

Same-Day Mode Choice Modeling with Household Vehicle Usage Simulation in Developing Countries

Xia Wan, Yi Zhang, Peter J. Jin, Bin Ran, Wei Wang, and Jun Chen

This paper presents a model of same-day mode choice at the household level for developing countries. A rule-based algorithm combining classical random utility maximization theory within a microsimulation framework is used. Modeling of private vehicle usage (including vehicle allocation and sharing use in household) is an essential component of this model because vehicle deficiency is common in developing countries. This model consists of four steps: (a) the allocation of private vehicles (car, motorcycle, and bicycle) in a household, (b) the mode choice of private vehicle users specified in the first step, (c) vehicle sharing in a household, and (d) the mode choice of individuals who do not use private vehicles. The adaptability of the model was improved by simulations on car, motorcycle, and bicycle usage. Discrepancies in the mode choice behavior of household members with and without the use of private vehicles are captured in this paper through different modeling methods. The rule-based algorithm, binary logit model, multinomial logit model, and mixed logit model were applied together in this four-step model. Travel diary survey data from 2007 from Bengbu, China, were used as an example for the validation test of this model. The results demonstrate that this model can accurately predict the mode choice of all household members in an internally self-consistent and theoretically credible manner for a midsize city in China. The proposed model is highly conducive to travel demand forecasting and transportation policy making.

Tour-based mode choice, which takes into account temporal-spatial constraints within a tour, has shown incremental advantages over trip-based models (1, 2). Researchers considered the interactions between tours within 1 day and developed the same-day mode choice model, making its adaptability more effective and efficient (3–5). The same-day or tour-based mode choice model has been well studied in developed countries. However, it has not been well studied in developing countries. In developing countries, mode choice decisions may differ considerably from those in developed countries because of differences in vehicle ownership, mobility needs, and people's travel and activity characteristics (6–9).

Most current mode choice models assume that cars are always available to all drivers if the household owns cars. This assumption

may be true in developed countries, but does not capture reality in developing countries (6, 8). The higher probability of car deficiency in developing countries probably requires operating vehicle allocation before mode choice. Car-deficient households are households where the number of drivers exceeds the number of cars. Moreover, three types of private vehicles (car, motorcycle, and bicycle) should be considered in the vehicle allocation step in households because this group more accurately represents the transportation situation in developing countries. In addition, the likelihood of household members switching travel modes within 1 day and the likelihood of vehicle sharing among household members have to be included in the mode choice model.

In this research, a same-day mode choice model, in which the activities of household members are predetermined, is established as a four-step model. The first step allocates vehicles at a household level. Household members are divided into two types: the members who use private vehicles and the ones who do not. The second step decides mode choices of individuals with private vehicles. The third step considers vehicle sharing among household members. The fourth step simulates mode choices of individuals without vehicles based on chronological order. The area studied in this paper is the urban area of Bengbu, China, a midsize city, which is a common type of city in China.

The scope of this paper is as follows. The following section reviews current studies on mode choices and applications in developing countries. The next section presents a model framework consisting of four subsections, with algorithms and detailed descriptions of models. The next section gives a case study of a mode choice model in the midsize city of Bengbu, China. The last section summarizes the conclusions and proposes future work.

LITERATURE REVIEW

In the past 30 years, the typical mode choice models have been trip-based models, in which the traditional random utility maximization theory is primarily adopted (1). In the Netherlands in the late 1970s, researchers developed tour-based models in which round-trip journeys are considered instead of single trips (9–12). The benefits of the tour-based models are that the temporal-spatial constraints within a tour are considered. Later, tour-based mode choice models were developed in Denmark, Poland, and the United States (2, 5, 13, and 14). Miller et al. summarized that there are some limitations of the tour-based mode choice models that preceded their model. These limitations include reliance on some tree logit forms, simplification of the definition and construction of tours, assumption of a main mode, separate calibration by purpose, and

X. Wan, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, and School of Transportation, Southeast University, Si Pai Lou 2, Nanjing 210096, China. Y. Zhang and P. J. Jin, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, Madison, WI 53706. B. Ran, W. Wang, and J. Chen, School of Transportation, Southeast University, Si Pai Lou 2, Nanjing 210096, China. Corresponding author: J. Chen, chenjun@seu.edu.cn.

Transportation Research Record: Journal of the Transportation Research Board, No. 2239, Transportation Research Board of the National Academies, Washington, D.C., 2011, pp. 23–33.
DOI: 10.3141/2239-04

use of explicit assumption about car availability rather than car allocation (1).

Recently, new modeling methods have been applied to build a mode choice model. Miller et al. proposed a new tour-based mode choice model in 2005. Their model assumes that if a private car is used in a tour, that car must be used for the entire trip chain because the car must be returned home at the end of the tour (1). In addition, some anchor points exist where individuals can change the travel mode. By using an ensemble of conditional and unconditional classifiers, Biagioni et al. adapted existing data-mining methods to further improve predication performance of the mode choice model (10). Ramadurai and Srinivasan investigated the dynamics and variability in the same-day mode choice decisions. This research concluded that the tour-based mode choice models are sensitive to state-dependence, which refers to influence of the previous episode on the current episode (3). Vovsha et al. compared the tour-based mode choice model systems applied in San Francisco, California (2001); New York (2002); Columbus, Ohio (2004); and Sacramento, California (2006). They suggested the concept of hybrid of probabilistic choice and that rule-based algorithms could be employed in future mode choice models (8).

Many travel demand forecasting models assume mode choice is mainly determined at the personal level. Even though the importance of representing group decision-making mechanisms of household behavior has been identified since the 1980s, studies about mode choice behavior at the household level are relatively new and limited (15). Zhang et al. developed a new household discrete choice model in 2009 to represent heterogeneous group decision-making mechanisms in choice behavior. Household utility function was defined to theoretically reflect household members' preferences and intrahousehold interactions. With the data collected in two Japanese cities in 2004, the effectiveness of the proposed household decision model was empirically confirmed (16). In addition, most present mode choice models assume that cars are always available for household members whenever they are needed. However, this assumption may not be true in some developing countries in which motor vehicles are not common in households. The tour-based mode choice model proposed by Miller et al. characterized the household interactions of vehicle allocation, ridesharing, and drop-off and pickup of household members (1). Anggraini et al. explicitly considered within-household interactions in activity-travel choices to refine ALBATROSS, a rule-based, activity-based modeling system (17). In 2004, the chi-square automatic interaction detection algorithm was applied in this research to derive a decision tree for the car allocation decisions in automobile-deficient households using a large activity diary data set recently collected in the Netherlands. The researchers found that the probability of the male getting the car is considerably higher than the female getting the car in many condition settings (17).

Mode choice in developing countries may differ significantly from mode choice in developed countries in several aspects. Choices could include vehicle types (two-wheelers versus four-wheelers), vehicle ownership levels, socioeconomic characteristics, perception of subjective factors, and variability in choice set. Srinivasan et al. investigated the differences between mode choice propensity for two-wheelers and four-wheelers and differences in sensitivities to travel time and cost across different user groups in India (18). Rajagopalan and Srinivasan analyzed household-level mode choice and modal expenditure decisions through a multiple discrete-continuous extreme value model in the context of Chennai city in India (19). Yagi and

Mohammadian studied the combined models of mode and destination choices for home-based tours within an activity-based model for Jakarta, Indonesia, and the simulation results showed that choice alternatives, the structure of the model, and key variables differ from those in the developed countries (20). Hence, there is a significant need for a mode choice model for developing countries, not just at the household level, and also for simulation of household private vehicle usage.

MODEL FRAMEWORK

Notation

Table 1 presents the notation used in formulating the mode choice model.

Algorithm of Mode Choice Model

The mode choice model proposed in this paper is an agent-based microsimulation framework. This model simulates the mode choice decision for the trips of all household members within 1 day. The decision is the output of the model. Activity schedules of all household members are regarded as one of the inputs to the model, which complies with the basic modeling sequence that the mode choice is made after the activity generation (12, 21). Household data, land use data, public policies, and road network data are also input factors. The model strives to capture interactions between household members, the dependence between trips of one person, and impacts of transportation environments, such as land use and road network, on mode choice.

Figure 1 is the flowchart of the proposed model. The model consists of four steps:

1. Private vehicle allocation at the household level. Vehicles discussed in this paper are cars, motorcycles, and bicycles.
2. Mode choice of household members who are allocated private vehicles. The household member must use the assigned vehicles for at least one trip within this day.
3. Search for the vehicle sharing use in the household. Vehicle sharing use contains two conditions:
 - Ridesharing (car, motorcycle, and bicycle). Some household members travel as a passenger, because of joint and escort activities.
 - Whether the assigned vehicle can be used by other household members when it is free at home.
4. The mode choice follows a chronological sequence for the rest of household members who are not allocated private vehicles.

Step 1. Vehicle Allocation in Households

Recently, Miller et al. and Anggraini et al. proposed that car deficiency should be considered, raising a vehicle allocation issue (1, 17). In developing countries, a car, used by only a small proportion of people, is not only a swift and convenient transportation tool, but also a symbol of social status and wealth. A motorcycle is a widely used vehicle, the advantages of which include savings in time and physical energy versus riding a bicycle. Riding a bicycle, however, is not restricted by a driver's age or ability. Therefore, the sequence

TABLE 1 Notation Used in Formulating Mode Choice Model

Notation	Description
h	Index for household $h=1, \dots, H$
ih	Individual i from household h
Ih	$\{Ih=1, 2, \dots, \text{number of members in household } h\}$, represents the set of household members in household h
I_{\max}	Maximum household size among all households
PV_{ih}	Private vehicle (car, motorcycle, and bicycle) allocated to individual i in household h . If PV_{ih} is equal to car, individual i is allocated with a car this day. The same allocation occurs with a motorcycle or bicycle. If PV_{ih} is equal to 0, no private vehicle is assigned.
$\#Car_h$	Number of cars in household h
$\#Ih_car$	Number of individuals who are eligible for using car (with driver's license and $PV_{ih} = 0$) in household h
$\#Motorcycle_h$	Number of motorcycles in household h
$\#Ih_motorcycle$	Number of individuals who are eligible for using motorcycle (with driver's license for motorcycle and $PV_{ih} = 0$) in household h
$\#Bicycle_h$	Number of bicycles (including electric bicycles) in household h
$\#Ih_bicycle$	Number of individuals who are older than 6 years but younger than 70, nondisabled, capable of riding bicycle, and $PV_{ih} = 0$ in household h
t_{ih}	Index for the trip t of individual i in household h
T_{ih}	Set of trips for individual i in household h
c_{ih}	Index for the trip chain c of individual i in household h
C_{ih}	Set of trip chain for individual i in household h
m	Index for travel mode $\{m = (\text{car} = 1, \text{motorcycle} = 2, \text{bicycle} = 3, \text{walk} = 4, \text{transit} = 5, \text{taxi} = 6, \text{company vehicle} = 7, \text{passenger} = 8 = M)\}$
$m_f(t, i)$	Set of feasible modes for trip t for individual i
$m(t, i)$	Travel modes for trip t for individual i
η_h	Shared normal error that is common to the utility of "not being used" of all this type of vehicle in household h , and $\sim MVN(0, \Sigma_1)$
η_{ih}	Unobserved random component of utility for individual i belonging to household h that is common across the same type of but different vehicles, and $\sim MVN(0, \Sigma_2)$
ξ_h	Independent part of "not being used" utility that varies across different vehicles in the household h and is distributed as $\sim IID(0, \pi^2/6\sigma_1^2)$, where σ_1 represents the Gumbel scale and is set up to 1
ξ_{ih}	Independent component of time-varying error of utility that varies across different vehicles in household h and is distributed as $\sim IID(0, \pi^2/6\sigma_2^2)$, where σ_2 represents the Gumbel scale and is set up to 1
η_{ijih}^m	Shared normal error that is common to individual i and mode m for mode choice of trip t_{ih} , and follows an MVN distribution $\sim MVN(0, \Sigma_3)$
ξ_{ijih}^m	Independent component of time-varying error of utility that varies across different trip t_{ijih} of individual i and mode m , and is distributed as $\sim IID(0, \pi^2/6\sigma_3^2)$, where σ_3 represents the gumbel scale and is set up to 1

NOTE: MVN = multivariate normal (distribution); IDD = independent and identically distributed.

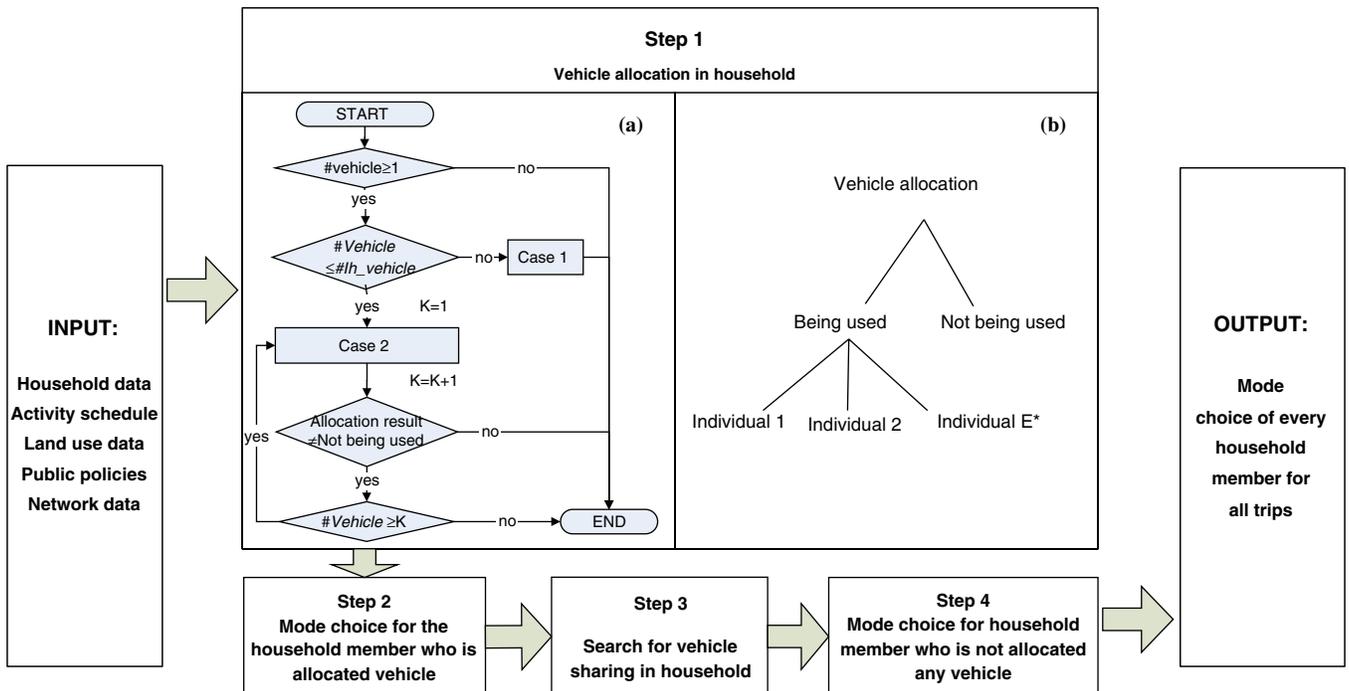


FIGURE 1 Flowchart of agent-based mode choice microsimulation model: (a) Step 1, vehicle allocation model of a household, and (b) model structure of mixed nested logit model in Step 1, Case 2. (Individual 1, 2, . . . , E are household members eligible for vehicle allocation.)

of vehicle allocation within households in developing countries would be as follows: car, motorcycle, and bicycle.

Step 1a in Figure 1 is the flowchart for the detailed sequence of vehicle allocation. Vehicles in this step can be car, motorcycle, or bicycle. First, if $\#Car_h$ is larger than 1, car allocation proceeds. Otherwise, the motorcycle allocation branch is operated. When $\#Car_h \geq 1$ and $\#Car_h \geq \#Ih_car$ (denoted as Case 1), a binary logit model is applied to decide whether all drivers will use the car. When $\#Car_h \geq 1$ and $\#Car_h < \#Ih_car$ (denoted as Case 2), a mixed nested logit model is operated to determine which car is allocated to which driver until all the cars have been specified. After the car allocation is finished, the motorcycle and bicycle allocation branches follow the same procedure, first for motorcycles, then for bicycles. If an individual i is allocated with a private vehicle, PV_{ih} is made to be equal to this type of vehicle; otherwise, PV_{ih} is 0.

Case 1. Binary Logit Model

When $\#Vehicle_h \geq 1$ and $\#Vehicle_h < \#Ih_vehicle$ (vehicle here is car, motorcycle, or bicycle), three binary logit models for different market segment (car, motorcycle, and bicycle) are established. These models solve the issue of whether the eligible members would use the correspondingly qualified vehicles.

With car allocation as an example, the alternatives of binary logit model are “to use this car” and “not to use this car” for individual i . The alternative with maximum utility is chosen with the effect of sociodemographic characteristics (e.g., gender, age), activity pattern and trip characteristics (e.g., tour number, activity type, longest trip distance), family attributes (e.g., income, family size), land use (e.g., density and diversity), road network, and public transit service. Meanwhile, the role that a person plays in a family is also captured to reflect the interactions between household members.

Case 2. Mixed Nested Logit Model

When $\#Vehicle_h \geq 1$ and $\#Vehicle_h \geq \#Ih_vehicle$, a mixed nested logit model is applied to process vehicle allocation. The output of the mixed nested logit model is to assign the vehicle to a certain household member or to no one. Every time the model is executed, the allocation of one vehicle is performed. If there are two cars in a household, the allocation of the first car is done, and the other car is assigned to an eligible household member in the second run. However, if nobody is assigned a car in the first run of the model, the car allocation would stop and the motorcycle allocation branch would begin. Index K is defined to monitor how many times a car, motorcycle, or bicycle is allocated.

The two-stage nested model depicted in Step 1b of Figure 1 is used to conduct the vehicle allocation within a household. The alternatives in the first stage of vehicle allocation are “being used” and “not being used.” This step is modeled by a binary utility structure. “Not being used” is chosen only if its utility exceeds that of the vehicles taken by someone in this family. If “being used” is selected in the first stage, the individual decided in the second stage is chosen on the basis of the utility maximization principle.

Utility specifications are as follows. The utility of the vehicle not being used is expressed by

$$U_{h_Nuse} = \beta_1 X_h + \eta_h + \xi_{ih} \quad (1)$$

The utility of the vehicle being used allocated to an individual i is expressed by

$$U_{ih} = \beta_2 X_{ih} + \eta_{ih} + \xi_{ih} \quad (2)$$

where

β_1 = vector of parameters associated with the vehicle not being used,

β_2 = vector of parameters associated with the vehicle being used,

X_h = vector of explanatory attributes affecting the vehicle not being used for household h (e.g., household size, household income),

X_{ih} = vector of explanatory attributes affecting vehicle allocation utility for individual i from household h (e.g., age, gender), and

$\eta_h, \eta_{ih}, \xi_h,$ and ξ_{ih} = error structure of the mixed nested logit model to show the correlations between repeated instances within household.

The error structure enables relaxing the assumption that error terms are independent and identical across alternatives and that error term variance is constant across households. Detailed descriptions of these four error variables are presented in Table 1.

The random utility maximizing behavioral framework is as follows. The utility maximization for “being used” and “not being used”: the alternative of “not being used” is chosen if $U_{h_Nuse} \geq 0$; otherwise, “being used” is chosen. The utility maximization for an individual deciding if the vehicle is used: the vehicle is allocated to individual i , if $U_{ih} > U_{jh}$ for all $j \neq i$, where j and i are eligible members in household h for this vehicle allocation.

The number of eligible people for one type of vehicle can vary from one household to another. Also, within one household, the number of eligible household members may be different when the mixed nested logit model is operated at different stages for the same type of vehicle. To address this issue, the choice set dimension of the mixed nested model is set to the maximum household size among all households, named by I_{max} . Thus, all households have a constant choice set dimension (I_{max}), but the number of members who exceed the real eligible members of household h ($\#Ih_vehicle$) are treated as virtual and ineligible members who are never allocated any vehicle. The person-level utility for ineligible members is set as a large negative number to exclude their choice for vehicle usage. To increase estimation efficiency and robustness, generic variables are specified in such a way that life-cycle and household role effects can be estimated as variables across households. The coefficients associated with these variances are estimated by using the simulated maximum likelihood technique. A detailed description of this method is given by Train (22). A Monte Carlo simulation with 2,500 pseudorandom normal draws was employed for estimation of this model.

Step 2. Mode Choice of Household Members with Vehicles Allocated

Step 2 conducts the mode choice of people allocated private vehicles in Step 1 within 1 day. This step is based on a rule-based assumption: the one who is allocated a private vehicle must use this vehicle for the longest trip.

First, the longest trip of individual i is enforced to use the allocated private vehicle PV_{ih} . If more than one trip of individual i within 1 day

has the same longest distance (for example, the home to work and work to home travel pattern has the same trip distance), all of these trips must use the allocated vehicle as the travel mode. However, there are some reasonable constraints: the private vehicle has to start from home and it must be sent back home after finishing the home-based trip chain. Therefore, once the trips in trip chain C_{ih} are enforced by the PV_{ih}, on the basis of the constraints mentioned above, the trips forming the closed loop with the enforced trips are valued by PV_{ih}.

Meanwhile, an individual who uses a private vehicle within 1 day has the probability to use other travel modes. Thus, the mode change of the private vehicle user within 1 day is considered in this model. In this case, there is an assumption that a person cannot be regarded as a driver using two different private vehicles within 1 day. Therefore, the possible alternatives for a private vehicle driver switching travel modes include transit, walk, taxi, or company vehicles. There are two types of anchor points where the driver can change travel modes: (a) the non-home-based subchain belonging to the enforced main chain that does not involve the enforced private vehicle and (b) the home-based trip chain of driver i that does not involve the enforced private vehicle.

A multinomial logit (MNL) model is used to model the mode choice of a private vehicle driver i for both two types of anchor points. The five alternatives of MNL model are to continue using PV_{ih}, to switch to transit, to walk, to take a taxi, or to use a company vehicle. The characteristics of the trip chain are described by the dummy variable Chain_{sub} and Chain_{Nenforce} in the model. Dummy variables PV_{ih_car}, PV_{ih_motorcycle}, and PV_{ih_bicycle} represent the private vehicles where the driver i is enforced. The choice results are regarded as the mode choices of all trips in the subchains or home-based trip chain. Then the mode choice of all trips of the private vehicle drivers in a household are decided.

In sum, the detailed procedures to determine the mode choice of a driver are as follows:

1. Take PV_{ih} as the travel mode for the longest trip of this driver. Next, value the home-based trips that form a closed chain with the enforced trip by PV_{ih} according to the constraints on private vehicle usage.
2. Decide the mode choice for the unenforced subchain-home-based chain belonging to the anchor point.

Step 3. Search for Vehicle Sharing in Household

Ridesharing in Household

If joint and escort activity exist in a household, ridesharing should be considered. In midsize cities in China, motorcycles and bicycles commonly have passengers. In this paper, because the activity schedules of all the members of the household are known, the joint and escort activities can be found. There is an assumption that only the individuals without private vehicles allocated can be passengers. The activity schedules of the passengers are matched with the drivers' schedules, and the mode choices of the corresponding trips are set as passengers.

Whether Assigned Vehicle Can Be Used by Other Household Member When It Is Not Occupied

Further vehicle sharing is investigated to decide whether the household members who are not allocated private vehicles in Step 1 can use the private vehicle when it is free at home. The sequence of the

vehicle-sharing searching in this step is car, motorcycle, and bicycle. Here, the car is used as an example.

The research of this case requires two conditions to be satisfied in a household h : (a) the household member eligible for the car use whose PV_{ih} equals zero exists, and (b) the individual qualifying for the first condition has a home-based trip chain when this car is free at home and the mode choices for all the trips of this trip chain are not decided. If a household member meets these two requirements, a binary logit model is used to decide to use or not to use the car for this tour. When more than one household member meets the two conditions, a random household member is chosen for the binary logit model. Subsequently, if someone in this household qualifies to use the car, further allocation is operated; otherwise, the allocation of other vehicles in this household is processed. The personal attributes and characteristics of the home-based trip chain are the important factors.

Step 4. Mode Choice of Household Members Not Allocated a Vehicle

The fourth step is to study the mode choices of household members who are not allocated private vehicles (PV_{ih} = 0). This type of household member makes choices from four alternatives: walk, transit, taxi, and a company vehicle for all of his or her trips, except those trips for which the mode has been decided in Step 3.

In this step, mode choice considers the effect of the previous trip's mode choice on the subsequent one. For every choice decision, an individual is assumed to select the mode that maximizes his or her utility of the trip. In terms of the chronological sequence, mode decision is repeated for this individual i trip-by-trip until all mode choices for his or her trips are acquired.

An order mixed logit model that captures the effect of the mode chosen in the previous trip on the current trip can be expressed as follows. The total utility of an alternative travel mode m for trip t_{ih} can then be expressed as

$$U_{t_{ih}}^m = V_{t_{ih}}^m + \sum_m \theta_{m'm} \sigma_{t_{j-1,ih}}^{m'} + \eta_{t_{ih}}^m + \xi_{t_{ih}}^m \quad (3)$$

where

$U_{t_{ih}}^m$ = utility for individual i in household h for trip t_j and mode m ;

$V_{t_{ih}}^m$ = deterministic component of utility for trip t_j of individual i and mode m , varying systematically as a function of the household characteristics, individual attributes, trip characteristics, and attributes of the modes;

m' = mode choice result of the previous trip, t_{j-1} ;

$\sigma_{t_{j-1,ih}}^{m'}$ = 1 if individual i made trip t_{j-1} with mode m' ; otherwise, it is equal to 0;

$\theta_{m'm}$ = parameters for the effect of travel mode m' of the last trip, t_{j-1} , on the mode choice of m on current trip t_j ; and

$\eta_{t_{ih}}^m$ and $\xi_{t_{ih}}^m$ = mixed logit error components used in this step to account for variability across users and correlation across error terms over repeated instances. The detailed meanings are presented in Table 1.

A maximum simulated likelihood estimation technique is used to solve this mode. The detailed algorithms and procedures can be found in work by Ramadurai and Srinivasan (3).

CASE STUDY

Data

The travel diary survey used for this case study covered 1 day, was conducted through home interviews, and was conducted at the household level. The data were collected in spring 2007 in Bengbu, with a sampling rate of approximately 3%. In the survey, all trips made by all household members (above 6 years old) from 5,478 households were recorded on a randomly selected weekday, along with household attributes, individual socioeconomic information, activity, and trip characteristics.

Rigorous data screening excluded records with inconsistent and incomplete data. The cleaned activity travel household samples used in this study produced 3,842 households. The model was estimated by 2,842 households randomly selected in these samples, and the remaining 1,000 households were used for validation testing.

Two hundred fifty-three households from the 2,842 estimation samples hold one car, and none of households have more than one car. Among the 253 households, 241 households have only one person with a driver's license. Twelve households have more than one member with a driver's license. This car ownership distribution is typical in a midsize Chinese city. Five hundred seventy-six households own a motorcycle, but less than 1% of these households have two motorcycles. Among the 1,922 households that

own bicycles, 579 households have more than two bicycles. The market shares of the travel mode of the estimation data are as follows: car (2.74%), motorcycle (5.11%), bicycle (23.25%), walk (33.68%), transit (26.39%), taxi (1.15%), company car (6.27%), and passenger (1.37%).

Estimation Results of Model System

For the estimation, the parameters of each model are calibrated step-by-step with the survey data.

Vehicle Allocation Modeling Results

The number of qualified samples for the binary logit model with respect to car, motorcycle, and bicycle is 241, 372, and 1,063, respectively. Because of the small sample size (12 households), the car allocation model between two or more eligible members in a household is not estimated. The parameters of mixed nested logit models for motorcycle and bicycle allocation are estimated by 211 and 1,438 samples, respectively.

Tables 2 and 3 present the estimation results of the allocation models for car, motorcycle, and bicycle. Variables with correlation of more than 0.3 are excluded from the specification.

TABLE 2 Vehicle Allocation Results: Binary Logit Model—Case 1

Variable	Car Coeff. (<i>t</i> -Stat.)	Motorcycle Coeff. (<i>t</i> -Stat.)	Bicycle Coeff. (<i>t</i> -Stat.)
Constant	1.788 (21.216)	1.119 (13.202)	0.345 (4.710)
Household Attributes			
Location: in the center of city			-0.145 (-3.965)
Number of household members > 2		1.477 (9.027)	
Number of motorcycles > 0			
Number of bicycles > 0		-1.313 (-7.237)	
Number of children > 0			0.185 (4.589)
High income (>50,000 yuan)			-0.231 (-8.243)
Personal Attributes			
Male	2.768 (20.567)	2.225 (40.233)	0.693 (22.546)
Age > 60			-0.528 (-12.317)
Company car available	-4.708 (-11.996)	-2.038 (-4.372)	
Education above university	0.887 (5.619)	1.019 (14.925)	-0.771 (-17.399)
Student			0.893 (2.109)
Trip chain > 1	-0.766 (-3.783)	3.846 (6.354)	1.235 (20.814)
Employee		1.274 (3.245)	0.817 (5.724)
Employer	2.314 (13.726)	1.679 (2.099)	
Have escort activity			0.908 (2.345)
Transit card hold		-1.019 (-7.345)	-1.639 (-9.814)
Distance of longest trip > 2 km			1.535 (7.893)
Distance of longest trip > 5 km	2.356 (4.128)	1.356 (4.128)	
Free parking	3.136 (7.893)		
Number of observations	241	372	1,063
L(0)	-166.7	-207.9	-735.7
L(<i>b</i>)	-58.57	-79.42	-309.9
Model fit (p^2)	0.649	0.618	0.579

NOTE: Coeff. = coefficient; *t*-stat = *t*-statistic. Of the three binary logit models, "not to use this vehicle" is set as reference alternative.

TABLE 3 Vehicle Allocation Results: Mixed Nested Logit Model—Case 2

Variable	Motorcycle Coeff. (t-Stat.)	Bicycle Coeff. (t-Stat.)
Household Role		
Head of the household	3.124 (7.892)	
Child of head of the household		1.270 (3.152)
Personal Attributes		
Male	1.709 (24.512)	0.982 (12.347)
Youth		0.783 (2.341)
Middle age		1.341 (4.213)
Age > 60 years		-1.524 (-6.210)
Company car available	-3.125 (-6.235)	-2.123 (-5.214)
Education above university * female ^a	0.791 (3.780)	
Distance from home to school > 2 km * student ^a		2.535 (10.091)
Work location is in the center of city * employee ^a		-0.784 (-2.178)
Have commute trip	0.654 (2.139)	0.817 (5.724)
Have multistop tour		1.508 (3.116)
Transit card hold * female ^a		-1.731 (-6.418)
Distance of longest trip > 2 km		1.535 (7.893)
Distance of longest trip > 5 km	2.231 (7.585)	
Not Being Used		
Constant	-2.456 (-6.805)	-0.782 (-2.124)
Number of workers in household	-0.523 (-3.173)	-0.213 (-2.092)
Distance from home to bus stop < 0.5 km		1.091 (1.974)
Low income household (< 50,000 yuan)		-1.127 (-9.134)
Home located in center of city		0.819 (2.871)
Number of observations	211	1,438
L(0)	-249.2	-1,919.6
L(b)	-107.6	-1,097.8
Model fit (ρ^2)	0.568	0.428

^aIndicates a combined variable, when both variables listed must be satisfied at the same time.

Tables 2 and 3 show the following:

- Of these five models, all parameters have expected signs.
- The overall goodness of fit of three binary logit models for car, motorcycle, and bicycle are 0.649, 0.618, and 0.579, respectively. The explanatory functions of the variables of car applied in these models are better than those for motorcycle, and the motorcycle functions are better than those for bicycle. This trend can be explained because most possession of cars and motorcycles exists only if there is demand in developing countries. However, it is common to have a bicycle in a household, even if it is not often used, making usage of bicycles more complicated.
- The overall goodness of fit in the nested logit models of motorcycles and bicycles is 0.568 and 0.428, respectively, indicating that the explanatory variables could well capture the allocation characteristics of motorcycles and bicycles in the household. The car, as a symbol of social status and wealth in a midsize city in China, is usually charged by a certain household member.
- The hypothesis of independence of error structure of the nested logit models for motorcycles across household members and over repeated allocations cannot be rejected [$LL_{i.i.d.model} = -107.558$, $LL_{mixed\ error\ model} = -102.917$, χ^2 (actual) = 4.614 < $\chi^2_{0.98}$ (critical, 4 d.f) = 9.488]. This result is partially because of a lack of repeated observations of motorcycle allocation in the same household. However, the proposed error structure provides a significant improvement in

model fit for the bicycle allocation model [$LL_{i.i.d.model} = -1145.326$, $LL_{mixed\ error\ model} = -1097.763$, χ^2 (actual) = 47.563 > $\chi^2_{0.98}$ (critical, 4 d.f) = 9.488].

Mode Choice of People with a Private Vehicle Allocated

After the enforcement regulation is applied to the individual allocated a vehicle whose PV_{ih} is not equal to 0, 1,247 samples for the estimation of mode choice model for the anchor point are obtained. To determine the mode choice result for the subchain or “not enforced chain” of the samples, the following hierarchy from top to bottom is used: private vehicle company car, taxi, transit, and walk. After estimation, the results of the MNL model are shown in Table 4, and the alternative “to continue using PV_{ih} ” is set as reference alternative.

The estimation result of MNL fits the data quite well, with ρ^2 being 0.467. However, the use of different private vehicles would present different mode-switching properties within a subchain or “not enforced chain” that is demonstrated by the model parameters. For instance, it is harder to make a switch between motorcycles and transit than between cars and transit (the coefficients of PV_{ih_car} and $PV_{ih_motorcycle}$ for the transit alternatives are -2.278 and -3.179, respectively).

TABLE 4 Estimated Result of Mode Choice MNL Model for Anchor Point

Variable	Interaction with Indicator Variable	Coefficient	t-Statistic
Constant (W)		-2.173	-18.213
Constant (TR)		-3.783	-8.946
Constant (TA)		-4.127	-4.761
Constant (CV)		-2.751	-7.721
PV _{ih} _car (TR)	Chain _{sub}	-2.278	-4.504
PV _{ih} _car * for work activity (CV) ^a	Chain _{sub}	3.421	6.810
PV _{ih} _car * for work activity (TA) ^a	Chain _{sub}	0.872	2.454
PV _{ih} _motorcycle (TR)	Chain _{sub}	-3.179	-12.896
PV _{ih} _bicycle * for shopping activity (W) ^a	Chain _{sub}	2.245	1.879
PV _{ih} _bicycle * for shopping activity (CV) ^a	Chain _{sub}	-3.115	-6.168
PV _{ih} _bicycle (TA)	Chain _{sub}	-0.741	-5.672
Female (W)	Chain _{sub}	1.567	2.177
Female (CV)	Chain _{sub}	-4.309	32.187
Multistop tour (TR)	Chain _{sub}	0.563	1.982
Recreation activity (CV)	Chain _{sub}	-5.121	-17.891
Anchor point in center of city (W)	Chain _{sub}	0.991	2.109
PV _{ih} _bicycle * older (W) ^a	Chain _{sub}	0.325	2.011
PV _{ih} _car (TR)	Chain _{Nonforce}	-2.278	-4.504
PV _{ih} _motorcycle (TR)	Chain _{Nonforce}	-3.179	-12.896
PV _{ih} _motorcycle * escort activity (W) ^a	Chain _{Nonforce}	-2.102	-4.116
PV _{ih} _bicycle * distance of longest trip > 5 km (TR) ^a	Chain _{Nonforce}	3.126	11.12
Recreation activity (W)	Chain _{Nonforce}	2.245	3.879
House located in suburban (TR)	Chain _{Nonforce}	-0.763	-2.106
Transit card (TR)	Chain _{Nonforce}	1.103	3.003
PV _{ih} _bicycle * older (W) ^a	Chain _{Nonforce}	0.69	6.908
PV _{ih} _car * for work activity (TA) ^a	Chain _{Nonforce}	0.296	1.793
PV _{ih} _motorcycle * for medical activity (taxi) ^a		0.162	2.712
PV _{ih} _bicycle * for medical activity (taxi) ^a		0.273	7.912
PV _{ih} _bicycle * distance of the longest trip (TR) ^a		0.038	9.987
Distance of the longest trip (W)		-0.212	-10.321
Student (w)		1.127	1.806
Number of observations		1,247	
L(0)		-2,007.9	
L(b)		-1,067.4	
Model fit (ρ ²)		0.467	

NOTE: W = walk; TR = transit; TA = taxi; CV = company vehicle. For individual *i*, PV_{ih} is equal to car note as PV_{ih}_car; PV_{ih} is equal to motorcycle note as PV_{ih}_motorcycle; PV_{ih} is equal to bicycle note as PV_{ih}_bicycle.
^a Indicates a combined variable, when both variables listed must be satisfied at the same time.

Issue of Shared Use Vehicles in a Household

Each of 513 individuals from 2,842 households makes at least one trip as a passenger in the survey day. The number of households having two drivers using only one vehicle on the same day is 13, which is less than 1% of 2,842 samples. More than 90% of private vehicles are used for commuting in midsize Chinese cities; therefore, it is difficult for other household members to use the car within normal work hours. Meanwhile, joint activities, such as shopping and recreation after work, may restrict the cases in which the vehicle is driven by other household members. In this study, the model for vehicle sharing is not calibrated because of the small sample size.

Mode Choice of People Without a Private Vehicle Allocated

The number of trips by household members who are not allocated a private vehicle are 20,116 after trips are subtracted that have mode choice of passenger or shared use of private vehicle. The 20,116 trips are estimated as samples of the mode choice model as described in the discussion of Step 4. Calibration results are listed in Table 5, with walk as the reference mode.

Inertial effects exist between the trips even without the use of a private vehicle: individuals are highly likely to choose a mode they have previously chosen. This result signifies an inherent rigidity among individuals against changing modes, especially for the tran-

TABLE 5 Estimated Result of Mixed Logit Model for People Without Private Vehicle Allocated

Variable	Transit		Taxi		Company Car	
	Coefficient	<i>t</i> -Statistic	Coefficient	<i>t</i> -Statistic	Coefficient	<i>t</i> -Statistic
Constant	-0.352	-5.471	-3.261	-11.382	-2.98	-7.891
Effect of Previous Mode						
Previous mode is walk			-1.239	-7.127		
Previous mode is transit * home-based trip ^a	1.135	11.783				
Previous mode is company car * home-based trip ^a					-0.712	-7.129
Previous mode is transit * not home-based trip ^a	2.269	13.011				
Previous mode is taxi * not home-based trip ^a			0.726	5.127		
Previous mode is company car * not home-based trip ^a					2.123	19.312
Previous mode is passenger * not home-based trip ^a	0.672	7.109	1.271	4.213		
Have mode shift before this trip	0.241	3.128	-0.149	-1.978		
Effect of Trip Characteristics						
Work correlation trip			0.927	3.126	3.863	5.761
Shopping activity trip			-2.763	-4.109	-7.638	-21.128
School activity trip * (kid, youth) ^a	0.126	2.163			-3.671	-6.626
Escort activity trip						
Recreation activity trip * older ^a	-1.109	-3.218	-3.258	-3.129	-4.237	-8.172
Recreation activity * middle age ^a	0.679	2.109				
Joint activity	-1.765	-4.213				
Medical activity	1.378	3.11	3.318	11.056		
Trip distance (km)	0.173	4.271	0.029	3.127		
Trip destination is suburban					2.781	4.318
Origins near bus stop in 500 m	0.763	3.187				
Money cost (yuan)	-0.731	-2.179	-0.137	-3.115		
Travel in peak hour			-0.531	-1.976		
Company car available			-3.117	-5.671	4.192	21.234
Effect of Personal Attributes						
Transit card holder	2.172	1.673				
Male	-0.598	-3.561	0.387	2.196		
In high-income household			2.234	4.213		
Number of observations			20,116			
L(0)			-25,899.8			
L(b)			-14,374.4			
Model fit (ρ^2)				0.445		

^aIndicates a combined variable, when both variables listed must be satisfied at the same time.

sit and walk modes. The goodness of fit of this model is 0.445; it fits quite well. The trip characteristics and personal attributes also show the apparent influences on the mode choices. All the terms in the proposed error structure were statistically significant. The empirical results imply the strong influence of unobserved preference heterogeneity, heteroskedasticity across alternatives, and serial correlation.

Model Validation

Validation of this model system was tested by a sample of 1,000 households that were randomly selected and set aside; there were 3,040 individuals and 8,664 trips in the sample. The market share of mode choice of the validation data is as follows: car (2.19%), motor-

cycle (5.23%), bicycle (22.79%), walk (34.37%), transit (27.29%), taxi (1.01%), company car (5.17%), and passenger (1.97%).

The agent-based, four-step mode choice modeling system built in this paper is implemented by Java software, and the results obtained in previous steps were embedded into the current step as input. In the second step, if the $\#Ih_car$ is larger than $\#Car_h$ of household h , the researchers randomly allocate the car to one of eligible members. In the third step, the case in which the household member shares a private vehicle is intentionally ignored. Vehicle allocation and mode choice results for the model validation test are presented in Tables 6 and 7.

Table 6 shows that more than 95% of the vehicle use prediction results are correct, and Table 7 shows that more than 88% of observed modes are chosen on average. Table 7 suggests that the model provides reasonable prediction ability for most alternatives (except taxi).

TABLE 6 Prediction Results for Validation Test, Vehicle Allocation (3,040 People)

Observed Result	Predicted Results (%)				
	Car	Motorcycle	Bicycle	No Use	Total
Car (2.69%, 82)	96.34	1.22	1.22	1.22	100.00
Motorcycle (5.74%, 175)	0.00	96.00	1.71	2.29	100.00
Bicycle (21.75%, 661)	0.00	0.61	92.28	7.11	100.00
No use (69.82%, 2,122)	0.09	0.05	3.49	96.37	100.00
Aggregate predicted shares	2.66	5.72	22.63	68.98	100.00

Three observations from these validation results are noteworthy. First, the prediction of the vehicle allocation model, which starts from cars and goes to motorcycles and then jumps to bicycles, is reasonable. The step-to-step simulation process improves the accuracy of the mode choice modeling and effectively confirms the use of the private vehicle within households. Second, household members are classified into two types: members allocated with private vehicles and members without, a classification that captures the different probabilities of mode choice changes of these two types of individuals and increases the accuracy of prediction of the model. Third, the prediction of taxi mode is poor, which may be attributed to the small sample size.

CONCLUSIONS AND FUTURE EXTENSIONS

This paper presents an agent-based, hybrid, four-step, same-day mode choice model. In this model, rules-based algorithms are combined with classical random utility maximization decision criterion within an explicit microsimulation framework to estimate mode choice at the household level in developing countries. Step 1 simulates vehicle (car, motorcycle, and bicycle) allocation within households by a binary logit model and a nested logit model. Step 2 combines a rule-based model and a discrete mode choice model to decide the mode choices of vehicle users. Step 3 simulates interactions among household members sharing vehicles. Step 4 applies a mixed MNL model to simulate the mode choices of individuals without private vehicles on a trip-by-trip basis. The effect of the previous mode on subsequent trips is considered in this model.

The benefits of the proposed model come from its abilities to model vehicle deficiency and the differences of mode choice behaviors among household members in developing countries. Through the addition of the vehicle allocation before mode choice predictions, the assumption that vehicles are always available is relaxed. The adaptability of the model system is improved by including the allocation of car, motorcycle, and bicycle according to their different functionalities in China. The discrepancies between the mode choices of household members with private vehicles and those without vehicles are solved by separately modeling this mode choice behavior with different models.

The proposed model has some limitations. First, the model needs the prediction from multiple models and a large sample set to support the mode estimation in the four steps. Second, the step-by-step framework requires a high level of accuracy in previous steps to reduce the deviation of the misfit transfer. However, given that the validation test shows the ability to model all trips made by all household members in an internally self-consistent manner from a midsize city in China, these costs are well worth paying.

There is still work to do in the conceptual and operational elaboration of this model. Because of a single set of limited data, this study did not conduct the calibration and test of car allocation and vehicle sharing submode. The validation of the submode in the model is needed, along with a study on whether the model fits the complicated transportation environments of larger cities. The proposed framework can be integrated with an activity generation model in future research to obtain additional insights about activity and travel behaviors.

TABLE 7 Prediction Results for Validation Test, Mode Choice (8,664 Trips)

Observed Result	Predicted Results (%)								
	Car	Motorcycle	Bicycle	Walk	Transit	Taxi	Company Car	Passenger	Total
Car (2.19%, 190)	95.79	0.00	0.53	1.58	0.00	0.53	1.58	0.00	100.00
Motorcycle (5.23%, 453)	0.00	94.49	3.08	1.32	0.22	0.66	0.22	0.00	100.00
Bicycle (22.79%, 1,974)	0.00	0.00	90.08	4.05	5.27	0.00	0.61	0.00	100.00
Walk (34.37%, 2,978)	0.00	0.30	2.99	89.46	5.74	0.10	0.84	0.57	100.00
Transit (27.29%, 2,364)	0.13	0.72	2.49	8.79	87.02	0.42	0.21	0.21	100.00
Taxi (1.01%, 86)	5.75	6.90	1.15	6.90	4.60	57.47	17.24	0.00	100.00
Company car (5.17%, 448)	0.45	1.56	2.68	7.14	5.58	1.34	78.35	2.90	100.00
Passenger (1.97%, 171)	0.00	0.00	1.75	2.34	5.26	3.51	1.17	85.96	100.00
Aggregate predicted shares	2.21	5.40	22.58	34.66	27.36	0.91	4.78	2.10	100.00

ACKNOWLEDGMENTS

The project is supported by grants from the Major Program of National Natural Science Foundation of China, the Program of National Natural Science Foundation of China, and the Major State Basic Research Development Program.

REFERENCES

1. Miller, E. J., M. J. Roorda, and J. A. Carrasco. A Tour-Based Model of Travel Mode Choice. *Transportation*, Vol. 32, No. 4, 2005, pp. 399–422.
2. Fosgerau, M. PETRA—An Activity-Based Approach to Travel Demand Analysis. In *National Transport Models: Recent Developments and Prospects* (L. Lundqvist and L.-G. Mattsson, eds.), Springer, Stockholm, Sweden, 2002.
3. Ramadurai, G., and K. K. Srinivasan. Dynamics and Variability in Within-Day Mode Choice Decisions: Role of State Dependence, Habit Persistence, and Unobserved Heterogeneity. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1977*, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 43–52.
4. Fujiwara, A., and J. Zhang. Development of a Scheduling Model for Car Tourists' 1-Day Tours. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1921*, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 100–111.
5. Bradley, M., M. Outwater, N. Jonnalagadda, and E. Ruiter. Estimation of an Activity-Based Microsimulation Model for San Francisco. Presented at 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.
6. Dissanayake, D., and T. Morikawa. Household Travel Behavior in Developing Countries: Nested Logit Model of Vehicle Ownership, Mode Choice, and Trip Chaining. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1805*, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 45–52.
7. Bowman, J. L., and M. E. Ben-Akiva. Activity-Based Disaggregate Travel Demand Model System with Activity Schedules. *Transportation Research Part A*, Vol. 35, No. 1, 2001, pp. 1–28.
8. Vovsha, P., J. Freedman, and M. Bradley. Tour-Based Mode Choice Modeling Techniques: US Practices. In *Proc., European Transport Conference, 2008*. Parsons Brinckerhoff, Inc., and MB Research and Consulting, Noordwijk, Netherlands, 2008.
9. Gunn, H. The Netherlands National Model: A Review of Seven Years of Application. *International Transactions in Operational Research*, Vol. 1, No. 2, 1994, pp. 125–133.
10. Biagioni, J. P., P. M. Szczurek, P. C. Nelson, and A. Mohammadian. Tour-Based Mode Choice Modeling: Using an Ensemble of Conditional and Unconditional Data Mining Classifiers. Presented at 88th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.
11. Daly, A. J., H. H. P. van Zwam, and J. van der Valk. Application of Disaggregate Models for a Regional Transport Study in the Netherlands. Presented at World Conference on Transport Research, Hamburg, Germany, 1983.
12. Ye, X., R. M. Pendyala, and G. Gottardi. An Exploration of the Relationship Between Mode Choice and Complexity of Trip Chaining Patterns. *Transportation Research Part B*, Vol. 41, No. 1, 2007, pp. 96–113.
13. Bowman, J. L. *Activity Based Travel Demand Model System with Daily Activity Schedules*. MS thesis. Massachusetts Institute of Technology, Cambridge, 1995.
14. Bowman, J. L. *The Day Activity Schedule Approach to Travel Demand Analysis*. PhD dissertation. Massachusetts Institute of Technology, Cambridge, 1998.
15. Timmermans, H. J. P., and J. Y. Zhang. Modeling Household Activity Travel Behavior: Examples of State of the Art Modeling Approaches and Research Agenda. *Transportation Research Part B*, Vol. 43, No. 2, 2009, pp. 187–190.
16. Zhang, J. Y., M. Kuwano, B. Lee, and A. Fujiwara. Modeling Household Discrete Choice Behavior Incorporating Heterogeneous Group Decision-Making Mechanisms. *Transportation Research Part B*, Vol. 43, No. 2, 2009, pp. 230–250.
17. Anggraini, R., T. A. Arentze, and H. J. P. Timmermans. Modeling Car Allocation Decisions in Automobile Deficient Households. Presented at European Transport Conference, Netherlands, 2008.
18. Srinivasan, K. K., G. N. Pradhan, and M. Naidu. Commute Mode Choice in a Developing Country: Role of Subjective Factors and Variations in Responsiveness Across Captive, Semicaptive, and Choice Segments. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2038*, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 53–61.
19. Rajagopalan, B. S., and K. K. Srinivasan. Integrating Household-Level Mode Choice and Modal Expenditure Decisions in a Developing Country: Multiple Discrete–Continuous Extreme Value Model. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2076*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 41–51.
20. Yagi, S., and A. Mohammadian. Joint Models of Home-Based Tour Mode and Destination Choices: Applications to a Developing Country. In *Transportation Research Record: Journal of the Transportation Research Board, No. 2076*, Transportation Research Board of the National Academies, Washington, D.C., 2008, pp. 29–40.
21. Krygsman, S., T. Arentze, and H. Timmermans. Capturing Tour Mode and Activity Choice Interdependencies: A Co-Evolutionary Logit Modelling Approach. *Transportation Research Part A*, Vol. 41, No. 10, 2007, pp. 913–933.
22. Train, K. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge, United Kingdom, 2003.

The Transportation in the Developing Countries Committee peer-reviewed this paper.