MICRO-SIMULATION FRAMEWORK FOR RESIDENTIAL LOCATION CHOICE IN INDIAN CONTEXT

Kumar Molugram
Associate Professor
Department of Civil Engineering, University College of Engineering,
Osmania University, Hyderabad, India

ABSTRACT

Residential location choice is an important step in the overall land use transportation planning of cites all over the World. The factors socio-economic characteristics, transportation system characteristics and characteristics of the neighborhood play a major role in the decision making process related to residential location choice. Rapid economic development in India and other developing countries, the income levels and living standards of people are improving on one side and the need for better transportation requirement is increasing on the other side. In view of the above, a study has been done to investigate the influencing factors on residential location choice in Thane City of Mumbai Metropolitan Region, India. The residential location choice model explains the decisions of households regarding residential location choice in the light of travel decisions. The calibration of a discrete choice residential location choice model is an important component of any integrated land use transport framework. In this study a revealed preference (*RP*) questionnaire design and stated preference (*SP*) experiment design for the development of a residential location choice model are presented. The *RP* and *SP* survey administration, data analysis and the methodology for development of the model are presented next. The structure, specification and calibration of the residential location choice models using revealed preference (*RP*) and stated preference (*SP*) data are presented. These models have been applied to the base year context of Thane city utilizing a micro-simulation framework. The micro-simulation model demonstrated here is like a building block, and additional models can be easily integrated into the model making complete land use transport micro-simulation model. This micro-simulation framework consists of two sub-models. The first model which is a socio-economic characterization module, assigns socio-economic characteristics to the basic/service employee using Monte Carlo technique based on the observed distributions of each of the characteristics. The second module allocates the residential sector to the basic/service employee. There after, the iterative process of Lowry land use model is followed to get the final distribution of population and employment in the study area. After performing several iterations of the process of Lowry land use model, the total population distribution and the service employment distribution were obtained for the study area. The observed and predicted population and service employment for base year were matching reasonably well. The predictive ability of all the four residential location choice models, viz., RP model, SP model, joint RP-SP model by sequential method, joint RP-SP model by simultaneous method were compared in this simulation framework. Three residential location choice models were used for finding the distribution of population and employment. It is observed that the RP model, SP model and RP+SP model predicted the distribution based on their model strengths and goodness of fit. The joint model was predicting almost very close to the observed distribution of base

Keywords: residential location choice, stated preference, Revealed preference, Micro-simulation

1. INTRODUCTION

The term "Micro-simulation" has appeared with increasing frequency in recent years and its implied definition varies with context. The preferred, broad definition, which has been adopted in the current research work is, "a modelling approach based on the representation of individual decisions, with all the possible outcomes linked to the characteristics of the relevant decision making unit (Orcutt et al. (1961); Aldana et al. (1973); Wilson and Powell (1974); Kreibich (1979); Goulias and Kitamura (1992); Raju et al. (1998); Subbarao and Krishna Rao (2001); Wegener and Spiekermann (2001); Hunt et al. (2001); Ton and Hensher (2001); Sugik and Miyamoto (2004); Ettema et al. (2005)). In order to use the knowledge about decision making units, all the processes must be integrated into a system or model. The model obtained through integrating a number of hypothesis, theories, or postulates into a unified one, provides an instrument to get the joint consequences of these hypotheses, theories or postulates. From a policy analysis point of view, models should provide the policy maker with the predictions about the combined effect of many factors acting on may interacting units. For this purpose, a single hypothesis or theory is rarely sufficient. It is possible only by linking several relevant theories together in a meaningful way.

Besides the direct usefulness as predictive instruments, travel demand models should have explanatory value. Such explanatory role is possible by linking knowledge of behaviour (decisions) of individual decision making units in terms of different decision processes. Proper linking and aggregation of the outcomes from these decisions result in the overall system behaviour. A model output shows a better performance if the model is developed in terms of bahaviour of the fundamental units incorporating the interdependency of various components. Many sub-models like residential location choice using stated preference and revealed preference developed earlier as part of the research. These sub-models, when tested, have shown reasonably good level of performance. These tests were done in an environment without considering the interdependency among them. Therefore, though the sub-models satisfied the validation tests on sample data, the success of the model depends on the proper linking of these sub-models into an integrated model. Integration of those sub-models using micro-simulation has been attempted in this work. This paper is discusses the micro-simulation part only using sub-models developed earlier (Kumar, M and K.V. Krishnarao, 2005, 2006).

2. RESIDENTIAL LOCATION CHOICE MODEL

The residential location choice model explains the decisions of households regarding residential location choice in the light of travel decisions. The calibration of a discrete choice residential location choice model is an important component of any integrated land use transport framework. The revealed preference (*RP*) questionnaire design and stated preference (*SP*) experiment design for the development of a residential location choice models were studied earlier (Kumar, M and K.V. Krishnarao, 2006). The calibrated parameters of the residential location choice model using revealed preference (*RP*), stated preference (*SP*) and joint modeling were taken for the present micro-simulation model.

2.1 Study Area

The Mumbai Metropolitan Region (MMR) of Maharashtra State, India has been selected as study region for examining the regional level model starting from a part of the region. Thane city of MMR had been chosen as the study area (part of MMR) for development of Revealed Preference (RP) and Stated Preference (SP) residential location choice model. The Mumbai Metropolitan Region (MMR) covers an area of about 4355 square kilometers. Mumbai, the capital of Maharashtra state and commercial capital of India, is a major constituent of MMR. The Greater Mumbai, a major part of MMR as per population, is a well-developed city covering an area of 468 square kilometers. The rest of MMR, which is in development stage, consists of five municipal corporations, fifteen municipal councils and a number of villages. With rapid urbanization and industrilisation during last three decades in MMR, the population has increased from 9.9 million in 1981 to 18.5 million in 2001.

Revealed preference and stated preference data were collected at various work places in Thane City of Mumbai Metropolitan Region (MMR), Maharashtra. Specifically for the development of residential location choice model. The Thane city contains 115 Traffic Analysis Zones according to MRTS Thane city study. In the work place based interview (face to face) survey information on socio-economic and travel characteristics were collected. Information on 1998 members at different work place locations in Thane City was collected. This resulted in 1750 valid samples, which is used for the calibration of *RP* and *SP* residential location choice models. The travel time, travel cost and travel distance skims for public transport were obtained from the network information of Thane for the development of *RP* residential location choice model. Where as for *SP* residential location choice all these variables are hypothetical. The choice set considered for the *RP* residential location choice model contained 11 geographical residential locations (sectors) in the study area numbered (1 to 11).

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

Table 1 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. The same RP and SP data were used for joint estimation by simultaneous and sequential methods.

Table 2 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-statistics are shown in brackets) at the 95% level. It was also observed that the estimated scale co-efficient μ is lower than one and is 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 same noise.

Table 1. Parameters of the pure SP and RP models

Variable	**SP Model**	**RP Model**	**remarks**
TT[1]	-0.0407	-0.0851	Generic variable
	(-16.5)	(-16.1)	
RV/INC[1]	-0.0517	-0.0106	Generic variable
	(-19.2)	(-1.5*)	
NBHI[1]	0.5812	0.2086	Generic variable
	(29.2)	(2.8)	
ACTP[1]	0.1966	-	Generic variable
	(2.9)		
PTAV[1]	-0.8475	-	Generic variable
	(-11.5)		
Structural parameters			
ρ^2	0.3827	0.1065	-
L(0)	-2426.015	-3836.35	-
L(θ)	-1497.624	-3427.86	-
Sample size	1748	1748	-

* t-test satisfied at 90% confidence level; TT –travel time; RV/INC – Rental value /HH Income; NBHI-Neighborhood index ; ACTP –Residential Accommodation type; PTAV – Public transport availability

There is a great similarity in the parameter estimates obtained by the two methods. 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 to be higher in sequential approach. These observations are in line with the earlier studies done by Gaudry et al. (1989).

Table 2. Parameters of mixed models 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	-

3. STRUCTURE OF MICRO-SIMULATION MODEL

The residential choice models developed in the earlier study (Kumar. M and Krishna Rao, K. V, 2006) explained above has been applied to the base year context of Thane city utilizing a micro-simulation framework. The methodology of allocating residential zones to basic employees, getting the associated population distribution and service employment, allocation of service employment and getting the distribution of additional population, and the associated iterative process employed is similar to the Lowry land use model philosophy. The micro-simulation model demonstrated here is like a building block, and additional models can be easily integrated into the model making complete land use transport micro-simulation model. This micro-simulation framework consists of two sub-models. The first model which is a socio-economic characterization module, assigns socio-economic characteristics to the basic/service employee using Monte Carlo technique based on the observed distributions of each of the characteristics. The second module allocates the residential sector to the basic/service employee. There after, the iterative process of Lowry land use model is followed to get the final distribution of population and employment in the study area. The following sections describe the details of this micro-simulation model.

4. SOCIO-ECONOMIC CHARACTERISATION

The basic idea in the present model is that the travel demand and population distribution are the cumulative effects of the decisions of all the decision making units in the study area. Therefore, it is necessary to have the complete information of all the decision making units on their socio-economic attributes which influence the choice process. Collecting/procuring such enormous data is practically not possible. However, this problem can be solved if a mechanism is evolved through which the required socio-economic attributes are simulated internally within the model without deviating much from the reality. To achieve this, a sub-model has been formulated to generate the characteristics of all the individual actors/decision making units based on the information available (existing revealed information) for a sample of the population. The distributions of the socio-economic characteristics as observed in the sample as well as the pattern of their choices in terms of residential location based on work place are assumed to be applicable to the study area. Therefore, it is possible to assign the attributes to an individual actor who is of the concern, based on the distributions observed from sample data. One such approach which enables this task is the Monte-Carlo simulation process.

Monte-Carlo simulation is carried out by obtaining the probabilities (p) and the cumulative probabilities (P) of the outcomes from the sample data. In this methodology, a random number (RN) drawn from uniform distribution between 0 and 1 is compared with the cumulative probabilities. If the random number value lies in between the cumulative probability values P_m and P_{m+1}(i.e, if $P_m < RN < P_{m+1}$), then the result is agreed to be (m+1) th outcome. If this process is continued for sufficiently large number of times, the distribution of the outcomes obtained from the simulation will represent the distribution observed in the sample. This process of generation of socio-economic characteristics using random numbers was explained as a flow chart shown in Figure 1.

In order to ensure that whether the Monte Carlo simulation is reproducing the observed distributions, the socio-economic characters of the individuals in the sample were generated as per the procedure described in Figure 1. The simulated and observed values of cumulative frequencies of each attribute of socio-economic characteristic were shown in Figure 2. It was observed from Figure 2 that the simulated values of various attributes were closely matching with the observed values.

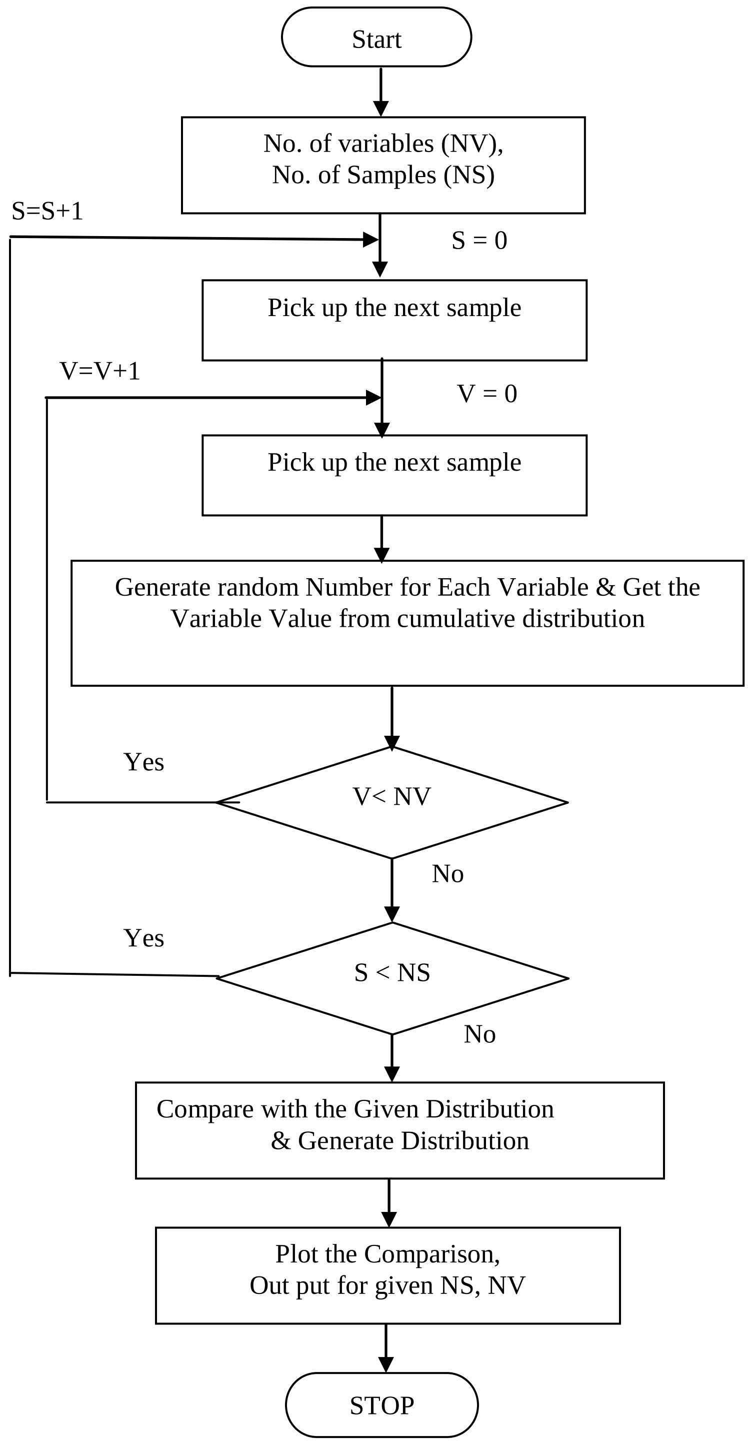

Fig. 1 Random number generation process for socio-economic characterization

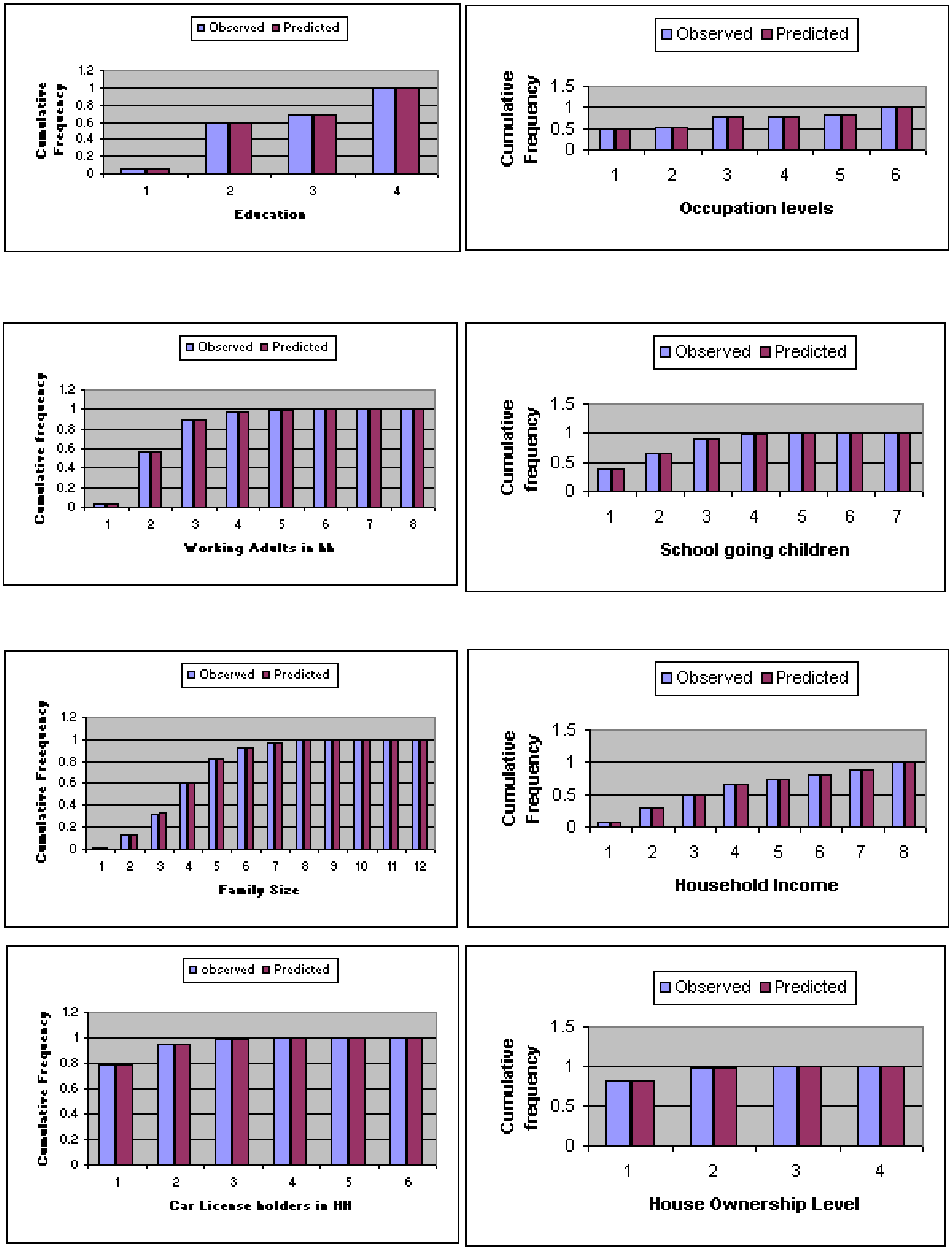

Fig. 2. Simulated and observed values of socio-economic variables

5. RESIDENTIAL LOCATION MICRO-SIMULATION MODEL

The integrated model of micro-simulation for residential location is supposed to provide the distribution of population and employment in the study area. Unlike other usual micro-simulation models, where only a sample is simulated, the present study aimed to simulate the total population of the study area. The socio-economic characteristics, choice pattern and variations in the choices as observed in the sample were logically used in a structured fashion to simulate the behaviour of the entire population. As the model uses the Monte-Carlo simulation, the model run produces the household socio-economic characteristics of population in the study area.

The process starts with generating socio-economic characteristics of all basic employees of the study area using socio-economic characterization model. The base year data on population, basic employment and service employment are provided in table 3 (as an observed data). The residential location choice model is applied and the probability of each of the basic employee choosing the residential sectors (9 sectors) were found out. Based on these probabilities the basic employees were assigned to the appropriate residential sector. The population in each sector is obtained by assigning the generated household size to each of the basic employees. After getting the distribution of population because of basic employment the requirement of service employment in each sector is worked out by using appropriate population serving ratios. The service employment thus worked out is allocated to the service zones in aggregate fashion using proportion of home to shop matrix. This home to shop matrix has been obtained from home interview survey data. Now, these service employees are treated as basic employees. The procedure of assigning socio-economic characteristics to these employees, allocating residential sectors obtaining the additional population distribution and service employment is now same as the one followed for the basic employees. This process is carried out iteratively until the incrementing population is negligible (comes down to single digit). The above procedure is depicted in the form of flow chart in Figure 3. Therefore, the same above process was repeated until the service employment was single number.

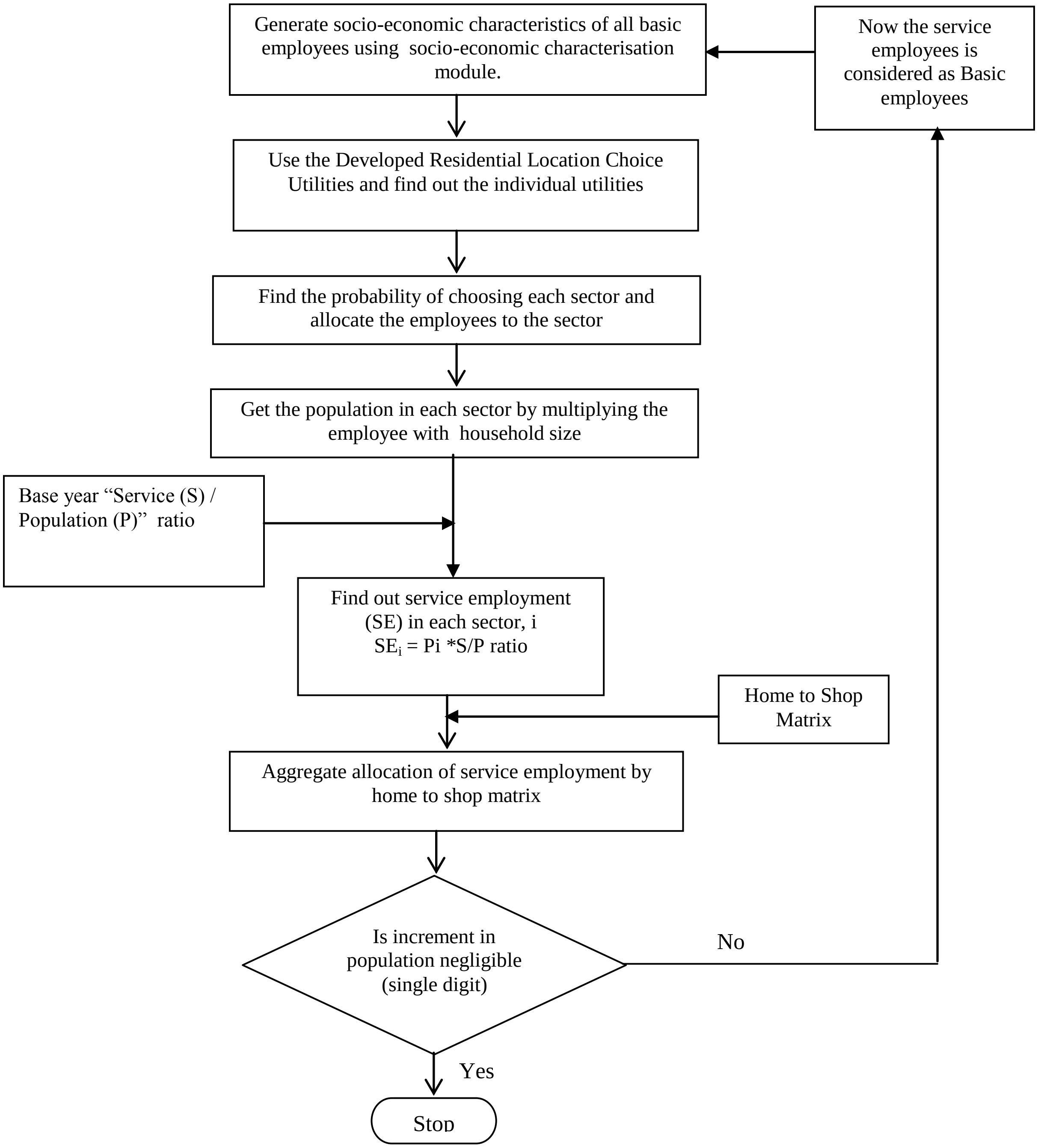

Fig. 3. Structure of Micro-simulation Model for residential location choice

6. RESULTS OF BASE YEAR MICRO-SIMULATION

After performing several iterations of the process explained in the preceding section, the total population distribution and the service employment distribution were obtained for the study area. The observed and predicted population and service employment for base year were matching reasonably well which are shown in Table 3 and Table 4 respectively. The predictive ability of all the four residential location choice models, viz., RP model, SP model, joint RP-SP model by sequential method, joint RP-SP model by simultaneous method were compared in this simulation framework.

Table 3. Population prediction for Base year by Different Models

Sector No.	Observed Values	Simulated Values			
		RP	**SP**	**RP+SP(si)**	**RP+SP(se)**
I	119754	172154.9	131983.1	138420.1	154975.9
II	360721	201990.6	263580.8	307819.8	298283.2
III	410462	195885.7	319931.2	343686.6	299628
IV	99753	158079.6	124545.7	104902.6	113213.3
V	112125	158686.4	130027.4	108834.5	115902.9
VI	28937	133241.8	66579.9	47473.9	57342.6
VII	26960	138612.8	71315.4	55399.2	67018.6
VIII	173998	186705.8	187037.9	208800.7	213269.7
IX	188049	175566.1	197177.2	184237.6	179953.1
Total	**1520757**	**1520924**	**1492178.6**	**1499575**	**1499587.3**

Note: Si – Simultaneous Method, Se – Sequential Method.

Table 4. Service employment prediction for Base year by Different Models

Sector No.	Observed Values	Simulated Values			
		RP	**SP**	**RP+SP(si)**	**RP+SP(se)**
I	2735	4644	3192	2865	3161
II	74722	67404	67145	70122	70102
III	16385	17404	19863	21763	21277
IV	5906	6314	5213	4562	4834
V	12086	13551	12434	11707	11849
VI	7168	7312	4176	3362	3825
VII	1350	1433	1627	1527	1486
VIII	7149	6688	5803	5628	5779
IX	18667	18137	16114	15135	15357
Total	**146168**	**142887**	**135567**	**136671**	**137670**

Note: Si – Simultaneous Method, Se – Sequential Method.

It can be observed from Figure 4 & Figure 5 that the prediction capability of RP model is poor, SP model is moderate and the joint model is good. The structural strength of any model influences the prediction capability. In this study the combined model was predicting much better than the models developed with individual data.

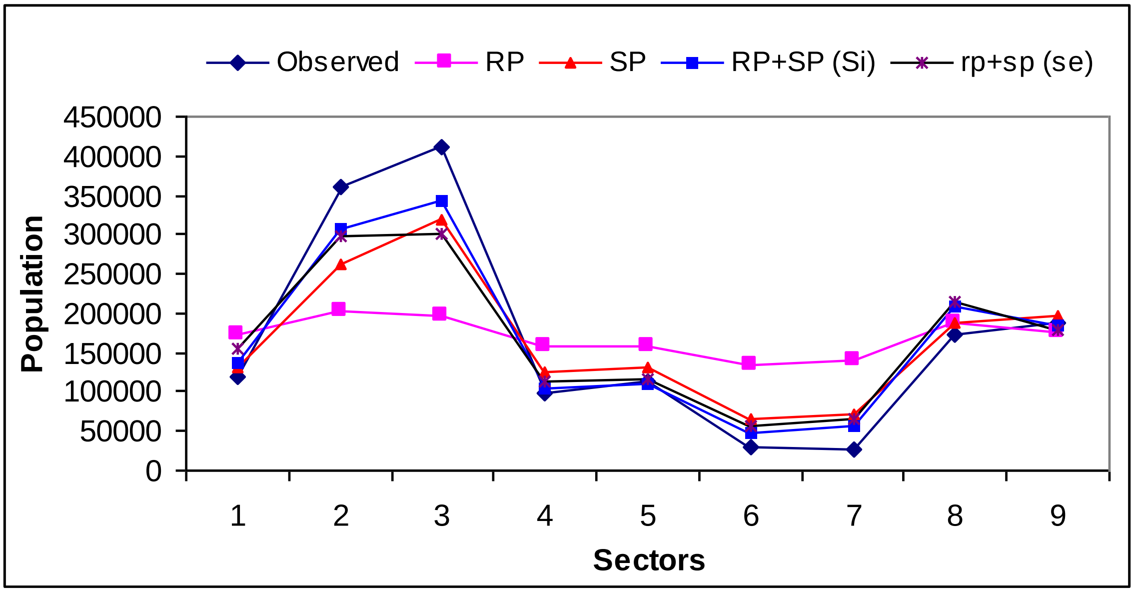

Fig. 4 Population prediction by Different Models

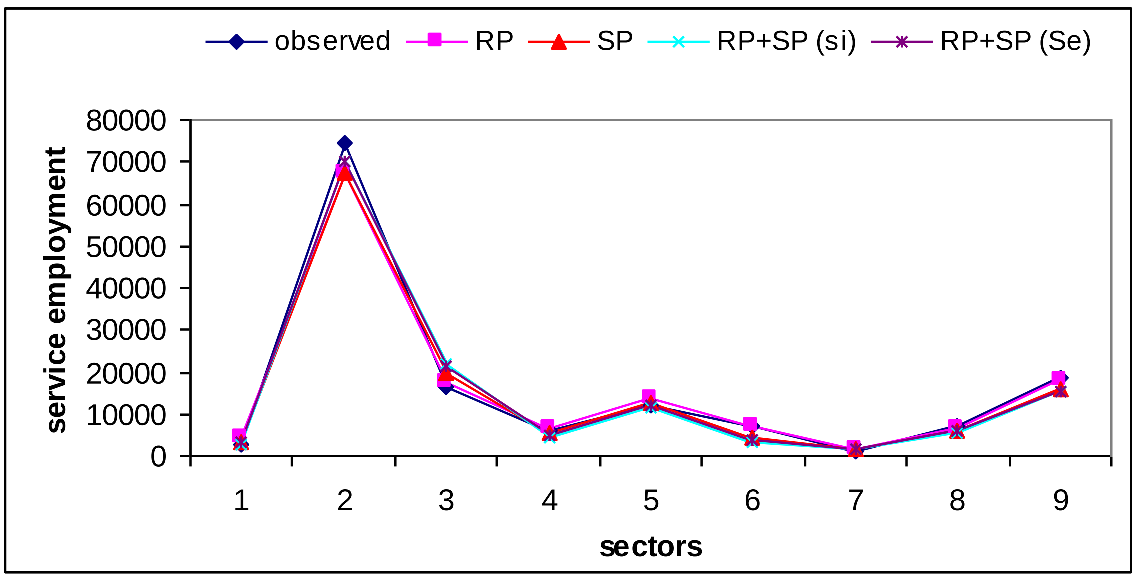

Fig. 5 Service Employment prediction by Different models

7. FORECASTING SCENARIO FOR FUTURE

The same micro-simulation framework can be used for future scenarios also. The socio-economic attributes applicable for that future scenario need to be obtained by the socio-economic charactrisation model. The difficulty lies in suggesting the distribution of the socio-economic variables for the future scenario.

8. CONCLUSIONS

This paper describes the application of micro-simulation model to Thane city of MMR, an Indian City. Integration of the sub-models in forming an operational model to represent the behaviour of individuals in residential location choice is presented in the light of available data. For any model to be identified as a useful tool, it is necessary to demonstrate that it can reproduce the observed behaviour. Three residential location choice models were used for finding the distribution of population and employment. It is observed that the RP model, SP model and RP+SP model predicted the distribution based on their model strengths and goodness of fit. The joint model was predicting almost very close to the observed distribution of base year.

9. REFERENCES

Aldana, E., deNeufville, R. and Stafford, J. H. (1973). "Micro-analysis of urban transportation demand, Highway Research Record 466, TRL, Washington DC., 1-11.

Ettema, D., Tamminga, G., Timmermans, H. and Arentze. T. (2005). "A micro-simulation model systems of departure time using a perception updating model under travel time uncertainty". Transportation Research, 39A, 325-344.

Goulias, G. K. and Kitamura, R. (1992). "Travel demand forecasting with dynamic micro-simulation". Transportation Research Record 1357, 8-17.

Gaudry, M. J. I.., Jara-Diaz, S. R. and Ortuzar, J. de D. (1989). "Value of time sensitivity to model specification". Transportation Research, 23B, 151-158.

Hunt, J. D., Donnelly, R., Abraham, J. E., Batten, C., Freedman, J., Hicks, T., Constinett, P. J. and Upton, W. J. (2001). "Design of a statewide land use transport interaction model for Oregon". In proceedings of 9th World Conference on Transport Research, Seoul, 6152.

Kribich, V. (1979). "Modelling car availability model split and trip distribution by Monte Carlo simulation: A short way to integrated models". Transportation, 8, 153-166

Kumar, M. and Krishnarao. K. V. (2005). "A Stated preference Experiment on Attractiveness of Residential Location Choice". In Proceedings of CUPUM-05, London.

Kumar, M. and Krishnarao. K. V. (2006). "Discrete Choice Models for Car Ownership, Residential Location Choice and Travel". Unpublished Ph.D. Thesis, Indian Institute of Technology, Bombay, India

Orcutt, G. H., Greenberger, M., Korbel, J. and Rivlin, A. M. (1961). "Micro-analysis of socio-economic Systems: A simulation study". Harper Brothers, New York.

Raju, K. A., Sikdar, P. K. and Dhingra, S. L. (1998). "Micro-simulation model for residential location and transport choice to work". Unpublished Ph.D. Thesis, Indian Institute of Technology, Bombay, India

Subbarao, Ch. V. and Krishanrao, K. V. (2001). "A micro-analytic land use model using artificial neural networks". In proceedings of 9th World conference on Transport Research, Seoul, 6122.

Sugik, N. and Miyamoto, K. (2004). Spatio-temporal aggregation effects and path dependence in a land use micro-simulation system". In proceedings of 10th World Conference on Transport Research, Turkey, 1014.

Ton. T. and Hensher, D. A. (2001). "Synthesising population data: the specification and generation of synthetic households in TRESIS 2.0". In proceedings of 9th World Conference on Transport Research, Seoul, 6403.

Wegener, M. and Spiekermann. K. (2001). "Simulation of activities, land use, transport and environment". In proceedings of 9th World Conference on Transport Research, Seoul, 6144.

Wilson, A.G., and Powell, C. E. (1974). "A new representation of the urban system for modelling and for the study of micro-level interdependence, working paper 64, Department of geography, University of Leeds, UK.