8 Auto Ownership and Use: An Integrated System of Disaggregate Demand Models

8.1 Introduction

This chapter presents the new model of the demand for personal-use cars and light trucks. The model is similar to previous models based on "disaggregate," or household-level, analysis in that it

- is estimated on a sample of households so as to capture the behavioral factors conditioning households' choices;
- describes each household's choices as depending on the characteristics of each class and vintage of vehicle (such as its operating cost, purchase price, seating capacity, etc.) as well as the characteristics of the household (such as its income, size, number of workers, etc.); and
- can be used in a simulation framework to provide consistent forecasts of total vehicle demand and use within a region.

In addition, the model has several important characteristics and capabilities that previous models have not incorporated. In particular, the model

- can forecast the number of vehicles owned and the number of miles traveled annually, by class and vintage of vehicle;
- explicitly recognizes the interdependence, or "interrelatedness," of a household's choice of how many vehicles to own and its choice of which class and vintage of vehicles to own;
- recognizes that a household's choice of how many and what class/vintage of vehicle to own depends on how much the household drives, and vice versa; and
- reflects the fact that each household chooses a particular make and model from within its chosen class of vehicles, but does so without requiring a specification of the demand for each make and model.

The model consists of a system of submodels that separately describe the number of vehicles owned, the class and vintage of each vehicle, and the miles traveled in each vehicle. While different submodels are used to describe different choices by the household, the system of submodels constitutes an integrated whole in that (1) its overall structure is derived from economic theory in which the household chooses **jointly** the number of vehicles to own, the class/vintage of each vehicle, the make/model of each vehicle, and the amount to drive each vehicle, and (2) the parameters of each submodel are estimated using techniques that account for the inter-

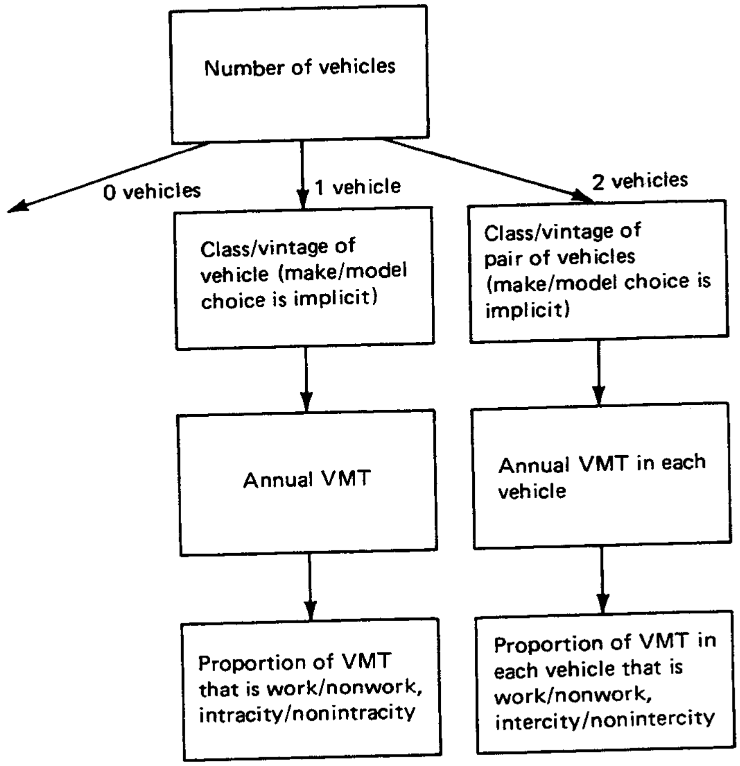

Figure 8.1
Household vehicle demand. Each box represents a separate submodel.

dependence among the households' choices. Consequently, the model embodies two significant advantages: the specification of different submodels for different choices facilitates understanding and use of the model; and consistency in the structure and estimation of the submodels ensures that the overall model accurately reflects the interdependent nature of a household's choices.

The structure of the model is most readily explained in terms of prediction. The model consists of seven submodels, each of which predicts one of the decisions represented in figure 8.1. First, the number of vehicles that the household owns is predicted. If the household is predicted to own no vehicles, then no further calculations are made. If the household is predicted to own one vehicle, the class and vintage of its vehicle is then predicted, followed by a prediction of annual VMT (vehicle miles traveled). Last, the proportion of miles that the vehicle will be driven by the household in each of four categories of travel is predicted. The four categories are short intracity work trips, other work trips, short intracity nonwork trips, and other nonwork trips.[1]

If a household is predicted to own two vehicles, then the model predicts the class and vintage of **each** of the vehicles. Next, the amount each vehicle is driven is predicted, followed by a prediction of the proportion of miles driven on each vehicle in each of the four categories of travel.

It is important to note that, while each household's decisions are predicted sequentially in the manner just described, the household is not assumed to **make** the decisions sequentially. Rather, the household simultaneously chooses the number of vehicles, class, and vintage of each vehicle and number of miles to drive, with each choice related to each other choice. The sequencing of the submodels is simply a convenience to the researcher and an aid to exposition, which is allowed by the particular form of the model (see the discussion on specification to follow and the general descriptions of GEV and continuous/discrete models in chapters 4 and 5).

While the model predicts choices of individual households, the ultimate purpose is to produce **aggregate** forecasts, such as the total number of vehicles by class and vintage in a region, state, or the nation as a whole. Aggregate forecasts for a particular geographic area are obtained by applying the model to each household in a sample of households from the area and aggregating the predictions for the sampled households. More specifically, for each household in a sample, the model is used to calculate the probability that the household will own a particular number of vehicles, each of a particular class and vintage, and the amount that the household will drive each of its vehicles. These calculations are performed for each sampled household separately, based upon the household's income, size, employment status, and location, as well as other factors, such as gas prices and fuel efficiency levels. Aggregate forecasts are obtained by taking a weighted sum of the predicted probabilities of each household in the sample, with the weights reflecting the sampling probability for the household. An illustration of this procedure is given in chapter 9, which presents forecasts for the state of California.

8.2 Model Specification

This section is relatively technical and need not be read for a general understanding of the model. General readers can skip to section 8.4, which describes the variables entering each submodel, the functional form relating the variables, and the estimated parameters of the functions.

The exact specification of the model, and its derivation from the economic theory of consumer behavior, is based on the recent work on GEV and discrete/continuous models discussed in chapters 4 and 5. The general form of the model is described first; the specific functions used in estimation are then presented as special cases and/or approximations of the general form.

We assume that each household chooses the number of vehicles to own, the class and vintage of each vehicle, the make and model of each vehicle, and the number of miles to travel in each vehicle so as to maximize its utility. The conditional indirect utility function of the household is denoted

$$V_{nc_nm_{c_n}} = f(Y, p_{nc_nm_{c_n}}, x_{nc_nm_{c_n}}), \tag{8.1}$$

where

n is the number of vehicles ($n = 0, 1, 2, \dots$);
c_n is the class/vintage combination of each of the n vehicles (i.e., c_n is a vector of length n whose elements denote the class/vintage combination of the vehicles);
m_{c_n} is the make/model of each of the n vehicles (i.e., m_{c_n} is a vector of length n whose elements denote the make/model of the vehicle(s), the make/model of each vehicle necessarily being within the class/vintage designated by c_n;
$V_{nc_nm_{c_n}}$ is the conditional indirect utility of the household, given that it owns n vehicles of class/vintage c_n and make/model m_{c_n}, expressed as a function of its income, the price of travel, and other explanatory variables;
Y is the income of the household;
$p_{nc_nm_{c_n}}$ is the cost per mile of traveling by vehicles of class/vintage c_n and make/model m_{c_n} (i.e., $p_{nc_nm_{c_n}}$ is a vector of length n whose elements are the cost per mile of traveling by vehicle(s) of class/vintage c_n and make/model m_{c_n}); and
$x_{nc_nm_{c_n}}$ is a vector of other explanatory variables, both observed and unobserved, that affect the utility that the household can obtain with n vehicles of class/vintage c_n and make/model m_{c_n}.

Given n, c_n, and m_{c_n}, the number of miles that the household will travel in the ith vehicle is, by Roy's identity, the (negative of the) derivative of the conditional indirect utility function with respect to the price per mile of traveling by the vehicle divided by the derivative with respect to income (see chapter 5 for a proof of this fact). This function can be denoted g_i:

$$\begin{aligned}\mathrm{VMT}^i_{nc_nm_{c_n}} &= -\frac{\partial V_{nc_nm_{c_n}}/\partial p^i_{nc_nm_{c_n}}}{\partial V_{nc_nm_{c_n}}/\partial Y} \\ &= g^i(Y, p_{nc_nm_{c_n}}, x_{nc_nm_{c_n}}), \qquad i = 1, \ldots, n,\end{aligned} \tag{8.2}$$

where

$\mathrm{VMT}^i_{nc_nm_{c_n}}$ is the miles traveled by the ith vehicle, given that the household owns n vehicles of class/vintage c_n and make/model m_{c_n}; and
$p^i_{nc_nm_{c_n}}$ is the cost per mile of traveling by the ith vehicle, given that the household owns n vehicles of class/vintage c_n and make/model m_{c_n}.

This equation represents the household's demand for travel in each vehicle, as a function of the cost of travel in all of its vehicles, the income of the household, and the other explanatory variables $x_{nc_nm_{c_n}}$.

The household chooses the number of vehicles to own, the class/vintage of each vehicle, and the make/model of each vehicle so that the conditional indirect utility function is maximized. That is, the household chooses n^* vehicles of class/vintage c_n^* and make/model $m^*_{c_n}$ if and only if

$$V_{n^*c_n^*m^*_{c_n}} > V_{nc_nm_{c_n}} \qquad \text{for all} \quad n, c_n, m_{c_n} \text{ other than } n^*, c_n^*, m^*_{c_n}.$$

Consequently, the probability that the household chooses n^*, c_n^*, and $m^*_{c_n}$ is simply

$$P_{n^*c_n^*m^*_{c_n}} = \mathrm{Prob}(V_{n^*c_n^*m^*_{c_n}} > V_{nc_nm_{c_n}} \text{ for all } n, c_n, m_{c_n} \text{ other than } n^*, c_n^*, m^*_{c_n}).$$

To evaluate this probability, recall that $V_{nc_nm_{c_n}}$ is a function of $x_{nc_nm_{c_n}}$, which includes both observed and unobserved variables. Combining the effects of all unobserved variables into one composite unobserved variable, take the mean of $V_{nc_nm_{c_n}}$ over this unobserved variable and label this mean $\bar{V}_{nc_nm_{c_n}}$. Then, without loss of generality, $V_{nc_nm_{c_n}}$ can be decomposed into its mean and deviations from the mean:

$$V_{nc_nm_{c_n}} = \bar{V}_{nc_nm_{c_n}} + e_{nc_nm_{c_n}}.$$

Assume that $e_{nc_nm_{c_n}}$ for all n, c_n, and m_{c_n} are jointly distributed in accordance with a particular Generalized Extreme Value (GEV) function:

$$\begin{aligned}&\mathrm{CDF}(e_{nc_nm_{c_n}} \text{ for all } n, c_n, \text{ and } m_{c_n}) \\ &= \exp\left\{-\sum_n \left[\sum_{c_n}\left[\sum_{m_{c_n}} [\exp(-e_{nc_nm_{c_n}})]^{1/(\theta)}\right]^{\theta/\lambda}\right]^{\lambda}\right\}\end{aligned} \tag{8.3}$$

where θ and λ are parameters. Then, as discussed in chapter 4, the probability that a household will choose n^* vehicles of class vintage c_n^* and make/model $m_{c_n}^*$ can be expressed in the following manner.

Decompose the mean utility that the household obtains from n vehicles of class/cintage c_n and make/model m_{c_n} into three parts:

$$\bar{V}_{nc_n m_{c_n}} = \bar{V}_n + \lambda \bar{V}_{c_n|n} + \theta\lambda \bar{V}_{m_{c_n}|nc_n},$$

where

$\bar{V}_n$ is the mean of $\bar{V}_{nc_n m_{c_n}}$ over class/vintages and makes/models;
$\bar{V}_{c_n|n}$ is the mean of $\bar{V}_{nc_n m_{c_n}}$ over makes/models minus $\bar{V}_n$ divided by λ; and
$\bar{V}_{m_{c_n}|nc_n}$ is $\bar{V}_{nc_n m_{c_n}}$ minus the mean of $\bar{V}_{nc_n m_{c_n}}$ over makes/models divided by $\theta\lambda$.

Note that $\bar{V}_n$ depends only on factors that vary over n, $\bar{V}_{c_n|nc_n}$ depends on factors that vary over n and c_n but not over m_{c_n}, and $\bar{V}_{m_{c_n}|nc_n}$ depends on factors that vary over n, c_n, and m_{c_n}.

Similarly, let the joint probability of a household choosing n^* vehicles of class/vintage c_n^* and make/model $n_{c_n}^*$ be written as the product of marginal and conditional probabilities:

$$P_{n^* c_n^* m_{c_n}^*} = P_{n^*} \cdot P_{c_n^*|n^*} \cdot P_{m_{c_n}^*|n^* c_n^*},$$

where

P_{n^*} is the marginal probability of n^* vehicles (marginal over all class/vintages and make/models);
$P_{c_n^*|n^*}$ is the probability of class/vintage c_n^* (conditional on n^*, and marginal over all make/models); and
$P_{m_{c_n}^*|n^* c_n^*}$ is the conditional probability of make/model $m_{c_n}^*$ (conditional on n^* vehicles of class/vintage c_n^*).

Then, given the above distribution for $e_{nc_n m_{c_n}}$, these marginal and conditional probabilities can be written as

$$P_{n^*} = \frac{\exp(\bar{V}_{n^*} + \lambda I_{n^*})}{\sum_n \exp(\bar{V}_n + \lambda I_n)}; \tag{8.4}$$

$$P_{c_n^*|n^*} = \frac{\exp(\bar{V}_{c_n^*|n^*} + \theta J_{c_{n^*}})}{\sum_{c_n} \exp(\bar{V}_{c_n|n^*} + \theta J_{c_n})}; \tag{8.5}$$

$$P_{m_{c_n}^*|n^* c_n^*} = \frac{\exp(\bar{V}_{m_{c_n}^*|n^* c_n^*})}{\sum_{m_{c_n}} \exp(\bar{V}_{m_{c_n}|n^* c_n^*})}; \tag{8.6}$$

where

$$J_{c_n^*} = \ln \sum_{m_{c_n}} \exp(\bar{V}_{m_{c_n}|n^* c_n^*}), \tag{8.7}$$

$$I_{n^*} = \ln \sum_{c_n} \exp(\bar{V}_{c_n|n^*} + \theta J_{c_n}). \tag{8.8}$$

The submodels that constitute the vehicle choice model are represented by equations (8.4) and (8.5). The vehicle quantity submodel, which describes the number of vehicles the household chooses to own, is given by equation (8.4). In this equation, the probability of owning a certain number of vehicles depends on elements of $\bar{V}_n$. In addition, the utility that the household would obtain by its choice of class/vintages affects the household's probability of choosing a certain number of vehicles; this is represented by the term I_n.

Equation (8.5) represents the submodel that describes the household's choice of class and vintage of each vehicle, given the number of vehicles it chooses to own. The probability of choosing a particular class and vintage depends on elements of $\bar{V}_{c_n|n}$. In addition, the utility that the household would obtain by its choice of make/model within a class/vintage affects the household's probability of choosing a certain class and vintage; this is represented by the term J_{c_n}.

The submodel that describes the miles traveled in each vehicle is represented by equation (8.2). The number of miles traveled depends on the household's income (Y), the cost per mile of travel in vehicles ($p_{nc_n m_{c_n}}$), and the other explanatory variables $x_{nc_n m_{c_n}}$.

This discussion is based on a general indirect utility function, labeled f in equation (8.1), which gives rise to general travel demand functions, labeled g in equation (8.2), and choice models that, at least with respect to the arguments within the logit form, are fully general (equations (8.4), (8.5), and (8.6)). The task now is to choose a specific indirect utility function and from it derive the specific functions used for the VMT submodel and the specific arguments used in the vehicle quantity and vehicle class/vintage submodels.

The following conditional indirect utility function has convenient properties that will become evident. Therefore, assume that each household faces a conditional indirect utility function of the form

No vehicles:

$V_{0c_0 m_{c_0}} = 0$, by normalization;

One vehicle:

$$V_{1c_1m_{c_1}} = (1/(1-\alpha_1))Y^{1-\alpha_1} + (1/\beta_1)\exp\{\delta_1 - \beta_1 p_{1c_1m_{c_1}}\} + \theta_{1c_1m_{c_1}}; \tag{8.9}$$

Two vehicles:

$$V_{2c_2m_{c_2}} = (1/(1-\alpha_2))Y^{1-\alpha_2} + (1/\beta_2)\exp\{\delta_2 - \beta_2 p^1_{2c_2m_{c_2}}\} - (1/\beta_2)\exp\{\delta_2 + \beta_2 p^2_{2c_2m_{c_2}}\} + \theta_{2c_2m_{c_2}};$$

and so on for $n = 3, 4, \ldots,$ where

$\alpha_1, \alpha_2, \beta_1, \beta_2$ are parameters;
δ_1 is a weighted sum of both observed and unobserved characteristics of the household, with the weights being parameters;
δ_2 is another weighted sum (not necessarily the same as δ_1) of characteristics of the household, with the weights being parameters;
$\theta_{1c_1m_{c_1}}$ is a weighted sum of both observed and unobserved household characteristics and characteristics of a vehicle of class/vintage c_1 and make/model m_{c_1}; and
$\theta_{2c_2m_{c_2}}$ is a weighted sum of household characteristics and characteristics of the **pair** of vehicles of class/vintage c_2 and make/model m_{c_2} (recall that c_2 is a vector denoting the class/vintage of each of the two vehicles, and similarly for m_{c_2}).

Using Roy's identity, the miles traveled in each vehicle, given the number, class/vintage, and make/model of the vehicles is

One vehicle:

$$\ln \mathrm{VMT} = \alpha_1 \ln Y - \beta_1 p_{1c_1m_{c_1}} + \delta_1; \tag{8.10}$$

Two autos:

$$\ln \mathrm{VMT}^i = \alpha_2 \ln Y - \beta_2 p^i_{2c_2m_{c_2}} + \delta_2, \qquad \text{for} \quad i = 1, 2. \tag{8.11}$$

Equations (8.10) and (8.11) are the specific functions used in the submodels describing vehicle miles traveled for one-vehicle and two-vehicle households. Note that because of the form of the two-vehicle indirect utility function, equation (8.11) represents the demand for travel in vehicle 1 or 2; that is, the forms of g^1 and g^2 in equation (8.2) are the same.

Since the conditional indirect utility function (8.9) is nonlinear in parameters of observed variables, a linear in parameters approximation to it in observed variables is used in estimation:

$$V_{0c_0m_{c_0}} = 0;$$

$$V_{1c_1m_{c_1}} = \psi_1 z_{1c_1m_{c_1}} + e'_{1c_1m_{c_1}}; \tag{8.12}$$

$$V_{2c_2m_{c_2}} = \psi_2 z_{2c_2m_{c_2}} + e'_{2c_2m_{c_2}};$$

where

$z_{nc_nm_{c_n}}$ $(n = 1, 2)$ are vector-valued functions of the observable characteristics of the household (i.e., Y and those entering δ_n and $\theta_{nc_nm_{c_n}}$) and observable characteristics of the class/vintage c_n and make/model m_{c_n} (i.e., $p_{nc_nm_{c_n}}$ and those entering $\theta_{nc_nm_{c_n}}$);
ψ_n $(n = 1, 2)$ are vectors of parameters; and
$e'_{nc_nm_{c_n}}$ $(n = 1, 2)$ are the errors in this approximation, which depend on both observed and unobserved variables.

Substituting the means of $\psi_1 z_{1c_1m_{c_1}}$ and $\psi_2 z_{2c_2m_{c_2}}$ over makes/models and vintages into equation (8.4) gives the specific functional form for the submodel that describes the number of vehicles the household will choose to own. This is the same as equation (8.4) with $\bar{V}_n$ replaced by a term that is linear in parameters.

Substituting the deviations of $\psi_1 z_{1c_1m_{c_1}}$ and $\psi_2 z_{2c_2m_{c_2}}$ from their means into equation (8.5) gives the specific functional form for the submodels that describe the class/vintages of the household's vehicles, given the number of vehicles the household will choose to own. This is the same as equation (8.5) with $V_{c_n|n}$ replaced by a term that is linear in parameters.

One final note is required to complete the discussion on model specification. There is no submodel in our estimated system that described the make/model of each of the household's vehicles; that is, equation (8.6) was not estimated. No such submodel was included for two reasons. First, forecasting vehicle demand with a model that includes a make/model component would necessitate developing forecasts of the characteristics of each future make and model of vehicle. This was considered to be infeasible, particularly for long run analysis. Second, the computations required to forecast with a make/model component are very burdensome because of the large number of makes and models available. As stated in section 7.2, previous applications of make/model models adopted several ad hoc forecasting procedures in order to keep these computations at a manageable level.

The household's choice of make and model of vehicle cannot be ignored, however, without biasing the results obtained in the class/vintage submodel. Recall that the class/vintage submodel (equation (8.5)) includes as an explanatory variable a term that reflects the utility that the household will obtain with its choice of make/model. This is the term J_{c_n} given by equation (8.7). If J_{c_n} is correlated with any other explanatory variable in the model, omitting it will bias estimates of the coefficients of the other variables; its inclusion prevents this bias.

Calculating J_{c_n} requires a make/model submodel (since, by equation (8.7), J_{c_n} is a function of the terms entering the choice probabilities for makes and models). McFadden (1978) has shown, however, that under fairly general assumptions J_{c_n} can be approximated by terms that are more directly observable. In particular, as the number of makes and models in a class/vintage grows large

$$J_{c_n} \to \ln(r_{c_n}) + \tfrac{1}{2}W_{c_n}^2, \tag{8.13}$$

where

r_{c_n} is the number of makes and models in class/vintage c_n, expressed as a proportion of the number of all makes and models; and
$W_{c_n}^2$ is the variance of $\bar{V}_{m_{c_n}|nc_n}$ around $\bar{V}_{c_n|n}$.

This approximation for J_{c_n} is adopted in the class/vintage submodel.

8.3 Data

The model was estimated on the sample of households that constitute the National Transportation Survey (NTS). This survey, administered by Cambridge Systematics, Inc., and Westat, Inc., and funded by the National Science Foundation, consists of 1,095 households contacted between mid-May and the end of June 1978. Households were sampled nationwide using a stratified random sampling procedure. The following information was obtained from each household:

- socioeconomic data, including household income and the age, sex, employment status, and education level of each household member;
- an inventory of the household's vehicles, specifying the make, model,

vintage, and date purchased of each vehicle currently owned;
- an estimate of the number of miles driven in each of the household's vehicles during the previous twelve months;
- information on whether the household had sold any vehicles during the previous twelve months, and, if so, the make, model, and vintage of each vehicle sold; and
- a one-day trip "diary" specifying the length, vehicle used, and purpose of each trip taken by a household member during the previous twenty-four hours.

Details of these data are described in Cambridge Systematics, Inc. (1980a).

The NTS household sample was augmented with data on the characteristics of more than 2,000 makes and models of 1967–1978 vintage vehicles. Data on front and rear shoulder room, luggage capacity, acceleration rate, braking distance, interior noise level, and repair record were obtained from *Consumer Reports*. Vehicle weight and turning radius were obtained from *Automotive News Almanac*. Prices for new vehicles were obtained from *Consumer Reports*, for used cars from the *Red Book*, and for used trucks from the *Used Truck Buyers Guide*. Gas mileage was obtained, when possible, from annual editions of *EPA Gas Mileage Guide for New Car Buyers*. Gas mileage of vehicles for which EPA figures were not available (particularly vintages prior to 1974, the year in which EPA began publishing such figures) were obtained from *Consumer Reports*; the *Consumer Reports* figures were converted to EPA-equivalent figures with an equation estimated on vehicles for which both EPA and *Consumer Reports* figures were available. Operating cost for each household was calculated using the gas price in the household's region (obtained from the *Oil and Gas Journal*) and the gas mileage, in miles per gallon, of the household's vehicle.

Complete details on the vehicle characteristics, particularly the way in which missing data were handled for each characteristic, are given in appendix A of Cambridge Systematics, Inc. (1980b).

Finally, for each NTS household, data were obtained on the number of public transit trips taken in the metropolitan area in which the household resides, and the population of the area. These data, collected from *National Urban Mass Transit Statistics*, were used to calculate the number of public transit trips taken per capita in the household's area of residence. This figure is used in the model as a measure of the quality of public transit service in the household's area of residence.

8.4 Estimation Results

As depicted in figure 8.1, seven submodels make up the complete model system. These submodels describe

- the number of vehicles each household chooses to own;
- the choice of class/vintage of vehicle for households choosing one vehicle;
- the choice of class/vintage of each vehicle for households choosing two vehicles;
- the number of miles traveled annually for households choosing one vehicle;
- the number of miles traveled annually by each vehicle for households choosing two vehicles;
- the proportion of miles traveled in each category of travel for households choosing one vehicle; and
- the proportion of miles traveled in each category of travel in each vehicle for households choosing two vehicles.

The estimation of each of these seven submodels is now described.

1 Vehicle Quantity Submodel

The vehicle quantity model calculates the probability that a household will choose to own a certain number of vehicles. This probability is given succinctly by equation (8.4). However, a more fully articulated description and interpretation is helpful.

A household is assumed to have a choice of owning zero, one, or two vehicles.[2] The probability of owning each number of vehicles depends on factors that reflect the household's need for vehicles and its willingness or ability to purchase vehicles. Let V_1 denote the weighted sum of factors that reflect a household's need or willingness to own one vehicle: $V_1 = \beta_1 z_1$, where z_1 is a vector of variables and β_1 is a vector of parameters to be estimated. Similarly, let V_2 denote the aggregate of factors that reflect a household's need or willingness to own two vehicles: $V_2 = \beta_2 z_2$, with analogous definitions. Given V_1 and V_2, the probability of owning one vehicle is specified as logit, namely[3]

$$P_1 = \frac{e^{V_1}}{e^{V_1} + e^{V_2} + 1}.$$

Note that as the need or willingness to purchase one vehicle, as captured in V_1, increases, the probability of owning one vehicle, P_1, also increases. The probability of owning two vehicles is

$$P_2 = \frac{e^{V_2}}{e^{V_1} + e^{V_2} + 1},$$

such that P_2 increases with V_2. The probability of owning no vehicles is simply 1.0 minus the probability of owning one or two vehicles:

$$P_0 = 1 - P_1 - P_2 = \frac{1}{e^{V_1} + e^{V_2} + 1},$$

such that when either V_1 or V_2 increases, P_0 decreases.

Among the variables entering V_1 and V_2 is one that represents the average utility that the household obtains in its choice of class and vintage of vehicle. (This is the term I_n in equation (8.4).) It is through this variable that interdependence of the vehicle quantity choice and the class/vintage choice is captured. Consequently, its inclusion is particularly important for the overall model specification.

Table 8.1 presents the variables entering the submodel and their estimated parameters, with the estimates obtained by the maximum likelihood method described in section 2.6.[4] Since factors that affect a household's need or willingness to own one vehicle also usually affect its need or willingness to own two vehicles, each explanatory variable enters both V_1 and V_2. Consequently, each variable is listed twice in the table, with a separate coefficient, or weight, estimated for owning one and two vehicles. With this information, the values of V_1 and V_2 that are implied by table 8.1 can be written explicitly as

$$\begin{aligned} V_1 = {} & 1.05 \text{ (log of household income)} + 1.08 \text{ (number of workers)} \\ & + .181 \text{ (number of members)} \\ & - .0009 \text{ (annual transit trips per capita in area)} \\ & + .635 \text{ (average utility in class/vintage choice given one vehicle)} \\ & - 1.79 \end{aligned}$$

and

Table 8.1
Vehicle quantity submodel

Explanatory variable	Estimated coefficient	t-statistic
1 Log of household income, entering one-vehicle alternative	1.05	3.69
2 Log of household income, entering two-vehicle alternative	1.57	3.52
3 Number of workers in household, entering one-vehicle alternative	1.08	3.78
4 Number of workers in household, entering two-vehicle alternative	1.50	4.78
5 Log of number of members in household, entering one-vehicle alternative	.181	.43
6 Log of number of members in household, entering two-vehicle alternative	.197	.39
7 Annual number of transit trips per capita in household's area of residence, entering one-vehicle alternative	−.0009	1.82
8 Annual number of transit trips per capita in household's area of residence, entering two-vehicle alternative	−.0021	3.42
9 Average utility in class/vintage choice	.635	7.14
10 Alternative-specific constant for one-vehicle alternative	−1.79	2.97
11 Alternative-specific constant for two-vehicle alternative	−4.95	5.19

a. Model: multinomial logit, fitted by maximum likelihood method. Alternatives: (1) no vehicles, (2) one vehicle, (3) two vehicles. Number of observations: 634. Log likelihood at zero: −700.23. Log likelihood at covergence: −475.03.

$$V_2 = 1.57\ (\text{log of household income}) + 1.50\ (\text{number of workers})$$
$$+ .197\ (\text{number of members})$$
$$- .0021\ (\text{annual transit trips per capita in area})$$
$$+ .635\ (\text{average utility in class/vintage choice given two vehicles})$$
$$- 4.95.$$

Given the values of V_1 and V_2, the probability of owning zero, one, or two vehicles is calculated using these formulas.

The variables entering V_1 and V_2 and the estimated values of their coefficients are interpreted as follows. The first two variables are the log of the household's income, entering the alternatives of one and two vehicles separately. The estimated coefficients of these variables are positive, and the coefficient for the two-vehicle alternative is larger than that for the one-vehicle alternative. This indicates, as expected, that an increase in a household's income will increase its probability of choosing two vehicles over one and its probability of choosing one vehicle over none. Since it is the log of income that enters rather than income itself, increasing a household's income changes the relative probability of owning a particular number of vehicles less for higher income households than for lower income households.

The third through eighth variables in the model capture the effects on vehicle quantity demand of the number of workers in a household, household size, and the quality of transit in the household's area of residence. Instead of measuring directly the quality of transit for each household (which would be essentially impossible), a proxy variable is used. This variable is the number of transit trips taken per capita in the household's area of residence. Insofar as more people ride transit in areas where transit quality is good than in areas of relatively poor transit service, this proxy is a reasonable measure of transit quality.

The estimated coefficients of the third through eighth enter with the expected signs and relative magnitudes. The estimated parameters indicate that

- an increase in the number of workers in a household increases the probability of choosing two vehicles over one and the probability of choosing one vehicle over none;
- an increase in household size increases the probability of choosing two

vehicles over one and of choosing one vehicle none; furthermore, the increase in these probabilities is less for larger households than smaller households; and
• an increase in the quality of transit in a household's area increases the probability of choosing one vehicle over two and the probability of choosing no vehicles over one.

The ninth explanatory variable is the term I_n discussed previously; it is a "feedback" variable from the class/vintage submodels and reflects the interdependence of the household's choice of how many vehicles to own with its choice of what class and vintage of vehicles to own. The variable enters with te anticipated sign and a large t-statistic, indicating its importance in explaining household's choices of how many vehicles to own.

The last two variables in the submodel are alternative-specific constants. The coefficients of these variables are chosen by the maximum likelihood method so that predicted shares of households in the three alternatives equal the actual shares in the estimation sample. Their signs and magnitudes have little interpretable content.

2 Class/Vintage Submodel for One-Vehicle Households

This submodel calculates, for each class and vintage of vehicles, the probability that a household that owns one vehicle will choose to own that class/vintage. The specification is given by equation (8.5) with $n^* = 1$, and is described more thoroughly in the following.

Denote V_i as a weighted sum of factors affecting the desirability to the household of owning a vehicle of class and vintage combination i (e.g., the desirability of a 1976 subcompact vehicle). This can be written as

$$V_i = \beta z_i,$$

where

z_i is a vector of characteristics of vehicles in class/vintage i and characteristics of the household; and
β is a vector of parameters to be estimated.

The probability that the household will choose to own a vehicle in class/vintage i is specified to be logit:

$$P_i = e^{V_i}/\sum_j e^{V_j},$$

where the summation in the denominator is over all possible classes and vintages of vehicles.

It is important to note that V_i includes variables that reflect the average utility that the household obtains in its choice of make and model of vehicle given that it chooses class/vintage i. (These are the variables $\ln(r_{c_n})$ and $(W_{c_n}^2)$ in equation (8.13), which are used to approximate J_{c_n} in equation (8.5).) These variables capture the interdependence of the class/vintage choice with the choice of a particular make and model of vehicle. Stated another way: Through these variables, the model incorporates the fact that a household chooses a make and model of vehicle within a class/vintage and chooses the class/vintage with knowledge of the makes and models within it.

As discussed in section 8.2, the exact function for the average utility associated with the make and model choice cannot be calculated without estimating a submodel of make/model choice. To avoid such estimation, the exact function is approximated by variables that McFadden (1978) has shown are arbitrarily close to the function when the number of makes and models of vehicles in each class/vintage is large. In particular, these variables are the number of makes and models within the class/vintage (expressed in logs) and the variance in the characteristics of makes and models around the average characteristics for the class/vintage (e.g., the variance of shoulder room around its average for makes and models within a given class/vintage).

For estimation, each household was assumed to have a choice among 12 classes of vehicles for each of 10 vintages, making a total of 120 alternatives from which to choose. The classes of vehicles are

1. subsubcompact domestic car (e.g., Chevette, Fiesta);
2. subcompact domestic car (e.g., Sunbird, Maverick);
3. compact domestic car (e.g., Firebird, Malibu);
4. midsize domestic car (e.g., Fairmont, Granada);
5. standard domestic car (e.g., Impala, Delta 88);
6. large domestic car (e.g., Olds 98, Lincoln);
7. small import car (e.g., Datsun 210, Toyota Corolla);
8. medium size import car (e.g., Datsun 510, Audi Fox);
9. luxury import car (e.g., all Mercedes, Datsun 810);
10. pickup trucks;
11. vans; and
12. utility vehicles (e.g., Jeep, Jimmy).

The 10 vintages are pre-1970, and the years 1970 through 1978.

For each class and vintage, the mean and variance of each vehicle characteristic were calculated over all makes and models in the class and vintage. In addition, the number of makes and models in each class/vintage was tabulated. These variables, along with socioeconomic characteristics of the household, entered the submodel.

Because of the large number of alternatives, estimation of this submodel on the full set of alternatives was considered infeasible. Consequently, a subset of the alternatives was employed for estimation. Fifteen alternatives were selected for each one-vehicle household. These alternatives included the household's chosen alternative (that is, the class/vintage that the household actually owned) and the alternative(s) that the household had chosen in the previous year (that is, the class/vintage(s) that the household owned in the previous year). If the household neither sold nor bought any vehicles during the year, then these two alternatives were the same, and that alternative was included only once. The remaining alternatives for each household were selected randomly.

This procedure of estimating on a subset of alternatives results, as described in section 2.6, in consistent estimates. Furthermore, tests indicate that, beyond a minimal number of alternatives, the estimated parameters are not sensitive to the number of alternatives included in estimation. A submodel with a given set of explanatory variables was estimated on fifteen alternatives and then separately on forty alternatives; essentially the same parameter estimates were obtained in both cases. (This result was also found by Cambridge Systematics, Inc., 1980b).

Table 8.2 presents the variables entering the submodel and their estimated parameters. For each class and vintage of vehicles, representative utility V_i can be written explicitly, using the information in table 8.2, as

$$
\begin{aligned}
V_i = {} & -.000380\,\begin{pmatrix}\text{average purchase price of}\\ \text{vehicles in class/vintage}\end{pmatrix} \times \begin{pmatrix}\text{dummy equaling one if}\\ \text{household's income is below}\\ \text{\$12,000, and zero otherwise}\end{pmatrix}\\
& -.00283\,\begin{pmatrix}\text{average purchase price}\\ \text{of vehicles in class/}\\ \text{vintage}\end{pmatrix} \times \begin{pmatrix}\text{dummy equaling one if}\\ \text{household's income is above}\\ \text{\$12,000, zero otherwise}\end{pmatrix}\\
& -.3209\,\begin{pmatrix}\text{average operating cost}\\ \text{of vehicles in class/}\\ \text{vintage i}\end{pmatrix} - 3.63\,\begin{pmatrix}\text{transaction cost}\\ \text{dummy for vehicles in}\\ \text{class/vintage}\end{pmatrix}\\
& + \text{other variables listed in table 8.2.}
\end{aligned}
$$

Table 8.2
Class/vintage submodel for one-vehicle households[a]

Explanatory variable	Estimated coefficient	t-statistic
1 Purchase price in dollars, for households with incomes less than or equal to $12,000	−.000380	2.50
2 Purchase price in dollars, for households with incomes greater than $12,000	−.000283	2.06
3 Operating cost, in cents per mile	−.3209	1.55
4 Transaction cost dummy	−3.63	19.3
5 Front and rear shoulder room, in inches, for households with three or fewer members	.0228	.704
6 Front and rear shoulder room, in inches, for households with more than three members	.0359	1.05
7 Luggage space, in cubic feet, for households with three or fewer members	.0447	1.28
8 Luggage space, in cubic feet, for households with more than three members	.1033	1.93
9 Horsepower, for households with incomes greater than $25,000	.0149	1.78
10 Log of the number of makes and models in the class/vintage (expressed as a proportion of all makes and models)	.544	3.40
11 Variance of front and rear shoulder room around its mean for the class/vintage	.00186	1.95
12 Foreign car dummy	−.472	1.41
13 Vintage 1976–1978 dummy	1.24	2.12
14 Vintage 1972–1975 dummy	.60	2.09
15 Vintage 1976–1978 dummy for households with incomes greater than $12,000	.916	1.50
16 Pickup truck dummy	.639	3.95
17 Van dummy	.380	.234

a. Model: multinomial logit, fitted by maximum likelihood method. Alternatives: class/vintages. Number of observations: 274. Log likelihood at convergence: −371.67.

Given the value of V_i for each class/vintage, the probability that the household will own a vehicle of any particular class/vintage is calculated using the formula for P_i stated previously.

The variables entering V_i and the estimated values of their coefficients are interpreted as follows. The first two variables relate to the average purchase price of vehicles in the class/vintage. Two purchase-price variables are entered to allow for the possibility that lower income households are more affected by price than higher income households. The first price variable is the average price of vehicles in the class/vintage for households with incomes less than or equal to $12,000 per year in 1978 dollars (and zero for higher income households). The second variable is the average price of vehicles in the class/vintage for households with incomes greater than $12,000 (and zero for lower income households). Annual incomes of $12,000 was used to divide the households because that was the average income of one-vehicle households in the sample.

The estimated parameters for these two variables have the expected signs and relative magnitudes. The negative coefficients for these price variables indicate that an increase in the price of vehicles in a class/vintage decreases a household's probability of choosing that class/vintage (all other things held constant). The coefficient for the higher income households is lower than that of lower income households, indicating that an increase in the price of vehicles in a class/vintage decreases the probability of choosing the class/vintage more for lower income households than for higher income households.

The next variable in the submodel is the average operating cost of vehicles in the class/vintage. Its negative coefficient indicates that an increase in the operating cost of vehicles in a class/vintage decreases the probability that a household will choose that class/vintage (all other things held constant).

The ratio of the operating cost coefficient to the purchase price coefficient is a measure of the amount that a household would be willing to pay, in the form of a higher purchase price, in return for a one-cent per mile reduction in operating cost.[5] For lower income households, this is $844. That is, lower income households would be willing to pay up to $844 in extra purchase price for a one-cent per mile reduction in operating cost. For higher income households, the figure is $1,134.

To evaluate these figures, consider the following. If a household were rational (in the economist's sense of the word) and considered only its

Table 8.3
Number of miles traveled annually at which model's estimate of willingness to pay for reduced operating costs is consistent with "rational" behavior

Discount rate r	Real growth in gas prices g	Life of vehicle in years L	Implied miles traveled annually (M)	
			Lower income households	Higher income households
.03	.00	10	9,774	13,125
.03	.02	10	8,870	11,911
.03	.00	8	11,872	15,942
.03	.02	8	10,981	14,745
.05	.00	10	10,730	14,408
.05	.02	10	9,772	13,122
.05	.00	8	12,807	17,197
.05	.02	8	11,872	15,942

discount rate in real terms (labeled r), the real growth rate in gas prices (g), the life of the vehicle (L), and the miles traveled each year (M) in its calculation of life cycle costs, then it would evaluate the present value of the life cycle costs of a vehicle to be

$$P + \frac{1 - e^{(r-g)L}}{r - g} M \cdot C,$$

where P is the purchase price of the vehicle and C is its operating cost per mile (both in dollars). If $\beta_1 P + \beta_2 C$ is estimated, then, given the expression above for life cycle costs, the number of miles traveled annually must satisfy

$$M = (\beta_2/\beta_1)((r - g)/(1 - e^{-(r-g)L})).$$

For lower income households, the ratio of the estimated coefficients of the operating cost and purchase price terms in table 8.2 is $(-.3209 \times 100)/-.000380 = 84{,}447$, where multiplication by 100 reflects the fact that operating cost enters the submodel in cents per mile rather than dollars per mile. The analogous figure for higher income households is 113,392. For various values of r, g, and L, the number of miles that the household must travel per year in order for the above equation to be satisfied is given in table 8.3.

These figures are somewhat high, though not greatly so. This implies that households either (1) have lower discount rates and/or expect higher growth rates for gas prices than indicated in the chart, (2) are rational but calculate life cycle costs differently than was assumed (e.g., they might consider the resale value of the vehicle after a certain number of years rather than its useful life and think that resale value increases with fuel efficiency), or (3) are not rational in the economists's sense of the word or, equivalently, do not know how to make the necessary present value calculations (e.g., they might think that operating cost reductions translate into greater dollar savings than they actually do).

The fourth explanatory variable is a transaction cost dummy. This variable takes the value of one for vehicles in a class/vintage that the household did not own in the previous year and zero for a class/vintage of the vehicle that the household did own in the previous year. Assuming that no household sells a vehicle within a class/vintage and buys another one in the same class/vintage, this variable represents the psychic, search, and other transaction costs associated with buying a new vehicle. As its coefficient and t-statistic indicate, these factors are very important in a household's choices.[6]

The fifth and sixth variables relate to the average interior dimensions of the vehicles in each class/vintage. Interior dimension is measured as front and rear shoulder room (in inches). Two variables were entered to allow for the possibility that large households are more concerned with interior dimensions than smaller households. The fifth variable is equal to the average shoulder room of vehicles in the class/vintage for households with three or fewer members (and zero for larger households). The sixth variable is equal to the average shoulder room of vehicles in the class/vintage for households with more than three members (and zero for smaller households).

The estimated parameters for these two variables also have the expected signs and relative magnitudes. The positive coefficients indicate that an increase in the shoulder room of vehicles in a class/vintage will increase a household's probability of choosing the class/vintage (all other things held constant). The coefficient for the larger households is larger than that of smaller households, indicating that an increase in the shoulder room of vehicles in a class/vintage increases the probability of choosing the class/vintage more for large households than for smaller households.

The seventh and eighth variables capture the effect of luggage space. As with shoulder room, two variables are entered to allow for the possibility that larger households are more concerned about luggage space than are smaller households. The estimated coefficients of these variables are positive, indicating that an increase in the luggage capacity of vehicles in a class/vintage increases a household's probability of choosing the class/vintage. Again, households with more than three members have a larger coefficient than smaller households, and hence an increase in luggage space increases a larger household's probability of choosing the class/vintage more than it does a smaller household's probability.

The ninth variable is the average horsepower of vehicles in each class/vintage for households with incomes over $25,000, and zero for households with lower incomes. This specification is based on the hypothesis that horsepower is a luxury for which high income households are willing to pay, but not low and middle income households. The positive coefficient of this variable indicates that an increase in the horsepower of vehicles in a class/vintage (with all other things held constant, including, significantly, the size of the vehicles as represented by shoulder room and luggage space) will increase the probability that high income households will choose the class/vintage. Furthermore, the ratio of the horsepower coefficient to the price coefficient indicates that high income households are willing to pay $53 for an extra unit of horsepower.

The tenth and eleventh variables capture the interdependence of a household's choice of class/vintage and its choice of make/model within the class/vintage. Recall that the specification of the class/vintage submodel (equation (8.5)) includes the term J_{c_n}, the utility that a household obtains in its choice of make/model, given its choice of class/vintage. However, since a submodel of make/model choice is not being estimated, this term is being approximated, as discussed previously (see equation (8.13)), by

1. the term $\ln(r_{c_n})$, where r is the number of makes and models in the class/vintage, expressed as a percentage of all makes and models (the tenth variable), and
2. variables that capture the variance of the utility of makes and models within a class/vintage around the mean utility for the class/vintage.

A number of variables were originally included in the approximation to the variance of the utility of makes and models. In particular, the variance of each vehicle characteristic was originally included in the submodel. However, the only variance found to have a significant positive coefficient[7,8] was that of shoulder room. This is the eleventh variable in the final submodel.

The importance of these last two variables is indicated by two things. First, the coefficient of $\ln(r_{c_n})$ is an estimate of θ, a parameter in the cumulative distribution function of the errors. To assure consistency with utility-maximization theory for all values of the explanatory variables, this term must be between zero and one (see McFadden, 1978). The estimated parameter is therefore within this consistent range. Second, the submodel was reestimated without the tenth and eleventh variables. The estimated coefficients of other variables changed considerably from those in table 8.2, indicating that ignoring the interdependence of the household's make/model choice and its class/vintage choice may seriously bias estimation results.

The remaining six variables are dummies for particular classes and vintages. The foreign car dummy (which takes the value of one for any class/vintage consisting of foreign vehicles and zero otherwise) enters with a negative coefficient. This indicates that (all other things held constant) households would prefer domestic vehicles to foreign vehicles. This implies that the penetration of foreign vehicles in 1978 was due to their characteristics (e.g., operating cost and price) being considered superior to those of domestic vehicles, rather than a proforeign bias. Furthermore, the negative coefficient of the foreign vehicles dummy perhaps reflects the difficulty of repairing and servicing foreign vehicles, rather than any antiforeign bias per se.

Two vintage dummies (the thirteenth and fourteenth variables) were entered to allow for the possibility that households prefer newer vehicles to older ones. The coefficient for the 1976–1978 vintage dummy is greater than that for the 1972–1975 vintage dummy, indicating that this is so. Furthermore, both these vintage dummies are positive, indicating that vehicles built after 1971 are preferred to those built in or prior to 1971.

Another vintage dummy for 1976–1978 vehicles was entered for households with incomes over $12,000. This specification reflects the hypothesis that a new vehicle is a luxury for which higher income households are more willing to pay than lower income households, and the positive estimated coefficient supports this.

The last two variables are dummies indicating that the class/vintage consists of pickup trucks and vans, respectively. The coefficients of these variables measure the difference in the utility that a household obtains from pickups and vans after all the other factors in the model have been accounted for. Since vehicle characteristics have different meanings for trucks and cars (e.g., shoulder room and luggage space), the positive coefficients of these dummy variables should not be interpreted as implying that pickup trucks and vans are preferred to cars. A direct comparison of utilities would be possible only if all characteristics (including luggage space) had the same meaning for cars and trucks, which they do not.

3 Class/Vintage Submodel for Two-Vehicle Households

A household that owns two vehicles has a choice of class and vintage for each of its two vehicles. Stated differently, a two-vehicle household has a choice of which **pair** of class/vintages to own. The third submodel calculates, for each pair of class/vintages, the probability that a household that owns two vehicles will choose to own that particular pair of class/vintages.

The submodel is specified similarly to that for class/vintage choice by one-vehicle households, except that now **pairs** of class/vintages are considered. In particular, the probability of owning pair i of class/vintages is specified to be logit:

$$P_i = \exp(V_i)/\sum_j \exp(V_j),$$

where $V_i = \beta z_i$ is a linear-in-parameters function of factors reflecting the desirability of class/vintage pair i, and the summation in the denominator is over all pairs of class/vintages.

Table 8.4 presents the list of variables reflecting the desirability of each class/vintage pair (i.e., elements of z_i) and the estimated coefficients of these variables (i.e., elements of β). As with the class/vintage submodel for one-vehicle households, estimation was performed on a subset of fifteen alternatives for each household. This subset included the class/vintage pair that the household owned in the current year and any class/vintage pair(s) that the household owned in the previous year (if different from the current class/vintage pair).

The explanatory variables of the model are similar to those of the one-vehicle submodel; however, the following differences should be noted:

1. Most explanatory variables (all except the ninth and thirteenth variables) are defined as the sum of the characteristics over both class/vintages in the pair. For example, the price variables are defined as sum of the average price of vehicles in one class/vintage in the pair **plus** the average price of vehicles in the other class/vintage. This specification implies that households are equally concerned with the characteristics of both of its

Table 8.4
Class/vintage submodel for two-vehicle households[a]

Explanatory variable	Estimated coefficient	t-statistic
1 Purchase price of both vehicles, summed in dollars, for households with incomes less than or equal to $12,000	−.000531	2.09
2 Purchase price of both vehicles, summed in dollars, for household with incomes greater than $12,000 and less than or equal to $20,000	−.000383	2.28
3 Purchase price of both vehicles, summed in dollars, for households with incomes greater than $20,000	−.0001713	1.29
4 Operating cost of both vehicles, summed in cents per mile, for households with incomes less than or equal to $20,000	−.441	1.96
5 Operating cost of both vehicles, summed in cents per mile, for households with incomes greater than $20,000	−.330	1.35
6 Number of transactions required to obtain pair of vehicles	−4.48	13.5
7 Front and rear shoulder room of both vehicles, summed in inches, for households with three or fewer members	.0370	1.47
8 Front and rear shoulder room of both vehicles, summed in inches, for households with more than three members	.0533	2.09
9 Expected absolute difference in shoulder room of two vehicles, in inches	.0240	2.01
10 Horsepower for both vehicles, summed for household with incomes greater than $25,000	.00954	1.43
11 Log of the number of pairs of makes and models in the class/vintage pair (expressed as a proportion of all possible pairs of makes and models)	.307	1.70
12 Number of foreign cars in pair	−.662	1.64
13 Dummy indicating at least one class/vintage in the pair is "prestigious" (see text for classification of prestigious vehicles)	1.20	2.50
14 Number of vintage 1976–1978 vehicles in pair	.155	.167
15 Number if vintage 1972–1975 vehicles in pair	.931	2.84
16 Number of vintage 1976–1978 vehicles in pair, for households with incomes greater than $12,000	1.35	1.49
17 Number of pickup trucks in pair	2.05	1.70
18 Number of vans in pair	.679	.56
19 Number of utility vehicles in pair	−2.89	3.26

a. Number of observations: 241. Log likelihood at convergence: −130.55.

vehicles. For example, an extra dollar in price is considered just as onerous to the household whether it occurs in one vehicle or the other. For situations in which this does not seem reasonable other specifications are used.
2. The price of a class/vintage pair was entered separately for three income groups: households with incomes below $12,000, those with incomes between $12,000 ad $20,000, and those with incomes over $20,000. This grouping expands on that used in the one-vehicle submodel by differentiating households with incomes over $20,000 from those with incomes between $12,000 and $20,000. (Not enough one-vehicle households had incomes in excess of $20,000 to allow this differentiation in the one-vehicle submodel.) As expected, the price coefficients are successively smaller in magnitude for successively higher income groups.
3. Similar to differentiation in price response for households with incomes over $20,000, operating cost was entered separately for households with incomes over $20,000 and for households with incomes less than $20,000. As expected, the magnitude of the operating cost coefficient is smaller for high income households that for low income households.

The ratios of operating cost coefficients to price coefficients (recall that these are measures of the amount that a household would be willing to pay for a one-cent per mile reduction in operating cost) are similar to those obtained in the one-vehicle submodel. For two-vehicle households with incomes less than $12,000, the figure is $830; for one-vehicle households the figure was $844. Two-vehicle households with incomes between $12,000 and $20,000 are willing to pay an estimated $1,151; the comparable figure for one-vehicle households was $1,134. Two-vehicle households with incomes over $20,000 are estimated by the model to be willing to pay as much as $1,926 for a one-cent per mile reduction in operating cost. (There is no comparable figure for one-vehicle households since, as stated, few one-vehicle households had incomes in excess of $20,000.)
4. The specification of the shoulder room variables reflects the hypothesis that two-vehicle households would like to own vehicles of different sizes, for example, a large family car and smaller car for commuting. If this is the case, then the utility that a household obtains from two vehicles with shoulder room S_1 and S_2, respectively, is

$$V = \alpha(S_1 + S_2) + \beta|S_1 - S_2|, \qquad \text{with} \quad \alpha > \beta.$$

The first term indicates that a household values extra shoulder room in either vehicle. The second term indicates that, for a given total amount of shoulder room, the household would prefer to have it distributed between the two vehicles in such a way that one vehicle has a lot while the other has comparatively little.

With this specification, a one inch increase in the shoulder room of the larger vehicle in a pair increases the utility that the household obtains from the pair by $\alpha + \beta$. A one inch increase in the shoulder room of the smaller vehicle increases the utility of the household by $\alpha - \beta$. The requirement that α exceed β ensures that an increase in the shoulder room of either vehicle will increase the household's utility.

While the household's utility depends on the shoulder room of the particular makes and models it chooses, variables entering the class/vintage submodel are averages over the makes and models within each class/vintage. For the first term in the preceding displayed equation, the average over pairs of makes and models within the two class/vintages is simply the sum of the average shoulder room for each class/vintage in the pair. However, for the second term, the average is **not** equal to the absolute value of the difference in the averages. That is,

$$E_{ij \in c^1 \times c^2}(S^i + S^j) = E_{i \in c^1}(S^i) + E_{j \in c^2}(S^j)$$

but

$$E_{ij \in c^1 \times c^2}|S^i - S^j| \neq |E_{i \in c^1}(S^i) - E_{j \in c^2}(S^j)|,$$

where

E is the expectation operator;
S^i is the shoulder room of make/model i; and
c^1, c^2 are the sets of makes and models in class/vintages 1 and 2, respectively.

However, under the assumption that shoulder room is normally distributed among makes and models within each class, it can be shown that[9]

$$E_{ij \in c^1 \times c^2}|S^i - S^j| = (S^i - S^j)\left(\Phi\left(\frac{S^i - S^j}{\sigma}\right) - \Phi\left(\frac{S^j - S^i}{\sigma}\right)\right) + 2\sigma \times \phi\left(\frac{S^i - S^j}{\sigma}\right), \tag{8.14}$$

where

Φ is the cumulative standard normal distribution;
ϕ is the probability density function for a standard normal deviate; and
σ is the variance of $(S^i - S^j)$.

The ninth variable in the submodel is this expression, with σ calculated under the assumption that the covariance of shoulder room in different class/vintages is zero.

The estimated coefficients of the shoulder room variables (the seventh through ninth variables) enter with the expected sign and with the appropriate relative magnitude. In particular, the coefficient for the sum of shoulder room for households with more than three members is larger than that for households with three or fewer members. Both of these coefficients are larger than the coefficient for the expected difference in shoulder room, which ensures that an increase in the shoulder room of either vehicle will increase the probability of a household choosing the class/vintage pair. The expected difference in shoulder room enters significantly, and omitting this variable substantially changes the estimated coefficients of other variables. We take this to indicate that the variable is important in explaining the choice of two-vehicle households.

5. Luggage space does not enter the submodel for two-vehicle households. In an initial specification, luggage space variables entered with negative coefficients and very small t-statistics. Since these households have two vehicles for carrying luggage, the result that they are not willing to pay for extra luggage space is perhaps not surprising.

6. The variance of shoulder room does not enter the submodel. In an initial specification, this variable entered with a negative sign and a very small t-statistic; as discussed previously, this variable must enter with a positive sign if it is to serve as an approximation to W^2 in equation (8.13). One reason for these results may be that the variance of shoulder room already enters the model indirectly through the expression for the expected difference in shoulder room (see equation (8.14)).

7. The model includes a dummy variable (the thirteenth variable) that indicates whether either of the class/vintages in the pair is a "prestigious" class/vintage. This variable was not originally included in the submodel; rather, it was suggested by the results of initial specifications. In particular, initial results indicated that households with two vehicles have a strong tendency to own at least one "prestigious" vehicle.

The determination of which class/vintages were considered prestigious was made on the basis of both a priori notions and trial-and-error experi-

Table 8.5
Vehicles designated as "prestigious" (denoted by ×)

Class	Vintage New (1978)	1 year (1977)	2 years (1976)	3 years (1975)	4 years (1974)
1 Subsubcompact domestic car	×				
2 Subcompact domestic car	×				
3 Compact domestic car	×	×			
4 Midsize domestic car	×	×			
5 Standard domestic car	×	×	×		
6 Large domestic car	×	×	×	×	
7 Small import car	×				
8 Medium size import car	×	×	×		
9 Luxury import car	×	×	×	×	×
10 Pickup trucks	×	×			
11 Vans	×	×	×		
12 Utility vehicles	×	×	×	×	

mentation. The final result was that the class/vintages in table 8.5 were classified as prestigious. This classification follows a definite pattern: The prestige of a vehicle is related to both how new it is and how large it is. All new vehicles are prestigious, and larger vehicles (e.g., classes 6 and 9) retain their prestige longer than smaller vehicles. This dummy variable (the thirteenth variable) enters significantly with a large, positive coefficient. Furthermore, omitting the variable substantially changes the estimated coefficients of other variables, and hence we conclude that this variable is relevant in explaining the choices of two-vehicle households.[10]

4 Annual VMT Submodel for One-Vehicle Households

This submodel predicts the number of miles traveled annually by one-vehicle households. It is specified as a regression equation of the form

$$\log(\text{VMT}) = \beta z,$$

where VMT is vehicle miles traveled annually, z is a vector of explanatory variables, and β is a vector of parameters.

The parameters were estimated with an instrumental variables approach. This approach was required, rather than ordinary least squares, because the regression equation includes the operating cost of the household's vehicle as an explanatory variable, representing the cost of travel to the household. Since a household chooses which vehicle it owns, it effectively chooses the operating cost that it faces when driving, namely, the operating cost of its chosen vehicle. Therefore, the operating cost that a household faces is an endogenous variable, so that estimation with ordinary least squares is biased. To avoid this endogeneity bias, instrumental variables estimation was applied.[11] The exogenous variables used to predict operating cost are

gas price in area of residence;
household income;
household size;
type of housing unit;
population of household's area of residence;
number of transit trips in area of residence;
number of adults, adolescents;
number of workers;
age of household head;
education level of household head;
sex of household head; and
distance to work.[12]

Table 8.6 presents the estimated submodel of annual VMT for one-vehicle households. Increases in household income, household size, or the number of workers in the household increase the expected number of miles that a household drives annually. An increase in the operating cost of a household's vehicle or the quality of transit in the household's area (as measured by the number of transit trips per capita) decreases the amount that the household expects to drive. Households in large urban areas drive more, all other things held constant, than households in small urban areas, and those in small urban areas drive more than those in rural areas. The geographic dummies for the northeastern, midwestern, and southern United States indicate that households in the western United States (the omitted region) travel more than comparable households in other parts of the United States.

5 Annual VMT Submodel for Each Vehicle for Two-Vehicle Households

This submodel describes the number of miles traveled in each vehicle owned by a two-vehicle household. It is specified as a regression equation, with the dependent variable being the log of vehicle miles traveled in each vehicle.

Table 8.6
Submodel for annual vehicle miles traveled for one-vehicle households[a]

Explanatory variable	Estimated coefficient	t-statistic
1 Log of household income, in dollars	0.1406	1.49
2 Operating cost, in cents per mile	−0.2795	2.63
3 Log of household size	0.2131	1.71
4 Number of workers in household	0.17777	1.61
5 Number of transit trips per capita in household's area of residence	−0.000258	0.78
6 Dummy indicating that household lives in an urban area with more than one million population	0.1163	0.377
7 Dummy indicating that household lives in an urban area with less than one million population	0.0477	0.283
8 Dummy indicating household lives in north-eastern United States	−0.179	0.93
9 Dummy indicating household lives in midwestern United States	−0.074	0.40
10 Dummy indicating household lives in southern United States	−0.167	0.89
11 Intercept	8.709	15.4

a. Model: regression equation, fitted by ordinary least squares with instruments replacing operating cost. Dependent variable: log of annual VMT. Number of observations: 226. R-squared: .114.

Table 8.7 presents the estimated submodel. The variables and the interpretations of their coefficients are essentially the same as in the VMT submodel for one-vehicle households. However, the submodel for two-vehicle households contains one additional variable: a dummy indicating whether the vehicle for which VMT is being described is the newer of the two vehicles that a household owns. The positive coefficient for this variable indicates that two-vehicle households drive their newer vehicles more than their older ones.

6 Submodel for the Proportion of VMT in Each Category for One-Vehicle Households

For some forecasting purposes it is useful to have VMT by purpose, or category, of travel. The submodel presented here predicts the proportion of

Table 8.7
Submodel for annual miles traveled in each vehicle for two-vehicle households[a]

Explanatory variable	Estimated coefficient	t-statistic
1 Log of household income, in dollars	0.276	3.70
2 Operating cost, in cents per mile	−0.0351	0.472
3 Dummy indicating that vehicle is the newer of household's two vehicles	0.432	5.16
4 Log of household size	0.0833	0.721
5 Number of workers in household	0.0284	0.456
6 Number of transit trips per capita in household's area of residence	−0.000421	2.20
7 Dummy indicating that household lives in an urban area with more than one million population	0.200	1.06
8 Dummy indicating that household lives in an urban area with less than one million population	−0.0920	0.876
9 Dummy indicating household lives in north-eastern United States	−0.174	1.18
10 Dummy indicating household lives in midwestern United States	−0.107	0.930
11 Dummy indicating household lives in southern United States	−0.648	0.541
12 Intercept	6.27	15.8

a. Model: regression equation, fitted by ordinary least squares with instruments replacing operating cost. Dependent variable: log of annual VMT. Number of observations: 419. R-squared: .117.

a household's VMT in each of four categories of travel:

1. intracity work trips;
2. intracity nonwork trips;
3. non-intracity work trips; and
4. non-intracity nonwork trips.

The submodel is specified as a logit model with four alternatives representing the four categories of travel. Specifically, the proportion of VMT in category i is specified to be

$$P_i = \frac{e^{V_i}}{e^{V_1} + e^{V_2} + e^{V_3} + e^{V_4}},$$

where V_i is a weighted sum of explanatory variables relating to alternative i, with the weights being parameters.

Unlike the previous submodels, the specification of this submodel was not derived in section 8.2. Its specification is not behavioral in that it is not derived from a household's utility maximization. It is also not integrated with the other submodels in the system; that is, there is no representation of the interdependence of a household's choice of how much to drive in each category with its choice of how much to drive in total or its choices of how many and what class/vintage of vehicles to own. Consequently, this submodel should be viewed as an approximation to the "true" behavioral model, an approximation that, unlike the "true" behavioral model, can be of practical use in forecasting with readily available data.[13] However, forecasts derived using this submodel must be viewed with caution.

The parameters of the submodel were estimated on data obtained from the trip diary of each sampled household. As discussed in section 8.3, the trip diary contains information on the length of each trip taken by a household member during a twenty-four hour period. Each trip was classified as being in one of the four categories listed previously, and the VMT summed by category. The proportion of VMT in each category was simply these VMT figures by category divided by the total VMT for all trips in the diary.

The definition of work trip was straightforward: Any trip whose origin or destination was the place of work was considered a work trip. The definition of intracity trips was less clear-cut. Knowing only the origin and destination of a trip (such as home for the origin and a store for the destination) and its length, the question was, What trips could be considered "intracity"? Two criteria were used. First, since intracity trips are those within an urban area and not those from rural areas to urban areas, only households that live in SMSAs were classified as taking intracity trips. All trips by rural households were classified as nonintracity. Second, since intracity trips are not those from urban areas to rural areas or from one urban area to another, and since these trips are generally longer than those within an urban area, only trips of fifteen miles or less (one way) were classified as intracity trips. Any trips over fifteen miles in length were classified as nonintracity. In summary, trips taken by urban households of less than fifteen miles in length (one way) were classified as intracity, and all others classified as nonintracity.

Table 8.8 presents the estimated submodel. Urban households faced all

Table 8.8
Submodel for proportion of VMT in each category for one-vehicle households[a]

Explanatory variable	Estimated coefficient	t-statistic
1 Proportion of household members who work, entering alternative 1	2.26	1.81
2 Proportion of household members who work, entering alternative 3	2.28	4.13
3 Gas price in household's area of residence, in cents per gallon, entering alternative 1	.0120	1.222
4 Gas price in household's area of residence, in cents per gallon, entering alternative 2	.00942	3.634
5 Gas price in household's area of residence, in cents per gallon, entering alternative 3	−.00237	.388
6 Household size, entering alternative 1	.242	.759
7 Household size, entering alternative 2	.0892	.434
8 Household size, entering alternative 3	.242	1.92
9 Alternative-specific constant for alternative 1	−9.88	1.47
10 Alternative-specific constant for alternative 2	−5.46	3.07
11 Alternative-specific constant for urban households for alternative 3	−2.59	4.71
12 Alternative-specific constant for non-urban households for alternative 3	−.790	.195

a. Model: multinomial logit, fitted by maximum likelihood method. Alternatives: (1) intracity work; (2) intracity nonwork; (3) nonintracity work; (4) nonintracity nonwork. Log likelihood at zero: −26,620. Log likelihood at convergence: −16,150.

four alternatives; however, since rural households could not, by definition, take intracity trips, only the third and fourth alternatives were included in estimation for these households.

The variables in Table 8.8 and their estimated coefficients are readily interpretable. The proportion of household members who work enters the work trip alternatives (1 and 3) with positive coefficients, indicating that an increase in the number of workers in a household increases the proportion of the household's total VMT that is driven on work trips.

The price of gas in the household's region[14] enters the first and second alternatives with positive coefficients and enters the third alternative with a negative coefficient. (The coefficient for the fourth alternative is normalized

Table 8.9
Submodel for proportion of VMT in each category for two-vehicle households[a]

Explanatory variable	Estimated coefficient	t-statistic
1 Proportion of household members who work, entering alternative 1	2.40	2.16
2 Proportion of household members who work, entering alternative 3	1.87	3.67
3 Number of transit trips per capita in household's area of residence, entering alternative 1	−.0118	1.135
4 Number of transit trips per capita in household's area of residence, entering alternative 2	−.0127	1.545
5 Gas price in household's area of residence, in cents per gallon, entering alternative 1	.00954	3.46
6 Gas price in household's area of residence, in cents per gallon, entering alternative 2	.0124	2.96
7 Gas price in household's area of residence, in cents per gallon, entering alternative 3	.0023	1.31
8 Household size, entering alternative 1	.154	.583
9 Household size, entering alternative 2	−.294	1.46
10 Household size, entering alternative 3	.242	2.16
11 Alternative-specific constant for alternative 1	−6.40	2.45
12 Alternative-specific constant for alternative 2	−4.60	1.56
13 Alternative-specific constant for urban households for alternative 3	−2.55	4.32
14 Alternative-specific constant for urban households for alternative 3	−4.35	3.201

a. Model: multinomial logit, fitted by maximum likelihood method. Alternatives: (1) intracity work; (2) intracity nonwork; (3) nonintracity work; (4) nonintracity nonwork. Log likelihood at zero: −46,730. Log likelihood at convergence: −29,040.

to zero.) This indicates that an increase in gas price results in a household decreasing its nonintracity trips (alternatives 3 and 4) relative to intracity trips. Furthermore, the relative magnitudes of the gas price coefficients indicate that, as the gas price rises, the household reduces its nonwork trips more than its work trips.

Household size enters the first three alternatives with positive coefficients. (Again the coefficient for alternative 4 is normalized to zero.) The relative magnitudes of the coefficients indicate that, as household size increases **and** the proportion of workers stays constant, the household increases its number of work trips relative to nonwork trips, and increases its intracity nonwork trips relative to nonintracity nonwork trips. This is again a reasonable implication, since an increase in the number of workers in a household (required to keep the proportion of workers constant when household size increases) will generally increase the number of work trips in proportion to the increase in number of workers, but an increase in household size will generally increase the need for shopping trips less than proportionately.

7 Submodel for the Proportion of VMT in Each Category for Two-Vehicle Households

A submodel similar to that described previously was also estimated for households owning two vehicles. Table 8.9 presents the estimation results. The definition of alternatives, the estimation method, variables, and results are essentially the same as for one-vehicle households. The only significant difference is that the submodel includes one variable that was not found to be significant for one-vehicle households. This variable is the number of transit trips taken per capita in the household's area of residence, entering the intracity trip alternatives as a measure of the quality of transit for intracity travel. The estimated coefficient of this variable is negative in each of the intracity trips alternatives, indicating that, as expected, increases in the quality of transit in urban areas decreases the amount that households drive on intracity trips relative to nonintracity trips.