Modeling Household Energy Consumption Behaviors by Considering Self-selection Effect and Heterogeneity

Biying YU[1], Junyi ZHANG [2] and Akimasa FUJIWARA[3]

[1] Doctor Candidate, Graduate School for International Development and Cooperation, Hiroshima University
(1-5-1 Higashi Hiroshima, Hiroshima 739-8529, Japan)
E-mail: yubiyingjapan@yahoo.co.jp

[2]Member of JSCE, Graduate School for International Development and Cooperation, Hiroshima University
(1-5-1 Higashi Hiroshima, Hiroshima 739-8529, Japan)
E-mail: zjy@hiroshima-u.ac.jp

[3] Member of JSCE, Graduate School for International Development and Cooperation, Hiroshima University
(1-5-1 Higashi Hiroshima, Hiroshima 739-8529, Japan)
E-mail: afujiw@hiroshima-u.ac.jp

Describing the household energy consumption pattern by considering the interaction between residential environment and energy consumption behaviors (e.g., ownership and usage of varied end-uses) is of much interest due to the existence of self-selection effect. Furthermore, because of the household budget or time constraint, it is expected that in-home energy consumption behavior and out-of-home energy consumption behavior (e.g., ownership and usage of vehicles) might be correlated with each other. These concerns motivate us to develop an integrated model of residential location choice and household energy consumption behaviors referring to ownership and usage of end-uses, which accounts for both in-home and out-of-home sectors. For the purpose of analysis, a household energy consumption survey which includes the information about households' energy consumption, ownership/usage of in-home appliances and vehicles, household and individual attributes, together with the built environment characteristics was conducted in Beijing in 2010, and finally 775 valid samples are collected. Based on the rich data, this paper presents a joint mixed Multinomial Logit-Multiple Discrete-Continuous Extreme Value (MNL–MDCEV) model of residential location choice, end-use ownership, and usage. Estimation results indicate that there is significant dependency among the choice dimensions and that self-selection effects as well as unobserved heterogeneity cannot be ignored when modeling residential environment and household energy consumption behavior interactions.

Key Words : *Household energy consumption behavior, residential location choice, Self-selection effect, heterogeneity, MNL-MDCEV*

1. INTRODUCTION

Transportation engineers and planners have assumed for several decades that there is an association between land-use development patterns and the travel behavior of individuals. This assumption has been confirmed in a lot of ways. Due to the truth that travel behavior is one part of household energy consumption behavior and it implicitly influence in-home energy use, therefore, it is reasonable to infer that land use patterns might play a role on household energy consumption behavior, motivated by the possibility that design policies associated with the built environment can be used to control, manage, and shape individual or household's behavior and aggregate energy demand. However, almost all analyses related with land-use and energy consumption recognize and control built environment attributes as direct explanatory factors of energy consumption behavior. While the observed and unobserved factors which make households select themselves to a specific residential environment and energy consumption pattern are not examined. Consequently, this study sheds light on exploring the interaction of household spatial location choice and household energy consumption behavior by incorporating the self-selection effect. An integrated model system termed mixed Multinomial Logit-Multiple Discrete-Continuous Extreme Value (MNL–MDCEV) model which covers residential location choice, end-use (including in-home appliances and out-of-home vehicles) ownership, and usage behavior, is presented here to identify the sensitivity of household energy consumption to changes in land use policy by considering a comprehensive set of built environment (BE) variables, socio-demographic variables as well as unobserved factors.

Essentially, the interaction between residential location and energy consumption behaviors occurs due to the existence of self-selection effect. In statistics, self-selection arises in any situation in which individuals select themselves into a group, causing a biased sample with nonprobability sampling. It is commonly used to describe situations where the unique characteristics of the people which make them to select themselves into the group which creates abnormal or undesirable conditions in the group. Self-selection generally results from two sources: attitudes and socio-demographic traits. An example of attitude-induced self-selection is the environmental awareness. Households who are environment friendly are more likely to choose living in the area with high street block density or with convenient public transit so as to reduce the possibility of owning and using private vehicle. Plus, energy-saving end-uses as well as efficient using style would be preferred in these households. With respect to the socio-demographic traits, it is easy to understand that low income households may choose to live in neighborhoods with ample transit service and hence not buy and use cars. In this case, it is not good transit facilities but households' economic constraints that have a true and direct influence on their choice of car ownership and usage. Therefore, it is very essential to consider the self-selection effect so as to describe household energy consumption behavior more accurately.

For the purpose of analysis, a household energy consumption survey which includes the information about households' energy consumption, ownership/usage of in-home appliances and vehicles, household and individual attributes, together with the built environment characteristics was conducted in Beijing in 2010, and finally 775 valid samples are collected. Based on the rich data, the integrated model is built to comprehensively understand household energy consumption behavior. This methodology may be characterized as follows: First, it explicitly considers and models the residential location choice decision jointly with the household energy consumption behavior choice (i.e., the ownership and usage of varied end-uses), which represents the correlation between these two. Such an integrated model provides a valuable tool for policy analysis, since it can predict how residential choices would change due to urban form design policies as well as estimate the household energy consumption behavior change. Second, both in-home energy consumption behavior (e.g., ownership and usage of electric appliances) and out-of-home energy consumption behavior (e.g., ownership and usage of vehicles) are included in this integrated model subject to the household budget constraint. Third, unobserved heterogeneity (i.e., sensitivity variations due to unobserved household/individual factors) are considered in residential location choice as well as household energy consumption behavior choice. Fourth, this methodology can easily control the self-selection effects caused by both social-demographic and unobserved attributes of households.

2. MODELING METHODOLOGY

The methodology adopted in this paper is enlightened from the model structure proposed by Pinjari et al.(2009) and the details are presented as follows.

(1) Residential location choice

Let i $(i = 1,2,\dots,I)$ denote the index for the households, j $(j = 1,2,\dots,J)$ denote the index for the spatial unit of residential choice. The residential location choice component is described by the familiar discrete choice formulation:

$$u_{ij} = \sum_{p} (c_p + \theta_p^{'}\pi_{ip})\, z_{ijp} + v_{ij} + \sum\nolimits_{k} \omega_{ijk} + \epsilon_{ij} \tag{1}$$

where, u_{ij} is the utility that the ith household chooses to settle in spatial unit j. z_{ij} is a set of built environment (BE) attributes associated with household i's decision on residence (such as land-use mix and activity accessibility). c_p is the pure influential effect on the residential choice behavior solely caused by the BE attribute p, and π_{ip} is a vector of observed household social-demographic characteristics (e.g., household income, household size and presence of children in the household) which make the household self-select to locate or not to locate in the spatial unit with the pth BE attribute in z_{ij}. v_{ij} and $\sum_k \omega_{ijk}$ are unobserved factors impacting household i's residential location choice which are used to represent households' heterogeneity. v_{ij} only contains those household-specific unobserved factors that influence sensitivity to residential choice, while ω_{ijk} includes only those household-specific unobserved factors that impact both residential choice and end-use k's ownership and usage. It is assumed v_{ij} follow normal distribution with a mean 0 and variance σ_v. ϵ_{ij} is an error term following an independently and identically Gumbel distribution.

(2) Energy consumption behavior choice

Assume that there are K different end-uses that a household can potentially allocate its money to. Let e_k be the expenditure consumption of end-use k $(k = 1,2,\cdots,K)$. If an outside goods which is always consumed is present, label it as the first goods with a unit price of one (see Bhat, 2008). In this study, the money derived from income deducting the energy expenditure is regarded as the outside goods, which is termed as disposal money. The utility that household i obtains from energy consumption is specified as the sum of the utilities got from spending money on each end-use, as shown below.

$$U_{ij}(e_k) = \frac{1}{\alpha_1}\exp(\xi_{ij1})e_{i1}^{\alpha_1} + \sum_{k=2}^{K}\frac{\gamma_k}{\alpha_k}\exp\left(\mu_k^{'}s_{ij} + \Delta_k^{'}x_i + \eta_{ijk} \pm \omega_{ijk} + \tau_{ijk}\right)\left\{\left(\frac{e_{ik}}{\gamma_k}+1\right)^{\alpha_k} - 1\right\} \quad (2)$$

Here, $U_{ij}(e_k)$ is the total utility derived from allocating a non-negative amount of the total budget to each end-use k, including savings. With the above utility function, it is assumed that a household maximizes its utility subject to its budget constraint that $\sum_{k=1}^{K} e_k = E$, where E is the total budget (e.g., expenditure, disposal income, or available time). As a result, the linearly competitive relationship among end-uses is reflected in the model. Note that only one type of budget constraints can be represented. This study only deals with household monetary budget constraint. φ_{ijk} is the baseline utility for money spent on end-use k which controls the discrete choice ownership decision in end-use k for household i living in spatial unit j. The parameter α_k and γ_k are the parameters to represent different levels of satiation effects or the shape of utility function. s_{ij} is a vector of built environment (BE) attributes with the corresponding coefficient $\mu_k^{'}$. π_{ip} is a set of observed household social-demographic characteristics and $\Delta_k^{'}$ is the coefficient vector. The η_{ijk} and ω_{ijk} are unobserved factors impacting household i's energy consumption behavior which is used to represent households' heterogeneity. η_{ijk} only explains those household-specific and end-use specific unobserved factors that influence household energy consumption behavior, while ω_{ijk} is the common components affecting the residential location choice and energy consumption behavior of end-use k. It is assumed that η_{ijk} and ω_{ijk} are all normally distributed with a mean 0 and variance $\sigma_{\eta k}$, $\sigma_{\omega k}$, respectively. Furthermore, the correlation among η_{ijk} (*k=2,3,...K*) is considered in this study and so does ω_{ijk} (*k=2,3,...K*). The error term τ_{ijk} is independently and identically Gumbel distributed. For identification, the baseline utility of the outside goods is denoted as $\exp(\xi_{ij1})$ because only utility differences matter.

(3) The integrated choice model

Join the previous two models together so as to comprehensively look on the residential location and household energy consumption behavior choice.

It is supposed that households choose a residential location j and energy consumption pattern referring to the ownership and usage of varied end-uses by making u_{ij} and $U_{ij}(e_k)$ jointly maximal subject to the income budget constraint. Due to the presence of the common component ω_{ijk} in the residential choice and energy consumption behavior model, the integrated model system arises by joining these two parts together. The "$\pm$" sign in front of ω_{ijk} term in the energy consumption behavior model means that the unobserved factors relating to residential location choice has a positive (+) or negative (-) effect on the ownership and usage of end-use k. In this integrated model, the self-selection effects caused by both observed and unobserved attributes are included. The integrated residential location and household energy consumption behavior choice probability can be derived by multiplying the probabilities of the two choice components.

Denote Γ as a vector that contains all the parameters to be estimated (i.e., c_p, $\theta_p^{'}$, $\mu_k^{'}$, $\Delta_k^{'}$, γ_k(*k=2,3,...K*), α_1, and the variances of the stochastic components v_{ij}, η_{ijk}, and ω_{ijk}), and $\Gamma_{-\sigma}$ as a vector of all parameters except the variance terms. Plus, let Λ_i be a vector that stacks v_{ij}, η_{ijk}, ω_{ijk} terms, and let Σ be a corresponding vector of variances. If individual i resides in spatial unit j, then define y_{ij} = 1, otherwise y_{ij} = 0. Given these interpretations, the likelihood function condition on the value of $\Gamma_{-\sigma}$ and Λ_i may be written as:

$$L_i(\Gamma_{-\sigma})|\Lambda_i = \prod_{j=1}^{J}\{P_i(j) \times P_i(e_1^*, e_2^*, e_3^*, \cdots, e_M^*, 0,0,\cdots,0)\}^{y_{ij}} \quad (3)$$

where, $P_i(j)$ is the probability of household i

choosing to live in the spatial unit j. And $P_i(e_1^*, e_2^*, e_3^*, \cdots, e_M^*, 0,0, \cdots, 0)$ is the probability household i chooses to own and use M alternatives from K end-uses.

Consequently, the unconditional likelihood function can be derived from Eq. (3) as:

$$L_i(\Gamma) = \int_{\Lambda_i} (L_i(\Gamma_{-\sigma}) | \Lambda_i) dF(\Lambda_i | \Sigma) \tag{4}$$

3. DATA

A survey was designed to collect the information about the in-home/out-of-home expenditures and energy consumption patterns of households in Beijing. The questionnaire contents were improved based on a pilot survey. The candidate households were first randomly visited based on a convenient sampling method and those who agreed to participate in the survey (nearly 2,000 households) were asked to fill in the questionnaires by a household member, who is most familiar with household energy consumption. Several days later those respondent households were visited again with small gifts and their answered questionnaires were checked by a face-to-face interview. As a result, we successfully collected the valid questionnaires from 775 households. The questionnaire contents include the following information.

(1) Household social demographic attributes: household size, income, composition of members, housing area, dwelling type.

(2) Individual attributes: household member's gender, age, education, car license ownership, employment status, commute mode, travel time for work/school.

(3) Built environment: distance, frequency for visit, travel mode, travel time to the nearest public station, supermarket, shopping mall, park, hospital, kindergarten, school.

(4) Ownership and usage of in-home appliances and vehicles: attributes (e.g., type, size, capacity) of appliances and vehicles, frequency and/or duration of usage per week in different seasons for appliances and vehicles.

(5) Energy consumption: monthly energy consumption or monetary expenditure spent on electricity, gas, and gasoline in four seasons.

Based on the aforementioned methodology and the rich survey data, an empirical analysis is done to explore the household energy consumption behavior by considering the interaction with residential location choice.

4. ESTIMATION RESULTS

Estimation results indicate that there is significant dependency among the choice dimensions and that self-selection effect and unobserved heterogeneity of households cannot be ignored when modeling residential environment and household energy consumption behavior interactions. Concrete explanation of the model results will be introduced during the presentation.

REFERENCES

1) Bhat, C.R.: The multiple discrete–continuous extreme value (MDCEV) model: role of utility function parameters, identification considerations, and model extensions, Transportation Research Part B 42 (3), pp. 274-303, 2008.
2) Pinjari, A.R., Bhat, C.R., Hensher, D.A.: Residential self-selection effects in an activity time-use behavior model, Transportation Research Part B 43, pp. 729-748, 2009.

(Received May 6, 2011)