

Modeling Participation and Consumption in the Greek Interurban Public Transportation Market

*Theodore Tsekeris, Centre for Planning and Economic Research (KEPE)
Athens, Greece*

Loukas Dimitriou, National Technical University of Athens

Abstract

This article investigates the Greek household travel demand for domestic interurban public transportation using cross-sectional micro-data from a countrywide Household Budget Survey. A number of limited dependent variable models, including the Heckit, Two-Part, and Double-Hurdle models, are implemented to jointly estimate the probability of selecting a specific mode and the amount of using it. The results provide useful insight into the existence of feedback relationships between the decision-making mechanisms of mode selection and amount of usage, and they demonstrate the predominant effect of income on the demand for coaches, railways, airplanes, and coast-wise sea ferries.

Introduction

The principal objective of this article is to investigate the consumption demand of Greek households for interurban public transportation services using cross-sectional data from a countrywide Household Budget Survey (HBS). The amount of money spent by households for trip making can provide a useful metric of their

travel demand for different passenger transportation modes. In particular, the magnitude of (e.g., monthly average) travel expenditures can incorporate such information as those related to the frequency and amount of trip making as well as the monetary cost of travel by each mode. The use of expenditure micro-data can facilitate the task of identifying major social and economic determinants of household budget allocation for different transportation modes. In particular, such data can allow determining different trip decision-making structures of travelers belonging to diverse economic, social, and demographic population groups and dissimilar geographical settings.

The present study concentrates on simultaneously identifying the main factors influencing the choice and usage of interurban public transportation services in Greece. The data about travel demand for air travel, coast-wise sea travel, and intercity coach and rail travel services refer to the household level and they are based on the (most recent) 2004–05 Greek HBS. The article provides an overview of the current econometric models used in the literature to estimate travel demand using micro-data on transportation expenditures. The methodological approaches employed in the current study to analyze intercity travel demand are examined, and the study data and the variables used in the model estimation are described. Results of the model estimation are presented and conclusions and policy implications are discussed.

Travel Demand Models Using Micro-data on Expenditures

Although the effects of several economic and sociodemographic factors on household travel choices have been well documented and studied using cross-sectional data from a variety of sources, very few studies have investigated the mechanisms of household spending behavior for trip making. These studies are principally based on the use of limited dependent-variable models. The most familiar type of such models is the Tobit model (Tobin 1958), which has been applied to a wide range of consumer demand studies using micro-data (Deaton 1997). Among them, Hagemann (1981) studied the household expenditures for vacation travel using the 1972–73 U.S. Consumer Expenditure Survey (CES). Nolan (2003) investigated urban household travel decisions in the Dublin area, separately with regard to petrol, bus, and taxi fares using the 1994–95 Irish HBS. Also, Thakuriah and Liao (2006) examined the household decisions on the daily short-distance (urban) overall travel expenditures using the 1999–2000 U.S. CES.

The use of Tobit models is based on the rather restrictive assumption that the same variables affect in the same way both the decisions of travelers to use or not to use a specific transportation mode and the intensity of using it. In this way, Tobit models can only capture *corner solutions*, which imply that all households are potential users and choose to use or not to use a particular mode due to lack of affordability (low income) or the high price of petrol or public transportation fare, or both (low income and high price). However, zero observations can sometimes be attributed to other factors, such as habitual (or true) nonusage of a specific transportation mode, infrequency of traveling with the particular mode, and reduced accessibility (e.g., due to lack of appropriate transportation infrastructure and public transportation services and geographical constraints, particularly in remote regions with low connectivity, such as islands).

The literature related to the treatment of potential bias in the travel demand estimation using micro-data due to the above factors is limited. Existing studies mostly concentrate on the effect of the habitual nonusage of a particular mode through application of the Heckman's two-stage sample selection estimator, also known as Heckit model or adjusted Tobit model, or the Limited Information Maximum Likelihood selection estimator (Heckman 1979). According to this methodology, the dichotomous choice of households to travel or not to travel with a specific mode dominates their decision about the intensity of using the particular mode. The Heckit model has been employed for analyzing the car petrol consumption behavior by Kayser (2000) using the U.S. Panel Study of Income Dynamics (PSID), Asensio et al. (2003a), who employed the 1990–91 Spanish HBS, and West and Williams (2007), who used the 1996–98 U.S. CES. Asensio et al. (2003b) also used the Heckit methodology for the estimation of household expenditures for urban public transportation services using the 1990–91 Spanish HBS. The solution of instrumental variable (IV) systems of travel demand equations based on micro-data can also be included in the category of the methods used for controlling the effect of selection bias (Bergantino 1997).

The present study implements the Heckit methodology for the disaggregate analysis of household expenditures for intercity passenger transportation. In addition, a different type of limited dependent variable model—the Double-Hurdle model—is employed here for comparison purposes. According to this model, the problem of potential users who do not use a transportation mode is examined with regard to factors concerning the habitual nonusage (first hurdle), as well as other factors that inhibit them from realizing a trip, such as personal or intrahousehold

constraints and reduced accessibility (second hurdle). Different versions of the Heckit and Double-Hurdle modeling methodologies are considered in this study (see below), in accordance with the existence of a feedback relationship between the decision-making mechanisms on choosing to use or not to use a specific mode and the amount of using the particular mode.

Methodological Approaches of the Study

The Heckit Model and the Two-Part Model

The consumption demand equation that describes the amount of expenditures y_i made by household i for using a specific transportation mode (e.g., for purchasing petrol or buying tickets or travel cards for using public transportation services), can be described as a latent dependent variable equation, which is typically solved using Ordinary Least Squares (OLS) among the subsample $y > 0$, as follows:

$$y_i^* = x_i\beta + u \tag{1}$$

where:

u is the error term that is assumed to be $\sim N(0, \sigma^2)$

x is a set of explanatory variables of the decision of each household i on the amount of expenditures to be spent

β is the corresponding vector of coefficients

In contrast with the actual outcome (i.e., the true expenditures made by a household i), the potential outcome is a latent variable y_i^* that is only partially observed. The non-zero observation values are assumed to be true observations of the potential outcome, but zero values indicate observations for which the potential outcome is missing (latent). The zero observations do not represent zero values for the potential outcome. In contrast with the Tobit model, rather than y^* being observed when $y^* > 0$, the y^* value is assumed to be observed based on the value of a second latent variable, z_i^* , where:

$$z_i^* = w_i\alpha + v \tag{2}$$

where:

v is the error term that is assumed to be $\sim N(0, \sigma^2)$

w is the set of selection variables of the decision of each household i on whether to use or not to use a particular mode

α is the corresponding vector of coefficients

Equation (2) is typically estimated using a binary probit (or logit) model, which provides the probability of household i to travel or not to travel with a specific mode. In this case, y is only observed if $z_i^* > 0$. Furthermore, the model is assumed to be governed by the following observability criteria:

$$z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0 \end{cases} \quad (3)$$

Equation (1), usually referred to as the consumption or outcome equation, and equation (2), usually referred to as the participation equation, together constitute the Heckit model. In contrast with the Tobit model, the participation and the consumption part of the Heckit model do not involve the same error structure. In addition, the selection variables w are not identical to the variables x of the consumption equation. Assuming that (u, v) has a bivariate normal distribution.

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho\sigma_u \\ \rho\sigma_u & \sigma_v^2 \end{pmatrix} \right] \quad (4)$$

where:

ρ is the correlation between u and v , the conditional equation providing the consumption part of the Heckit model can be written as follows:

$$E(y_i | z_i = 1) = x_i \beta + \rho \sigma_u \left[\frac{\phi(w_i \alpha)}{\Phi(w_i \alpha)} \right] = x_i \beta + \rho \sigma_u \lambda \quad (5)$$

where:

$\lambda = \left[\frac{\phi(w_i \alpha)}{\Phi(w_i \alpha)} \right]$ is the Inverse Mill's Ratio (IMR) that denotes the nonselection hazard

If the estimate $\hat{\lambda}_i$ is significant, then $H_0 : \rho=0$ can be rejected, which means that there is selection bias. Otherwise, the second term of equation (5) that includes the IMR is removed, and the Heckit model is reduced to the Two-Part Model (2PM), also referred to as the hurdle or complete dominance model (Duan et al. 2003).

In the 2PM, all zero observations are generated by the mode selection decision and since a household chooses a specific mode, then it would have a certain level of usage of this particular mode. This implies that there is only dominance of the selection part to the consumption part, rather than both dominance and dependence, as it is implied in the Heckit model. In such a situation, there is no systematic feedback relationship between the participation and the consumption decision of households for using a specific transportation mode. In the case of employing the 2PM with $\ln(y)$, the actual outcome of using a particular transportation mode can be predicted as follows:

$$E(y_i) = \Phi(w_i \alpha) \exp(x_i \gamma + \sigma_\gamma^2 / 2) \tag{6}$$

where:

- γ is the set of coefficients of the regression equation describing the amount of using a specific transportation mode
- σ_γ^2 is the variance of the regression equation

Both the Heckit and 2PM can be generally described within the following set of relationships:

$$y_i = \begin{cases} y_i^* = x_i \beta + u & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0 \end{cases}, \text{ with } z_i^* = w_i \alpha + v \text{ and } z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0 \end{cases} \tag{7}$$

The 2PM can be considered as more efficient than the Heckit when the specification conditions of the latter model are not met (i.e., $H_0 : \rho=0$ holds). In such cases, the 2PM avoids possible sources of inefficiency that can be present in the Heckit

due to multicollinearity between the IMR and the regressors of the consumption equation.

The Double-Hurdle Model

The Double-Hurdle model, proposed by Cragg (1971), provides a statistical counterpart of the aforementioned theoretical structure described in (7). More specifically, this model postulates that the usage of a transportation mode (i.e., a positive amount of expenses for it) presupposes that household members overcome two hurdles: (1) be potential users, dependent on the willingness to use the mode, and (2) actually use the mode, dependent on the ability to access the particular mode. In comparison to the rules followed in (7), the Double Hurdle model can be generally described as follows:

$$y_i = \begin{cases} y_i^* = x_i \beta + u & \text{if } z_i^* > 0 \text{ and } y_i^* > 0 \\ 0 & \text{otherwise} \end{cases}, \quad \text{with } z_i^* = w_i \alpha + v \quad \text{and} \quad z_i = \begin{cases} 1 & \text{if } z_i^* > 0 \\ 0 & \text{if } z_i^* \leq 0 \end{cases} \quad (8)$$

Such a theoretical structure can be appealing to the case of analyzing the household demand for interurban public transportation, where potential users may not be able to use a particular service for more than one reason. For instance, an individual may not want to use the intercity coach or railway (nonparticipation) because of the habit of the license-holding members to move with private car. On the other hand, it could be that the individual would like to use the intercity coach or train (i.e., he or she is a potential user), but is constrained from accessing the particular transportation service due to the absence of a station/terminal close to the area of residence or the low frequency of the related services.

A Double-Hurdle model with dependent or independent error terms u and v , referred to here as D-DHM and I-DHM, respectively, can be alternatively adopted, according to whether the hypothesis $H_0 : \rho = 0$ can be rejected or not rejected, respectively. The D-DHM assumes an endogenous feedback effect from the intensity of consumption to the participation decision (Deaton and Irish 1984). In the case that the error terms u and v are independent ($\rho = 0$), the I-DHM of Cragg (1971) is obtained. The Tobit model can be regarded as a nested version of the I-DHM with $\Phi(w_i \alpha) = 1$. Similarly to the case of the Heckit (or 2PM), the estimation of correlation ρ is also based here on the joint bivariate normal distribution of the error terms u and v , as described in relationship (4).

Description of the Study Data and Model Variables

The present study focuses on the analysis of the domestic interurban travel behavior of households, in relation to four different public transportation modes (coach, rail, airplane, and sea ferry). The current analysis is based on fare and travel card expenditure data, as they are originated from the 2004–05 Greek HBS. Sea travel expenses involve transporting a vehicle on the ferry such that origin and destination circulation travel could be accommodated by auto or motorcycle. This survey includes information about the travel (and nontravel) expenditures made by 6,555 households (about 2/1000 of the total population), using a multilayer stratified sampling methodology to ensure the representation power of the given sample at the level of Regions (NUTS II). Due to the very small proportion of households using intercity railway services, which has resulted in a share of less than 1 percent of the total travel expenditures, coach and rail service expenses have been aggregated here into a single category (interurban land public transportation modes) to avoid potential small-sample bias problems.

Although the current study is based on cross-sectional data, it uses information about the spatial structure of prices in the model specification. Specifically, the income variable is expressed here by the total monthly household expenditures, which typically provide a proxy of the household permanent income, weighted by the region's Relative Consumer Price Index (RCPI) for each transportation mode. The RCPI is given by the ratio of the CPI for each mode (coach and rail, airplane, and sea ferry) at a specific region to the CPI for all transportation modes at the particular region.

The CPI for all transportation modes has been calculated as the weighted sum of the CPIs of urban and interurban bus and rail services, sea ferries, airline services, taxi tariffs, and automobile petrol. The corresponding weights were provided by the NSSG and they refer to the relative contribution of the monetary travel cost of each transportation mode to the total CPI for all transportation modes in 2004. The information about the regional price variations of the taxi tariffs and public transportation fares was obtained from the Greek Ministry of Transport and Communications, while the information on the regional variation of petrol prices was obtained from the Greek Ministry of Development.

In addition to the income variable, a range of other sociodemographic and economic variables has been used in the present study to represent the characteristics of each household. Also, a HBS population density index was used to demonstrate contiguous sets of local areas with varying levels of residential density and total population size (see EUROSTAT 2005). Finally, the ratio of the gross regional product to the gross domestic product as well as regional dummy variables have been used to represent local-specific effects. The population density variables and region-specific dummies can capture local effects on the range and spatial distribution of destinations, and the frequency of using sea ferry and other transportation services. Further insights into the differential behavior of traveling by different modes could be obtained by complementing HBS related to trip expenditure information with travel surveys, which can typically provide more insight into the structure of trip characteristics.

The variables used in the model estimation, their symbols, and their basic descriptive statistics are shown in Table 1. The specification of the models was based on the log-likelihood performance criteria, the convergence behavior of the likelihood-maximizing estimator, and the need for assuring the identification of the Heckit model. The participation and consumption equations basically employ the same variables. The two equations include different variables for describing household composition and economic situation (see below). These variables have been employed to demonstrate possible interaction effects of the employment of different household members and the size and aging structure of each family on the decision to choose and the amount of using a specific mode. In accordance with the standard econometric formulation of similar studies related to the analysis of household travel (and nontravel) expenditures (Bergantino 1997; Asensio et al. 2003a; Asensio et al. 2003b), the consumption equation of all interurban public transportation modes follows a semilog linear form by employing the logarithms of the dependent (monthly travel expenditures by mode) and income variables.

Table 1. Variables Used in the Model Estimation

<i>Variable</i>	<i>Symbol</i>	<i>Mean</i>	<i>Std. Dev.</i>
Coach and rail expenses (Euros/month)	<i>land</i>	1.559	8.613
Air travel expenses (Euros/month)	<i>air</i>	1.206	13.267
Sea travel expenses (Euros/month)	<i>sea</i>	2.736	11.739
Income variable for coach and rail users (Euros/month)	<i>income(l)</i>	2068.611	1441.663
Income variable for airplane users (Euros/month)	<i>income(a)</i>	2020.51	1393.926
Income variable for sea ferry users (Euros/month)	<i>income(s)</i>	1968.77	1358.232
Gender (male=1)	<i>sex</i>	0.75	0.433
Age of the head of household (hoh)	<i>age</i>	54.259	17.2
Hoh age < 30 y.o.	<i>age20</i>	0.075	0.263
Hoh age between 30 - 40 y.o.	<i>age30</i>	0.158	0.365
Hoh age between 40 - 50 y.o.	<i>age40</i>	0.183	0.386
Hoh age between 50 - 60 y.o.	<i>age50</i>	0.182	0.386
Hoh age between > 60 y.o.	<i>age60</i>	0.402	0.49
No car owned	<i>ncar0</i>	0.34	0.474
One car owned	<i>ncar1</i>	0.523	0.5
Two cars owned	<i>ncar2</i>	0.12	0.325
Three or more cars owned	<i>ncar3</i>	0.017	0.13
Density > 500 inhab. / sq.km and population > 50000 inhab.	<i>hden</i>	0.423	0.494
Density between 100 – 500 inhab. / sq.km and population > 50000 inhab.	<i>mden</i>	0.03	0.171
Density < 500 inhab. / sq.km and population < 50000 inhab.	<i>lden</i>	0.547	0.498
Hoh working full-time	<i>ftime</i>	0.529	0.499
Number of children (< 13 y.o.)	<i>up13</i>	0.378	0.750
Number of teenagers (between 13 y.o. and 18 y.o.)	<i>teen</i>	0.110	0.355
Number of adults	<i>adults</i>	1.797	0.563

Table 1. Variables Used in the Model Estimation (cont'd.)

<i>Variable</i>	<i>Symbol</i>	<i>Mean</i>	<i>Std. Dev.</i>
Both hoh, spouse and at least one other member working	<i>hhe1</i>	0.037	0.189
Both hoh and spouse work, no other member working	<i>hhe2</i>	0.181	0.385
Either hoh or spouse, and least one other member work	<i>hhe3</i>	0.062	0.241
Either hoh or spouse work, no other member working	<i>hhe4</i>	0.265	0.442
At least two members working excl. the hoh and spouse	<i>hhe5</i>	0.015	0.123
One member working excl. the hoh and spouse	<i>hhe6</i>	0.067	0.249
None working in the household	<i>hhe7</i>	0.373	0.484
Expenses for motorcycle purchase (proxy for ownership)	<i>motexp</i>	0.006	0.077
Region's share to GDP	<i>pgdp</i>	23.123	20.219
None, primary or lower secondary education	<i>ledu</i>	0.548	0.498
Upper secondary, post-secondary nontertiary education	<i>medu</i>	0.288	0.453
Higher education (all refer to the hoh)	<i>hedu</i>	0.163	0.369
One person, aged 65 years or more	<i>hht1</i>	0.109	0.311
One person, aged 30 to 64 years	<i>hht2</i>	0.067	0.25
One person, under 30 years	<i>hht3</i>	0.027	0.162
One person with children (< 18 y.o.)	<i>hht4</i>	0.016	0.125
Couple without children, older person aged 65 or more	<i>hht5</i>	0.153	0.36
Couple without children, older person under 65 years	<i>hht6</i>	0.089	0.285
Couple with one child (< 18 y.o.)	<i>hht7</i>	0.077	0.266
Couple with two children (< 18 y.o.)	<i>hht8</i>	0.103	0.304
Couple with more than two children (< 18 y.o.)	<i>hht9</i>	0.024	0.153
Single parent or couple with at least one child > 18 y.o.	<i>hht10</i>	0.236	0.425
Any other type of household	<i>hht11</i>	0.1	0.3
Number of members active occupied	<i>eam</i>	1.005	0.957
Ownership of second residence	<i>seres</i>	0.189	0.392

d indicates dummy variable (0 or 1)

Results of the Estimation

The first hypothesis tested in this study refers to the similarity of the decision-making mechanisms of mode selection and usage, in terms of the existence of significant correlation between the error terms of the two (participation and consumption) equations. According to the results of the maximum likelihood ratio tests (see Table 2), the error terms of the participation and consumption equations in both types of limited dependent variable models are found to be zero correlated, based on the 5 percent significance level of the χ^2 distribution, for the cases of land and air transportation, in contrast with the case of sea transportation. Namely, these findings fail to reject the hypothesis that the 2PM and I-DHM, where the decision to choose a specific transportation mode does not affect the amount of using it, are acceptable alternatives to the Heckit and D-DHM, respectively, where the decision to choose a specific mode is influenced by the amount of using it, for coach and rail and airplane, in comparison to the sea ferry.

Table 2. Results of the Maximum Likelihood Ratio Test
(under the null hypothesis of zero correlation between the error terms of the participation and consumption equations in the two types of models)

<i>Dependent Variable</i>	<i>Type of Models</i>	<i>Maximum Likelihood Ratio Test Results^a</i>	
<i>land</i>	Heckit – 2PM	$\chi^2(1) = 0.15,$	$p > \chi^2 = 0.699$
	D-DHM – I-DHM	$\chi^2(1) = 2.50,$	$p > \chi^2 = 0.114$
<i>air</i>	Heckit – 2PM	$\chi^2(1) = 0.75,$	$p > \chi^2 = 0.386$
	D-DHM – I-DHM	$\chi^2(1) = 1.96,$	$p > \chi^2 = 0.162$
<i>sea</i>	Heckit – 2PM	$\chi^2(1) = 9.46,$	$p > \chi^2 = 0.002$
	D-DHM – I-DHM	$\chi^2(1) = 157.34,$	$p > \chi^2 = 0.000$

a. $\chi^2(1)_{.05} = 3.841$

The selection of the I-DHM for the case of coach and rail and airplane implies that participation and consumption decisions for these modes are not made simultaneously, and the utility related to the amount of using these modes does not significantly affect the decision to choose them. Moreover, the selection of the 2PM for these modes reinforces the above result, implying that the decision to choose

coach and rail and airplane is separate from the amount of using them. On the contrary, a significant relationship exists between the utility of selecting and using sea ferries. This relationship can be attributed to the features of the particular mode, such as operating speed and frequency of services, the highly seasonal character of the sea travel demand and, especially, the concept of many holiday-makers that the journey by sea ferry toward islands constitutes part of their vacation. Furthermore, the outcome of the likelihood ratio tests demonstrates that sample selection bias is not a significant problem in the case of coaches and railways and airplanes in comparison to the case of sea ferries.

Table 3 presents the estimated parameters of the participation equations of the 2PM and I-DHM for the intercity land and air passenger transportation categories, and the Heckit and D-DHM for the sea passenger transportation category. The effects of the explanatory variables on the mode selection decisions are generally found to be similar between the 2PM and the I-DHM (for land transportation, and, particularly, for air transportation) and the Heckit and the D-DHM (for sea ferries), in terms of both their direction and significance. This outcome signifies that there is no other significant factor (hurdle) that influences the selection of interurban public transportation modes, except of nonparticipation and corner solutions. In particular, the income variable has a significant positive impact on the selection of coach and train, airplane, and sea ferry.

The presence of children, teenagers, males, and older persons (as household heads) and, particularly, couples without children with the oldest person aged 65 or more (*hht5*) reduces the probability of selecting one of these transportation modes. These results signify the impact of personal and intrahousehold constraints, mostly related to aging structure, on the mode selection in the specific market. On the contrary, high population densities, second home ownership, and higher education level increase the probability of selecting these modes. The effect of other variables is mixed between different modes. More specifically, the number of private cars tends to decrease the probability of selecting coach and train, but it increases the probability of selecting sea ferry and airplane (when number of cars <3). Motorcycle purchase expenses are found to have a significant positive effect on selecting land and sea passenger transportation modes in comparison to selecting airplane. Hence, private vehicle owners have an increased propensity to select sea ferry, since such a mode allows origin and destination circulation travel by their own car or motorcycle.

Table 3. Estimated Parameters of the Participation Equations

	<i>land</i>	<i>air</i>	<i>sea</i>	<i>land</i>	<i>air</i>	<i>sea</i>
	<i>2PM</i>	<i>2PM</i>	<i>Heckit</i>	<i>I-DHM</i>	<i>I-DHM</i>	<i>D-DHM</i>
<i>income(l,a,s)</i>	0.460**	0.613**	0.490**	0.334**	0.613**	0.440**
<i>sex</i>	-0.175**	-0.075	-0.122**	-0.229**	-0.075	-0.157**
<i>age30</i>	-0.553**	-0.221	-0.192**	-0.530**	-0.221	-0.192**
<i>age40</i>	-0.387**	-0.253*	-0.204**	-0.339**	-0.253*	-0.195**
<i>age50</i>	-0.477**	-0.328**	-0.345**	-0.431**	-0.328**	-0.319**
<i>age60</i>	-0.623**	-0.134	-0.422**	-0.501**	-0.134	-0.403**
<i>ncar1</i>	-0.436**	0.128	0.173**	-0.530**	0.128	0.106
<i>ncar2</i>	-0.758**	0.048	0.215**	-0.835**	0.048	0.168*
<i>ncar3</i>	-0.852**	-0.045	0.497**	-0.879**	-0.045	0.441**
<i>medu</i>	0.088	0.257**	0.314**	0.101	0.257**	0.312**
<i>hedu</i>	0.151**	0.284**	0.369**	0.142*	0.284**	0.364**
<i>mden</i>	0.354**	-0.323	0.278**	0.359**	-0.323	0.371**
<i>hden</i>	0.101*	0.152*	0.081*	0.100	0.152*	0.088*
<i>ftime</i>	-0.239**	0.001	-0.256**	-0.149	0.002	-0.278**
<i>up13</i>	-0.171**	-0.102*	-0.111**	-0.155*	-0.102*	-0.106**
<i>teen</i>	-0.033	-0.246**	-0.141**	-0.011	-0.245**	-0.137**
<i>adults</i>	0.010	-0.234**	-0.108**	0.006	-0.234**	-0.139**
<i>hhe1</i>	0.216	-0.067	0.165	0.317*	-0.067	0.201
<i>hhe2</i>	-0.083	0.173	0.193*	-0.038	0.173	0.206*
<i>hhe3</i>	0.138	-0.040	0.271**	0.227	-0.040	0.296**
<i>hhe4</i>	0.017	-0.107	0.087	0.039	-0.107	0.087
<i>hhe5</i>	-0.125	-0.006	-0.236	-0.057	-0.006	-0.226
<i>hhe6</i>	-0.075	0.022	0.169*	-0.007	0.022	0.169*
<i>motexp</i>	0.610**	-0.075	0.926**	0.594*	-0.075	0.939**
<i>seres</i>	0.354**	0.036	0.257**	0.358**	0.036	0.270**
<i>pgdp</i>	0.002	-0.026**	-0.001	-0.005	-0.026**	-0.001

* indicates $0.05 < p\text{-value} < 0.1$

** indicates $p\text{-value} < 0.05$, using the *t*-test statistics

Table 4 shows the estimated parameters of the consumption equations of the 2PM and I-DHM for the intercity land and air passenger transportation categories, and the Heckit and D-DHM for the sea passenger transportation category. In contrast with the participation equations (see Table 3), the explanatory variables of the consumption equations of the 2PM and I-DHM (for land and air transportation) and the Heckit and D-DHM (for sea ferries) are not always found to have the same effect on travel demand in terms of both their direction and significance. Thus, there are possibly significant additional factors that affect the amount of using the Greek interurban public transportation modes, except those of nonparticipa-

tion and corner solutions. Namely, there are underlying factors that do not allow potential users to actually use these public transportation services.

Table 4. Estimated Parameters of the Consumption Equations

	<i>land</i>	<i>air</i>	<i>sea</i>	<i>land</i>	<i>air</i>	<i>sea</i>
	<i>2PM</i>	<i>2PM</i>	<i>Heckit</i>	<i>I-DHM</i>	<i>I-DHM</i>	<i>D-DHM</i>
<i>income(l,a,s)</i>	0.182**	0.146**	0.406**	0.999**	0.543**	0.572**
<i>sex</i>	-0.046*	-0.044**	-0.045**	0.175	-0.035	0.039
<i>age30</i>	-0.190**	-0.086**	0.094	-0.303	-0.216	0.182
<i>age40</i>	-0.145**	-0.099**	0.126	-0.328	-0.298	0.156
<i>age50</i>	-0.146**	-0.107**	-0.068	-0.261	-0.368	-0.061
<i>age60</i>	-0.211**	-0.066*	0.105	-0.644	-0.275	0.148
<i>ncar1</i>	-0.127**	-0.014	0.275**	0.390*	-0.302	0.425**
<i>ncar2</i>	-0.231**	-0.009	0.211	0.406	-0.190	0.327*
<i>ncar3</i>	-0.330**	-0.051	0.538**	-0.052	-0.513	0.704**
<i>hden</i>	0.029	0.031*	-0.078	-0.028	0.106	-0.084
<i>mden</i>	0.150**	-0.049	0.202	0.102	0.352	-0.373
<i>ftime</i>	-0.092**	-0.017	0.144	-0.668**	-0.340	0.189
<i>motexp</i>	0.341**	-0.009	0.181	0.363	0.466	0.261
<i>pgdp</i>	0.001	-0.007**	0.023**	0.028	-0.001	0.032**
<i>medu</i>	0.038	0.041**	0.067	-0.074	-0.039	0.111
<i>hedu</i>	0.078**	0.064**	0.144	0.184	0.146	0.192*
<i>hht1</i>	-0.102**	0.075**	-0.280	-0.355	-0.398	-0.592**
<i>hht2</i>	-0.027	0.119**	-0.216	0.505	-0.254	-0.277
<i>hht3</i>	0.244**	0.016	0.133	0.345	-0.419	0.332
<i>hht4</i>	-0.074	-0.029	-0.071	-0.149	2.089**	-0.352
<i>hht5</i>	-0.045	0.061*	0.202	0.094	-0.153	0.185
<i>hht6</i>	-0.083**	0.066**	0.341**	0.321	0.063	0.362*
<i>hht7</i>	-0.189**	0.078**	0.333**	-0.612	0.246	0.331*
<i>hht8</i>	-0.181**	0.029	0.192	-0.397	0.158	0.185
<i>hht9</i>	-0.167**	0.021	0.108	-0.107	0.503	0.040
<i>hht10</i>	-0.057*	0.014	0.017	-0.530**	-0.213	0.005
<i>eam</i>	0.003	0.01	-0.059	-0.041	0.037	-0.081
<i>seres</i>	0.122**	-0.001	-0.057	0.060	-0.104	-0.031

* indicates $0.05 < p\text{-value} < 0.1$

** indicates $p\text{-value} < 0.05$, using the *t*-test statistics

Such factors can be related to personal, intrahousehold, and geographical constraints, as well as accessibility conditions, which are associated with the service and infrastructure characteristics of each individual mode. These situations may arise due to car unavailability, needs for escort trips involving children and elderly people, reduced connectivity of public transportation services in low-density areas and isolated regions (e.g., islands), and shortage of ferry services in winter season. The major factor that positively and significantly affects the usage of all interurban public passenger transportation modes is income, as in the case of mode selection. This outcome verifies that all these modes are normal goods (i.e., their usage increases as income levels increase). The Double Hurdle models produce larger income coefficients for all modes in the consumption equations than those produced by the Heckit and 2PM approaches. On the contrary, the latter models generally produce larger income coefficients in the participation equations than those produced by the former models. These differences imply that income plays a more significant role in travel expenditures when the market demand curve is estimated over the entire population (travelers and nontravelers), as in the Double Hurdle models, rather than when the relevant population consists only of travelers, as in the Heckit and 2PM models.

Among other factors, the increase of age is found to reduce the amount of using air and land passenger transportation modes. In all models used in the study, car ownership and motorcycle purchase expenses are found to positively affect the amount of sea travel. This outcome verifies the complementary relationship between private vehicle and sea ferry. On the other hand, car ownership is found to have a negative impact on the amount of air travel, which implies the substitution relationship between airplane and car usage for long-distance trips. Also, the 2PM demonstrates a significant negative influence of car ownership on the amount of using coach and rail, while the I-DHM shows such a negative effect only in the case of high car ownership levels (>2 cars).

Moreover, couples with children, as they are represented by variables *hht7*, *hht8*, and *hht9*, have a negative influence on the amount of coach and rail travel, but a positive influence on the amount of air and sea travel, in all models. Other household characteristics, which are not explicitly represented in the current model specification, could be having an impact on the selection and intensity of using different interurban public transportation services. In particular, the travel behavior of couples with children may be related to the pricing structure of transportation companies, which is implicitly captured here through the spatial variation of

prices, and the value of time, which is partially represented by the income variable. This is because such households may consider care-giving costs, which, although they are independent of travel costs, can influence mode choice in favor of faster travel modes (i.e., airplane in comparison to coach and rail), as it was obtained from the estimation results.

Conclusions and Policy Implications of the Study

This article provided an econometric analysis of the interurban public transportation demand of Greek households. Two modeling approaches pertaining to different assumptions about consumer behavior were implemented based on the micro-data of the 2004 Greek HBS. These approaches refer to (1) the Heckit or 2PM, dependent on whether there is feedback between the participation and consumption decisions, and (2) the Double Hurdle model with dependent or independent errors. The major difference between these two types of limited dependent variable models lies in the treatment of zero observations. The decision about the selection of the appropriate model to be implemented by researchers employing similar datasets based on Household Budget (or Consumer Expenditure) Surveys is not straightforward, and direct comparison of the results is difficult to be made. This is because the Heckit and 2PM models only include travelers, while the Double Hurdle models use all observations. The study results provide valuable insights into how such different assumptions can affect theoretical developments in travel behavior and modeling. The Double Hurdle models can be considered as more general than the Heckit and 2PM approaches, since, unlike the latter models, they assume that some nontravelers could be induced to use a transportation mode if some changes were to occur in their sociodemographic situation or economic status. Likewise, travelers may stop using a specific mode if they undergo changes in some of the above factors. On the other hand, the Heckit and 2PM models investigate such changes at the one or the other direction only for the existing population of travelers.

The significant differences observed in the estimated coefficients of the consumption equations of the two types of models suggest that there are possibly other factors than those related to corner solutions and habitual nonparticipation, which affect the amount of using interurban public transportation modes. Such factors can refer to personal, intrahousehold, geographical, and accessibility constraints. The identification of these factors help better explain the reasons for which some potential users do not actually consume public transportation service. The

increase of income level was found to significantly and positively affect both the participation and consumption decisions for all modes. Other factors such as car ownership, motorcycle purchase expenditures, age, household composition, residential density, and education level were also found to significantly influence public transportation demand in diverse ways. The estimation results reject the existence of feedback relationship between the selection and amount of using coach and rail and airplane, in comparison to the sea ferry, where the two decisions are found to be correlated. This relationship possibly reflects the formation of consumption habits and a positive utility associated with the sea travel.

The study results have several implications concerning those involved in public transportation policy making and planning. These implications refer to the design and evaluation of suitable measures for different transportation market segments at three levels of policy. The first and most general policy level corresponds to the entire population of travelers and nontravelers, and it relates to the results of the Double Hurdle models. The second policy level corresponds to the whole population of travelers, and it relates to the Heckit and 2PM models. The third policy level concerns specific target groups of potential or actual travelers, based on, for example, their household, income, education, or employment status, and dependent on whether the market demand curve is estimated over the entire population (of travelers and nontravelers). Particularly in the context of the Greek interurban public transportation market, the present results suggest that the increasing levels of car ownership are likely to have a positive impact on the modal share for sea ferry, in comparison to that for coach and rail and airplane. Current demographic trends related to the rapid aging of the population are also expected to adversely affect the usage of air and land passenger transportation modes. Targeted measures for promoting air and sea travel are anticipated to be mostly acceptable by families with children.

The demand models described in this study can be regarded as belonging to the family of trip generation models. Therefore, substitution or complementarity relationships between different transportation modes are not explicitly represented in these models, except of the effect of private vehicle ownership on interurban travel demand. Nonetheless, the modal share of urban travel can arguably affect decisions concerning the participation and consumption of interurban public transportation modes. These effects may be related to seasonal demand variations (vacation vs. nonvacation periods), and involve access-related complementarity and activity- or destination-based substitution among urban and interurban

transportation modes. An appropriate treatment of such relationships would involve the use of systemwide approaches, which give rise to a number of structural consumer demand equations to represent interactions among the household budget shares for urban and interurban travel by alternative modes.

References

- Asensio, J., A. Matas, and J-L Raymond. 2003a. Petrol expenditure and redistributive effects of its taxation in Spain. *Transportation Research Part A: Policy and Practice* 37 (1): 49–69.
- Asensio, J., A. Matas, and J-L Raymond. 2003b. Redistributive effects of subsidies to urban public transport in Spain. *Transport Reviews* 23 (4): 433–452.
- Bergantino, A. 1997. Estimating Engel curves for transport expenditures: Evidence from UK household budget data. *International Journal of Transport Economics* 24 (2): 279–305.
- Cragg, J. G. 1971. Some statistical models for limited dependent variables with applications to the demand for durable goods. *Econometrica* 39: 829–844.
- Deaton, A., and M. Irish. 1984. Statistical models for zero expenditures in household budgets. *Journal of Public Economics* 23: 59–80.
- Deaton, A. 1997. *The analysis of household surveys: Microeconometric analysis for development policy*. Baltimore, MD: Johns Hopkins University Press.
- Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse. 1983. A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics* 1: 115–126.
- EUROSTAT. 2005. *Household budget surveys*. HBS Working Group, Revised Document No. E2/HBS/153-E/2003/EN, Directorate D-2, Eurostat, European Commission, Luxemburg.
- Hagemann, R. P. 1981. The determinants of household vacation travel: some empirical evidence. *Applied Economics* 13 (2): 225–234.
- Heckman, J.J. 1979. Sample selection bias as a specification error. *Econometrica* 47: 153–161.
- Kayser, H. 2000. Gasoline demand and car choice: estimating gasoline demand using household information. *Energy Economics* 22 (3): 331–348.

- Nolan, A. 2003. The determinants of urban households' transport decisions: A microeconomic study using Irish data. *International Journal of Transport Economics* 30 (1):103–132.
- Thakuriah, P., and Y. Liao. 2006. Transportation expenditures and ability to pay: Evidence from Consumer Expenditure Survey. *Proceedings of the Transportation Research Board 85th Annual Meeting*. National Research Council, Washington, DC.
- Tobin, J. 1958. Estimation of relationships for limited dependent variables. *Econometrica* 26: 24–36.
- West, S.E., and R. C. Williams. 2007. Optimal taxation and cross-price effects on labor supply: Estimates of the optimal gas tax. *Journal of Public Economics* 91 (3–4): 593–617.

About the Authors

THEODORE TSEKERIS (tsek@kepe.gr) is a research fellow in transportation economics at the Centre for Planning and Economic Research (KEPE) in Athens, Greece. He holds a bachelor's degree in civil engineering from City University, London, postgraduate diploma and master's degree in transportation from the University of London, U.K., and a Ph.D. in transportation from the National Technical University of Athens, Greece. His research experience spans the areas of transportation planning, management and economics, travel demand analysis, and evaluation of advanced technologies.

LOUKAS DIMITRIOU (lucdimit@central.ntua.gr) is a research associate in the Department of Transportation Planning and Engineering at the National Technical University of Athens (NTUA), Greece. He holds a diploma in civil engineering with specialization in transportation from the NTUA. His research activities include the design and analysis of transportation systems, game-theoretic models of economic behavior, artificial intelligence, and econometric applications in transportation.