Coping with Insufficient Data:
The Case of Household Automobile Holding Modeling
by
Ryuichi Kitamura and Toshiyuki Yamamoto

It is often the case that typically available data do not contain all the variables that are desired for the analysis of the behavior of interest. In case of the analysis of household automobile holding behavior, for example, information on the cost of holding an automobile is rarely available in data, but has to be generated by the analyst based on a set of assumptions. In fact, unless the data have been collected specifically to analyze household automobile holding behavior, information that is needed to determine the cost of automobile holding and operation—e.g., make and model of the automobile, acquired new or used, purchase price, fuel consumption rate, or insurance costs—is typically unavailable.

A possible approach when data are insufficient is to develop a theoretical model which, based on external principles, embodies relationships among observed variables. In this study, a utilitarian model of household automobile holding is developed based on the assumption that a household holds an optimum number of automobiles at the time of observation. A unique feature of the model is the adoption of the notion of base auto ownership cost. This is the minimum expenditure that is required per unit time to hold an automobile, and each household is assumed to spend a nonnegative amount of money in addition to the base cost to hold a better automobile that offers more amenities. With the utilitarian assumption that the household optimizes its vehicle holding and use, the model expresses the utility of automobile holding in terms of income and household size, without requiring variables that can hardly be measured, e.g., unit cost of auto and transit travel.

Let

n_A = number of automobiles
n_H = number of adult household members
Y = household income
M_A = mobility by automobile per adult household member (person-km)
M_T = mobility by public transit per adult household member (person-km)
A = auto amenities expenditure per automobile
X = expenditure per adult household member for other goods
p_A = auto variable cost (per person-km)
p_T = transit variable cost (per person-km)
q = unit auto amenity cost
$\tilde{C}$ = base auto cost per unit time
p = price of other goods
$\underline{M}(Z)$ = minimum mobility per household (person-km)
$\bar{M}(Z)$ = maximum mobility per automobile (km)
Z = vector of household attributes

where automobile amenities expenditure refers to the amount spent over the minimum amount to hold a higher quality automobile or an automobile with more options.

Assuming that utility is produced by traveling to engage in activities, by consuming auto amenities, and by consuming other goods, let the utility of a household, given the number of automobiles it owns ($n_A > 0$), be

$$U = U(M_A, M_T, A, X \mid n_A) = \left(M_A \left(\frac{n_A}{n_H} \right)^{\eta} \right)^{\alpha} M_T{}^{\beta} A^{\gamma} X^{\delta} \tag{1}$$

where $\left(n_A/n_H\right)^{\eta}$ is a modifier that represents the effect of auto availability on the utility of auto mobility.

Now, taking the logarithm of Eq. (1), let the household's automobile holding behavior be depicted as

$$\text{Max } \ln U(M_A, M_T, A, X \mid n_A) = \alpha \ln M_A + \alpha\eta \ln(n_A/n_H) + \beta \ln M_T + \gamma \ln A + \delta \ln X$$

Subject to

$$\begin{aligned} & n_H p_A M_A + n_H p_T M_T + n_A(qA + \tilde{C}) + n_H pX = Y \\ & n_H (M_A + M_T) \geq \underline{M}(Z) \\ & \frac{n_H M_A}{n_A} \leq \bar{M}(Z) \end{aligned} \tag{2}$$

where the first condition represents income constraint, the second condition indicates that the minimum mobility requirements be met, and the third condition represents the ceiling on the use of household automobiles in terms of total vehicle kilometers. The Lagrangean is given as

$$\begin{aligned} \mathcal{L} = & \alpha \ln M_A + \alpha\eta \ln(n_A/n_H) + \beta \ln M_T + \gamma \ln A + \delta \ln X \\ & -\lambda(n_H p_A M_A + n_H p_T M_T + n_A(qA + \tilde{C}) + n_H pX - Y) \\ & -\mu(\underline{M}(Z) - n_H(M_A + M_T)) - \rho(n_H M_A/n_A - \bar{M}(Z)) \end{aligned} \tag{3}$$

In this paper, only unbounded solution ($\mu^* = \rho^* = 0$) is examined.

Normalizing the utility function by letting $\delta = 1$, the first-order conditions for the optimum are, by differentiating the Lagrangean with respect to M_A, M_T, A and X,

$$M_A = \frac{\alpha}{\lambda n_H p_A}, \quad M_T = \frac{\beta}{\lambda n_H p_T}, \quad A = \frac{\gamma}{\lambda n_A q}, \quad X = \frac{1}{\lambda n_H p}. \tag{4}$$

Substituting these into the income constraint,

$$\lambda = \frac{\alpha + \beta + \gamma + 1}{Y - n_A \tilde{C}} = \frac{K_1}{Y - n_A \tilde{C}} \tag{5}$$

where $K_1 = \alpha + \beta + \gamma + 1$. The optimum solution is

$$M_A^* = \frac{Y - n_A \tilde{C}}{n_H p_A} \frac{\alpha}{K_1}, \; M_T^* = \frac{Y - n_A \tilde{C}}{n_H p_T} \frac{\beta}{K_1}, \; A^* = \frac{Y - n_A \tilde{C}}{n_A q} \frac{\gamma}{K_1}, \; X^* = \frac{Y - n_A \tilde{C}}{n_H p} \frac{1}{K_1} \tag{6}$$

The indirect utility function is obtained as

$$U^* = K_1 \ln(Y - n_A \tilde{C}) - \alpha \ln n_H p_A + \alpha\eta(\ln n_A - \ln n_H) - \beta \ln n_H p_T - \gamma \ln n_A q - \ln n_H p \\ + \alpha \ln \alpha + \beta \ln \beta + \gamma \ln \gamma - K_1 \ln K_1 \tag{7}$$

Rearranging this,

$$\begin{aligned} U^* &= K_1 \ln(Y - n_A \tilde{C}) - \gamma \ln n_A + \alpha\eta \ln n_A - (\alpha + \beta) \ln n_H - \alpha\eta \ln n_H - \ln n_H \\ &\quad - \alpha \ln p_A - \beta \ln p_T - \gamma \ln q - \ln p + \alpha \ln \alpha + \beta \ln \beta + \gamma \ln \gamma - K_1 \ln K_1 \\ &= K_1 \ln(Y - n_A \tilde{C}) - (\gamma - \alpha\eta) \ln n_A - (\alpha + \beta + \alpha\eta + 1) \ln n_H + K_2 \end{aligned} \tag{8}$$

where

$$K_2 = -\alpha \ln p_A - \beta \ln p_T - \gamma \ln q - \ln p + \alpha \ln \alpha + \beta \ln \beta + \gamma \ln \gamma - K_1 \ln K_1 . \tag{9}$$

Note that there is no element in K_2 that is associated with the attributes of the household.

In case the household does not hold an automobile ($n_A = 0$), let the household behavior be formulated as

$$\begin{aligned} &\text{Max } U(M_T, X) = \beta \ln M_T + \delta \ln X \\ &\text{Subject to} \\ &\qquad n_H p_T M_T + n_H p X = Y \\ &\qquad n_H M_T \geq \underline{M}(Z) \end{aligned} \tag{10}$$

The indirect utility function is obtained as

$$\begin{aligned} U^* &= \beta(\ln \beta + \ln Y - \ln n_H p_T - \ln(\beta + 1)) + (\ln Y - \ln n_H p - \ln(\beta + 1)) \\ &= (\beta + 1) \ln Y - (\beta + 1) \ln n_H - \beta \ln p_T - \ln p + \beta \ln \beta - (\beta + 1) \ln(\beta + 1) \\ &= (\beta + 1) \ln Y - (\beta + 1) \ln n_H + K_3 \end{aligned} \tag{11}$$

where

$$K_3 = -\beta \ln p_T - \ln p + \beta \ln \beta - (\beta + 1) \ln(\beta + 1) \tag{12}$$

In sum, the indirect utility of automobile holding is given as

$$U^* = \begin{cases} (\beta + 1) \ln Y - (\beta + 1) \ln n_H + K_3 & \text{if } n_A = 0 \\ (\alpha + \beta + \gamma + 1) \ln(Y - n_A \tilde{C}) - (\gamma - \alpha\eta) \ln n_A & \\ \quad - (\alpha + \beta + \alpha\eta + 1) \ln n_H + K_2 & \text{if } n_A = 1, 2, \ldots \end{cases} \tag{13}$$

Letting $C = K_3 - K_2$ and introducing random error terms, the $\varepsilon_i(n_A)$, redefine the indirect utility as

$$U_i^*(n_A) = \begin{cases} (\beta + 1) \ln Y - (\beta + 1) \ln n_H + C + \varepsilon_i(0) & \text{if } n_A = 0 \\ (\alpha + \beta + \gamma + 1) \ln(Y - n_A \tilde{C}) - (\gamma - \alpha\eta) \ln n_A & \\ \quad - (\alpha + \beta + \alpha\eta + 1) \ln n_H + \varepsilon_i(n_A) & \text{if } n_A = 1, 2, \ldots \end{cases} \tag{14}$$

where *i* refers to the household. With the assumption that the $\varepsilon_i(n_H)$ are i.i.d. Gumbel, the parameters can be estimated by formulating a multinomial logit model of automobile ownership with the indirect utility function as defined in Eq. (13).

For this purpose, it may be more convenient to rewrite (13) in terms of relative indirect utility, by adding $(\alpha+\beta+\alpha\eta+1)\ln n_H$ to the right-hand side:

$$U_i^*(n_A)=\begin{cases}(\beta+1)\ln Y+\alpha(1+\eta)\ln n_H+C+\varepsilon_i(0) & \text{if } n_A=0\\ (\alpha+\beta+\gamma+1)\ln(Y-n_A\tilde{C})-(\gamma-\alpha\eta)\ln n_A+\varepsilon_i(n_A) & \text{if } n_A=1,2,\ldots\end{cases}\qquad (14')$$

Note that the differences in utility across different values of n_A remain unchanged between Eq. (14) and Eq. (14').

Eq. (14) may appear strange at first because the relative utility of owning no automobile increases with the number of adult household members. Note, however, that when household income is given, then owning an automobile may become less advantageous as the number of adult members increases because income per adult member decreases with more adult members. On the other hand, the relative utility of owning more automobiles increases when $(\gamma-\alpha\eta)$ is negative, or, when $\alpha\eta>\gamma$.

Assuming that the maximum number of automobiles a household may hold is $n_{\max}$, the probability that household *i* will hold n_A automobiles can be expressed as

$$p_i(n_A)=\begin{cases}\exp\{(\beta+1)\ln Y+C\}/K & n_A=0\\ \exp\{(\alpha+\beta+\gamma+1)\ln(Y-n_A\tilde{C})-(\gamma-\alpha\eta)\ln n_A-\alpha(1+\eta)\ln n_H\}/K & \\ & n_A=1,2,\ldots,n_{\max}\end{cases}\qquad (15)$$

where

$$\begin{aligned}K&=\exp\{(\beta+1)\ln Y+C\}\\&\quad+\sum_{n_A=1}^{n_{\max}}\exp\{(\alpha+\beta+\gamma+1)\ln(Y-n_A\tilde{C})-(\gamma-\alpha\eta)\ln n_A-\alpha(1+\eta)\ln n_H\}\end{aligned}\qquad (16)$$

Unknown parameters are α, β, γ, η and *C*. The sum, $(\alpha+\beta+\gamma+1)$, can be estimated as the coefficient of $\ln(Y-n_A\tilde{C})$, $\alpha(1+\eta)$ as the coefficient of $\ln n_H$, and $(\gamma-\alpha\eta)$ as the coefficient of $\ln n_A$.

The variable, $\tilde{C}$, is also unknown. It is proposed that alternative values be postulated for $\tilde{C}$ when estimating the model, and the one that offers the largest likelihood value be used as the value of $\tilde{C}$. Thus, the minimum cost of holding an automobile is estimated through the exercise of estimating the model of this study.

Importantly, the formulation does not require that p_A, p_T and p be known. Since these variables are not in typically available data sets and since it is not easy to empirically determine the values of these variables, use of the theoretical formulation shown here aids in the quantification of the indirect utility function of household automobile ownership.

Yet, there are at least two major issues that remain. First, boundary solutions must be incorporated into the estimation process. Second, it is logical to assume that the parameters α,

β and γ, and possibly η, vary across households depending on their attributes. Once we allow them to vary, however, the constant term, *C*, of Eq. (14') varies from household to household. Estimating the parameters, then, requires that p_A, p_T and p be known.