.9)	0.773 (14.2)	-
L(0)	-6764.66	-6764.65	-
L(c)	-3453.73	-3453.73	-
L(θ)	-5291.64	-5279.05	-
No. of observations	3496	3496	-

There is a great similarity in the values of each corresponding parameter in Table **6.8**. This conforms empirically that both mixed estimation approaches produce consistent estimates. However, the sequential method yields parameters with higher t-statistics. The general goodness of fit is also higher in the sequential approach. The scale factor *μ* is observed higher in sequential approach. The same observation was matching with the earlier studies done by Gaudry et al. (1989).

12 Conclusion

The discrete choice modeling offers considerable advantages like economy of data collection, transferability, policy sensitivity and flexibility over conventional approaches. Various attributes of household characteristics and system characteristics have been shown to have a statistically significant influence on residential location preferences in Thane city of MMR. Several transportation-related attributes have an effect on the attractiveness of residential locations in Thane City of MMR. The NBHI in particular has been shown to have an impact on residential location choice.

Some respondents have correlated cost with the neighbourhood quality and therefore selected more expensive alternatives more readily even though respondents were told to assume that all unmentioned attributes were the same across all alternatives. The models of residential location choice bahaviour resulting from this work can be used to asses the impacts of changes to the transportation system in Thane city of MMR. The stated preference techniques that were used were found to be very useful in many ways. A useful data set with good statistical properties was obtained easily and quickly with very little cost. There is still some concern that all the attributes presented to the respondents proved to have a significant influence simply because values for these factors were specified. The following specific conclusions were made from this study.

- Travel time, rental value, household income, and neighbourhood Index were the most influencing attributes in residential location choice modelling.
- The RP and SP residential location choice model developed and calibrated in this study exhibits good results in terms of goodness-of-fit measures.
- The neighbourhood index is the representative of locational facilities, which is playing an important role in residential location choice model**.**
- In mixed estimation with RP and SP data, the calibration results were improved than RP and SP individual models. The scale factor was observed to be less than 1 but close to one indicating that RP and SP data sets are equally significant.

The developed models can be used for residential allocation in any land use mode.

References

Abraham, J E and Hunt, J D (1997) Specification and estimation of a nested logit model of home, workplace and commuter mode choice by multiple worker households. *Transportation Research Record* 1606, 17–24.

Bates, J (1988) Economic issues in stated preference analysis. *Journal of Transportation Economics and policy* 22 (1), 59-69.

Ben-Akiva, M and Bowman, J L (1998) Integration of an activity-based model system and a residential location model. *Urban Studies* 35 (7), 1131–1153.

Ben-Akiva, M and Morikawa, T (1990) Estimation of switching models from revealed preferences and stated intensions. *Transportation Research A*, 24(6), 485-495.

Bradley, M A and Daly, A J (1991) Estimation of logit choice models using mixed stated preference and revealed preference information. Preprints 6th International Conference on Travel Behaviour, Quebec, Canada.

Bradley, M A and Daly, A J (1992) Use of logit scaling approach in stated preference analysis. In proceedings of *World Conference on Transport Research*, Lyon, 811-823.

Bradley, M A and Daly, A. J (1997) Estimation of logit choice models using mixed stated-preference and revealed-preference information. Understanding Travel Behaviour in an Era of Change, eds. P. Stopher and M. Lee-Gosselin, *Elsevier Science*, Oxford, 209-231.

Bradley, M A and Kroes, E (1990) Forecasting issues in stated preference survey research. 69th TRB Annual Meeting, Washington DC, January 1990, USA.

Daly, A J (1987) Estimating tree logit models. *Transportation Research B*, 21(4), 251-267.

(ALOGIT) Daly, A J (1992) ALOGIT 3.2 User's Guide. *Hague Consulting Group*, The Hague, Netherlands.

Gabriel, S A and Rosenthal, S S (1989) Household location and race: estimates of a multinomial logit model. *The Review of Economics and Statistics* 17 (2), 240–249.

Harris, B (1996) Land use models in transportation planning: a review of past developments and current practice. available *www.bts.gov/other/MFD_tmip/papers/landuse/compendium/dvrpc_appb.htm>.*

Hensher, D A and Louviere, J J (1982) Design and analysis of simulated choice or allocation experiments in travel choice modelling. *Transportation Research Record*, 890, 11-17.

Hunt. J D McMillan, J D P and Abraham, J E (1994) Stated preference investigation of influences on attractiveness of residential locations. *Transportation Research Record* 1466, 79–87.

Johnson, N and Kotz, S (1970) Distribution in Statistics: Continuous Univariate Distributions. Chapter 21, *John Wiley:* New York.

Kocur, G. T., Adler, T., Hyman, W. and Aunet, B. (1982). "Guide to forecasting travel demand with direct assessment". Report No. UMTA-NH-11-001-82, *Urban Mass transportation*, US Department of Transportation, Washington, D.C.

Lerman, S R (1975) A disaggregate behavioural model of urban mobility decisions, *Ph.D. dissertation*, Massachusetts Institute of Technology.

Louviere, J J Henly, D H Woodworth, G Meyer, R J Levin, I P, Stoner, J W Cury, D and Anderson, D A (1981) Laboratory-simulation versus revealed preference methods for estimating travel demand models. *Transportation Research Record* 794, TRB, Washington D.C., 42-51.

McFadden, D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics**.** *Academic Press*, New York, 105-142.

McFadden, D (1978) Modelling the choice of residential location. In A. Karlquist, L Lundquist, F Snickars and J W Weibull (eds), Spatial Interaction Theory and Planning Models, North-Holland: Amsterdam.

Onaka, J and Clark, W A V (1983) A disaggregate model of residential mobility and housing choice. *Geographical Analysis* 19, 287–304.

Ortuzar, J D and Willumsen, L G (1994) Modelling Transport. John Wiley and Sons, New York.

Pearmain, D and Swanson, J (1990) The use of stated preference techniques in the quantitative analysis of travel behaviour. Proceedings of the IMA Conference on Mathematics in Transport, University of Cardiff, September 1990, Wales.

Quagley, J M (1985) Consumer choice of dwelling, neighbourhood and public services. *Regional Science and Urban Economics* 15, 41–63.

Sermons, M W (2000) Influence of race on household residential utility. *Geographical Analysis* 32 (3), 225–246.

Sermons, M W and Koppelman, F S (2001) Representing the differences between female and male commute behavior in residential location choice models. *Journal of Transport Geography* 9, 101–110.

Swanson, J Pearmain, D and Loughead, K (1992) Stated preference sample sizes. Proceedings 20th PTRC Summer Annual Meeting, University of Manchester Institute of Science and Technology, London.

Timmermans, H Borgers, A Dijk, J and Oppewal, H (1992) Residential choice behaviour of dual earner households: a decomposistional joint choice model. *Environment and planning* 24 A, 517-533.

Waddell, P (1993) Exogenous workplace choice in residential location models: is the assumption valid?. *Geographical Analysis* (25), 65–82.

Waddell, P (1996) Accessibility and residential location: the interaction of workplace, residential mobility, tenure, and location choices. Presented at the *Lincoln Land Institute TRED Conferenc*e. Available at: <http://www.odot.- state.or.us/tddtpan/modeling.html>.

Weisbrod, G Lerman, S and Ben-Akiva, M (1980) Tradeoffs in residential location decisions: Transportation vs. other factors. *Transportation Policy and Decision Making* (1), 13-26.