Proceedings of the Eastern Asia Society for Transportation Studies, Vol. 5, pp. 2463 -2469, 2005

RESIDENTIAL AND TRAVEL MODE CHOICES IN THE DEVELOPING WORLD: A COMPARATIVE STUDY OF ELEVEN DEVELOPING CITIES

Metin SENBIL
COE Researcher
Graduate School of International Development and Cooperation, Hiroshima University
Higashi Hiroshima
739-8527 Japan
e-mail: senbil@hiroshima-u.ac.jp

Akimasa FUJIWARA
Professor
Graduate School of International Development and Cooperation, Hiroshima University
Higashi Hiroshima
739-8527 Japan
Fax: +81-82-424-6921
e-mail: afujiw@hiroshima-u.ac.jp

Junyi ZHANG
Associate Professor
Graduate School of International Development and Cooperation, Hiroshima University
Higashi Hiroshima
739-8527 Japan
Fax: +81-82-424-6919
e-mail: zjy@hiroshima-u.ac.jp

Abstract:

Using person-trip survey data collected in eleven developing cities throughout the world, a comparative study is conducted in a comparative manner by elucidating the neighborhood characteristics compiled from the person trip data sets. To do so, data mining strategy is adopted to identify three different neighborhood types, namely, compact, transit-oriented and suburban neighborhoods. To all of the cities, we apply a generic choice model to reveal the underlying factors of modal uses. The study is concluded with a discussion on the inferences as estimation results.

Keywords: Developing city, neighborhood choice, mode choice

1. INTRODUCTION

Residential location is a prime determinant of almost all of the travel decisions made by households. In addition to short term transportation decisions such as those on daily trip chains, long term travel decisions such as decisions on automobile ownership are generally centered around the residential location. Thus, in this study, our focus is on the household choice made on the residential location and the commute trip mode. To do so, we define three alternative residential areas- compact, transit-oriented and suburban, differences of which are captured with respect to the organization of the land use in the residential zone and the transportation infrastructure. We hypothesize that residential areas, which are compact and transit-oriented are more sustainable than the suburban one, which is thought to foster the private car ownership and its intensive use in environmentally and economically harmful ways to the society. Thus, it becomes important to identify the causes that affect the selection of the residential location and the commute trip mode, and incorporate the results in a more comprehensive framework of environmental management when devising land use strategies to promote urban sustainability.

There have been many efforts to elucidate to incorporate these close relationships into models of residential location and travel choices. One earlier example is the study by Anas (1981), who estimates a Multinomial Logit model on joint residential location and travel mode choice. One of the latest attempts, Ben Akiva and Bowman (1998) has come from the newly emerging activity based travel demand approach incorporating the latest developments in the travel demand field and the nested logit model that controls for the IIA characteristic of the Multinomial Logit model. Almost all of the attempts emphasize the need to control for the IIA even in the travel mode selection and choice set generation from an urban area normally comprised of tens if not hundreds of zones to select.

Departing from the usual practice, in this study, we offer a simultaneous residential and mode choice model based on person-trip data. No land use data collected are used in the analyses, instead we have derived the land use data from the person-trip data by data mining. In this regard, we first present preliminary analyses in the second section, and in the third section we give detailed information on our methodology. Two statistical methodologies we use are an automatic identification detector, CHAID and the discrete choice modeling. With the first methodology we intend to differentiate land uses of transportation zones, and with the second methodology we estimate the utility functions of compact, transit-oriented and metropolitan residential areas and the travel modes, private car and public transit, based on the information available from the person-trip data sets of eleven developing cities around the world. The classification and estimation results are presented in forth section, and the fifth section concludes the study.

2. JICA 11 PERSON-TRIP DATA

The data sets we have used in our analyses are the person-trip data sets of eleven cities of the developing countries around the world (Table 1). The data has been administered by Japan International Cooperation Agency (JICA) and collected between 1996 and 2000. Table 1 supplies the information on populations, surveyed households and individuals, and one day person-trips. As can be noticed from Table 1, surveyed individuals constitute a very tiny fraction of the populations with the ratios that range from one to six percent. Number of trips per person ranges between 1.97 (Cairo) and 2.71 (Kuala Lumpur).

Table 1. Person-Trip Data Sets

City	Population	Survey year	# of Households	# of Persons	# of Trips
Belem, Brazil	1,782,394	2000	6,889	24,043	59,529
Bucharest, Romania	2,150,000	1998	32,888	67,509	143,311
Cairo, Egypt	14,400,000	2001	41,962	136,070	268,360
Chengdu, China	3,090,000	2000	14,537	31,188	70,199
Damascus, Syria	3,078,190	1998	17,202	38,490	81,698
Jakarta, Indonesia	20,964,000	2000	100,864	425,237	1,083,280
Kuala Lumpur, Malaysia	1,390,800	1998	27,331	80,560	218,460
Managua, Nicaragua	1,200,000	1998	8,089	24,854	54,138
Manila, Philippines	9,454,000	1996	60,752	231,889	471,035
Phnom Penh, Cambodia	1,152,000	2000	6,446	18,664	40,369
Tripoli, Lebanon	330,900	2000	1,321	3,608	7,615

Some of the socio-economic characteristics of the individual cities are given in Table 2. Household sizes reported in Table 3, shows compliance with the generally known socio-economic conditions of the developing countries. Extended families or several families in one household are not uncommon in these countries. Then it might be said that the pattern of daily activities in these countries might be significantly different from the developed world where household sizes are smaller. Especially the daily activities such as shopping, social and recreation might be easily affected by the household size and composition and this might affect the number of home-based trip chains. Unfortunately, in the literature we do not have much study that has reported any information on the household composition and the activity-trip generation. Another point to be highlighted in Table 2, is the contradiction between the household income and the car ownership for the city of Tripoli, which is a city of Lebanon. For Tripoli, the high ratio of automobile ownership might a result of past richness of the country before the Civil War that raged the whole country in the early 1980s, according to this supposition, it might be the case that the private car stock of Tripoli might be aged more than 20 years.

Table 2. Socio-economic Characteristics of JICA-11 Person Trip Data Sets

	# of household members		Household income		Age		Sex[a]		Car ownership[a]	
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Belem	5.31	2.10	508.09	700.35	29.63	16.86	.49	.50	.19	.39
Bucharest	3.28	1.36	432.40	577.29	38.87	19.26	.48	.50	.43	.50
Cairo	4.93	1.71	94.51	143.22	25.98	15.18	.65	.48	.20	.40
Chengdu	2.89	1.76	179.79	129.94	38.10	16.65	.54	.50	.07	.25
Damascus	5.58	2.14	258.69	159.84	31.31	18.09	.69	.47	.30	.46
Jakarta	3.67	1.55	245.15	220.80	29.76	15.67	.54	.50	.19	.39
Kuala Lumpur	4.75	1.80	867.55	720.55	28.95	17.93	.61	.49	.68	.46
Managua	6.19	2.80	235.74	306.46	26.07	15.30	.51	.50	.19	.40
Manila	4.42	1.65	434.72	560.92	28.40	16.38	.47	.50	.13	.33
Phom Penh	5.89	2.23	82.69	76.76	NA[b]	NA[b]	NA[b]	NA[b]	.15	.35
Tripoli	5.36	2.18	4.05	1.30	26.25	15.99	.65	.48	.71	.45

a. Represented by dummy variables: 1 stands for male in Sex, for having at least one car in Car ownership.

b. Not Available

The time use in JICA 11 cities is portrayed in Table 3[1]. In the table, it can be seen that the work activity duration generally stays above eight hours per day; the longest average work duration is experienced in the city of Managua, which is approximately two hours longer than the shortest average work duration which is observed at Damascus. On the other hand, Manila has the longest average commute trip duration. For the non-commute trips, Chengdu and Manila are the two cities which stay above than all other cities. We find a low level correlation between commute trip duration and the departure from home (-0.13, p =0.00) when all data sets are pooled. When we compute the same correlation for different cities, we find the highest absolute correlation value for Damascus (-0.16, p = 0.00), followed by Jakarta (0.15,p=00); the lowest correlation is found for the city of Belem (-0.07,p = 0.00). In addition to this, total trip duration is shorter for the cities of Belem, Tripoli and Damascus than the other cities significantly. There is a very time negative correlation between total trip duration and the departure from the home when all trip are considered as well as when we exclude the cases in which individuals have also pursued commute trips.

[1] Note that the person-trip data is tabulated at the individual level, however we have the information indexed at the household level (see below).

Table 3. Time Use in JICA 11 Cities

a	Time of departure from home to start the day[b]	Time of return to home to retire the day[b]	Time spent out-of-home[c]	Total duration for work[c]	Total duration for trips[c]	Average commute trip duration[d]
Belem	9:25	16:03	6.73	8.74	.87	24.09
Bucharest	8:28	15:44	6.91	8.43	1.27	39.78
Cairo	7:25	15:37	7.93	8.16	1.22	43.99
Chengdu	8:19	14:46	6.71	8.30	1.39	37.43
Damascus	9:09	15:41	6.90	7.47	.98	29.54
Jakarta	7:38	15:15	7.20	8.06	1.33	39.92
Kuala Lumpur	8:02	17:08	9.11	8.80	1.34	34.30
Managua	8:30	16:11	7.57	9.55	1.32	40.84
Manila	6:44	15:47	6.31	9.02	1.40	48.58
Phom Penh	NA	NA	NA	NA	NA	NA
Tripoli	7:36	14:59	7.45	7.82	.78	25.66

a. Non zero values are taken into consideration

b. time of day in 24:00 system

c. In hours

d. In minutes

Contrary to the above information supplied by the Tables 1-3, Table 4 gives information on the number of daily activities, along with derived variables about residential zone compactness. As noted in the introduction section of this study, one of the residential land use type is compact land use, which generally contains most of the activities within the residential area, hence trip made are contained in the residential zone. But this has to be approached with caution as we do not have exact information on the residential zones, so we contain ourselves to the transportation analysis zones that are devised for each city, and we assume that the transportation analysis zone in which household resides as the residential zone. Because of this reason, compactness measure might inherently bear some bias in case of zones which are too big or too small with respect the residential zone itself. In fact this drawback constitutes one of the most important reasons why we have employed CHAID (see below) by including activity-trip variables in order to detect different land uses by using person-trip data sets. According to the compactness ratio for all trips, Jakarta, Phom Penh and Manila constitute the cities which have compact residential zones.

Table 4. One-day Activities and Compactness of the Residential Zone

	# of activities		Compactness dummy for commute trips[a]		Compactness ratio for all trips[b]	
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Belem	2.47	1.14	.20	.40	.33	.45
Bucharest	2.12	.63	.10	.30	.21	.40
Cairo	1.96	.35	.18	.39	.28	.45
Chengdu	2.25	.78	.34	.47	.33	.46
Damascus	2.12	.58	.20	.40	.26	.43
Jakarta	2.51	.97	.33	.47	.47	.48
Kuala Lumpur	2.71	1.75	.16	.36	.25	.42
Managua	2.18	.71	.12	.32	.21	.40
Manila	2.03	.93	.28	.45	.40	.48
Phom Penh	2.16	1.47	.33	.47	.45	.50
Tripoli	2.11	.56	.27	.44	.32	.46

a. 1 represents the case that commute trip is contained in the residential zone, 0 represents else.

b. The ratio of the number of trips which have both origins and destinations in the residential zone to the total number of trips.

In the next section, we devise a methodology to differentiate three neighborhoods, which are compact, transit-oriented and suburban. To differentiate these residential areas from each other, we have used daily activity trip characteristics, which are given in Table 3 and Table 4.

In Table 4, three variables are given which refer to activity locations visited throughout the survey day with respect to the residential zone.

3. METHODOLOGY

As mentioned in the previous sections, the person-trip data sets do not contain information about the land use of the individual cities. The only information available is on the trip made and the personal characteristics. Besides, not all of the eligible household members are surveyed. With the available information, to classify zones with respect to three broad categories of compact, transit-oriented and metropolitan neighborhoods, we make use of the classification routines known as CHAID (Chi-Squared Automatic Identification Detector). Before giving details on the classification method, it will be helpful to devise the characteristics of the land use categories.

Compact neighborhood is an area where different activities are reachable within walk or short motorized trips from the residential location and the trip chains formed are generally contained in close vicinity of the residential area. In compact neighborhood, either most of the trips are contained in the residential zone or intra-zonal are very short in duration by any mode and mostly by non-motorized modes. Transit-oriented neighborhood is well served by the transit services and the activities are in general in close vicinity of the residential area, the trip chains are mostly reliant on the transit services. Suburban neighborhood accommodates a bedroom community. Most of trips especially commutes are to other zones and by motorized private modes. A significant feature of metropolitan residential area is that a good amount of the trips goes to other areas with long durations.

In order to detect and classify the residential areas into three sets, we have make use of the CHAID first developed by Kass(1981) and later refined by Biggs (1991). An algorithm is available in commercial statistical analysis package SPSS called as Answer Tree.

Regarding the discrete choice, we propose a Mixed Logit model for estimating the choice model. As noted by Train (2003), Mixed Logit model approximates every random utility model and obviates three basic drawbacks of Logit model by allowing for random taste variation, unrestricted substitution patterns, and correlation in unobserved factors over time (p. 138). The mixed logit model is briefly explained below. The individual utility consists of two parts observed (V) and unobserved parts (ε).

$$U_{inm} = V_{inm} + \varepsilon_{inm} \tag{1}$$

where the subscripts *i, i, m* represents individual, *i*, land use, *l*, and commute trip mode, *m*, respectively. From Eq. 1, by assuming unobserved part is distributed as IID extreme value, the logit probability, *Pr* of a neighborhood type, *n*, and mode choice, *m*, then reads as (*23*):

$$Pr_{inm}(\boldsymbol{\beta}_i) = \frac{e^{\boldsymbol{\beta}'\mathbf{x}_{inm}}}{\sum_{n\in N, m\in M} e^{\boldsymbol{\beta}'\mathbf{x}_{inm}}} \tag{2}$$

Note that this probability is for certain values of parameter vector $\boldsymbol{\beta}$. Hence the unconditional probability over all possible values of parameter vector $\boldsymbol{\beta}$ depends on its probability distribution, *g* (see Train, 1997)

$$P_{inm}\left(\boldsymbol{\theta}^*\right)=\int P_{inm}\left(\boldsymbol{\beta}_i\right)g\left(\boldsymbol{\beta}_i\middle|\boldsymbol{\theta}^*\right)\mathbf{d}\boldsymbol{\beta}_i \tag{4}$$

integration of parameter vector over all of possible values rules out the possibility of heterogeneity across individuals. The log-likelihood function of a sample based on individual probability as in Eq. 4 then becomes

$$\ell(\boldsymbol{\theta})=\sum_i \ln\left(P_{inm}(\boldsymbol{\theta})\right) \tag{5}$$

We approximate the probability, P_{nl}, by simulation and then maximize the log-likelihood function (generally named as simulated log-likelihood). The probability, P_{nl}, is approximated by summing randomly selected values for coefficient vector many times and by taking the average of the resultant value. It has been shown by Hajivassiliou and Ruud (1994) that the bias in simulated log-likelihood decreases with increasing number of repetitions.

Let there are K observed variables some of which have random parameters distributed as IID standard Uniform distribution, i.e., $\beta_{ik} \sim U(\mu_k, \sigma_k)$, hence a random draw of β_{ik} becomes $\beta_{ik}=\mu_k+\sigma_k s_{ik}$, with s_{ik} a random draw from standard Uniform distribution (Bhat, 2001). With this, Eq. 1 becomes as (by stacking dual choice nm in c):

$$U_{ic}=\alpha_c+\sum_1^K \beta_{ick}x_{ick}+\varepsilon_{ic}=\alpha_c+\sum_1^K \mu_k x_{ik}+\sum_1^K \sigma_k s_{ik}x_{ick}+\varepsilon_{ic} \tag{6}$$

In Equation 6, random parameters have standard variations, σ, have to be estimated but those which are not random have standard deviations equal to zero. With Equation 6, the Logit function of Eq. 2 becomes as:

$$L_{ic}=\frac{e^{\alpha_c+\sum_1^K \mu_k x_{ik}+\sum_1^K \sigma_k s_{ik}x_{ick}}}{\sum_j e^{\alpha_j+\sum_1^K \mu_k x_{ik}+\sum_1^K \sigma_k s_{ik}x_{ijk}}} \tag{7}$$

The probability of land use in a cell is obtained by integrating out the random parameters from R_n, that is

$$P_n\left(\boldsymbol{\mu}^*,\boldsymbol{\sigma}^*\right)=\int_{s_{n1}=-\infty}^{s_{n1}=\infty}\int_{s_{n2}=-\infty}^{s_{n2}=\infty}\cdots\int_{s_{nK}=-\infty}^{s_{nK}=\infty}\frac{e^{\alpha_c+\sum_1^K \mu_k x_{ik}+\sum_1^K \sigma_k s_{ik}x_{ick}}}{\sum_j e^{\alpha_j+\sum_1^K \mu_k x_{ik}+\sum_1^K \sigma_k s_{ik}x_{ijk}}}dU\left(s_{n1}\right)dU\left(s_{n2}\right)\cdots dU\left(s_{nK}\right) \tag{8}$$

where $U(\cdot)$ is the standard cumulative Uniform distribution; with this, the log-likelihood can be written as:

$$\ell(\boldsymbol{\mu},\boldsymbol{\sigma})=\sum_n \log\int_{s_{n1}=-\infty}^{s_{n1}=\infty}\int_{s_{n2}=-\infty}^{s_{n2}=\infty}\cdots\int_{s_{nK}=-\infty}^{s_{nK}=\infty}\frac{e^{\alpha_c+\sum_1^K \mu_k x_{ik}+\sum_1^K \sigma_k s_{ik}x_{ick}}}{\sum_j e^{\alpha_j+\sum_1^K \mu_k x_{ik}+\sum_1^K \sigma_k s_{ik}x_{ijk}}}dU\left(s_{n1}\right)dU\left(s_{n2}\right)\cdots dU\left(s_{nK}\right) \tag{9}.$$

We estimate the log-likelihood function by using LIMDEP 8.0.

REFERENCES

Anas, A. (1981) The estimation of multinomial logit models of joint location and travel mode choice from aggregated data, Journal of Regional Science, Vol. 21, pp. 223-242.

Ben Akiva, M. and Lerman, S. (1985) Discrete Choice Analysis. The MIT Press, Cambridge, MA.

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

Bhat, C. R. (2001) Quasi-Random Maximum Simulated Likelihood Estimation Of The Mixed Multinomial Logit Model. Transportation Research B, Vol. 35, 2001, pp. 677-693.

Biggs, D.B. de Ville, and Suen, E. (1991), A method of choosing multiway partitions for classification and decision trees. Journal of Applied Statistics, Vol.18, pp.49-62.

Hajivassiliou, V. and Ruud, P. (1994) Classical estimation methods for LDV models using simulation. In Handbook of Econometrics, Vol.4. R. Engle and D. McFadden, eds., Elsevier Science, New York, pp.2384-2438.

Kass, G. (1980), An exploratory technique for investigating large quantities of categorical data. Applied Statistics, Vol.29(2), pp.119-127.

Train, K. (1997) Recreation Demand Models With Taste Differences Over People. Land Economics, Vol. 74, 1997, pp. 230-239.

Train, K. (2003) Discrete Choice with Simulation. Cambridge University Press, New York.