PB2001-103210

# Vehicle Availability Modeling

**July 1998**

TMIP

*Travel Model Improvement Program*

REPRODUCED BY: NTIS
U.S. Department of Commerce
National Technical Information Service
Springfield, Virginia 22161

U.S. Department of Transportation
**Federal Highway Administration**
**Federal Transit Administration**
**Bureau of Transportation Statistics**
**Assistant Secretary for Transportation Policy**

U.S. Environmental Protection Agency

# Travel Model Improvement Program

The Department of Transportation, in cooperation with the Environmental Protection Agency and the Department of Energy, has embarked on a research program to respond to the requirements of the Clean Air Act Amendments of 1990 and the Intermodal Surface Transportation Efficiency Act of 1991. This program addresses the linkage of transportation to air quality, energy, economic growth, land use and the overall quality of life. The program addresses both analytic tools and the integration of these tools into the planning process to better support decision makers. The program has the following objectives:

1. To increase the ability of existing travel forecasting procedures to respond to emerging issues including; environmental concerns, growth management, and lifestyles along with traditional transportation issues,

2. To redesign the travel forecasting process to reflect changes in behavior, to respond to greater information needs placed on the forecasting process and to take advantage of changes in data collection technology, and

3. To integrate the forecasting techniques into the decision making process, providing better understanding of the effects of transportation improvements and allowing decision makers in state governments, local governments, transit operators, metropolitan planning organizations and environmental agencies the capability of making improved transportation decisions.

This program was funded through the Travel Model Improvement Program.

Further information about the Travel Model Improvement Program may be obtained by writing to:

TMIP Information
Metropolitan Planning Branch (HEPM-30)
Federal Highway Administration
U.S. Department of Transportation
400 Seventh Street, SW
Washington, D.C. 20590

# Table of Contents


**1.0 Introduction** ..... 1
- 1.1 Background ..... 1
- 1.2 The Importance of Vehicle Availability Modeling ..... 2
- 1.3 The Range of Vehicle Availability Model Types ..... 4

**2.0 Basic Practice** ..... 7
- 2.1 An Aggregate Model – Milwaukee ..... 7
- 2.2 Basic Practice Disaggregate Models ..... 9

**3.0 Advanced Practice** ..... 15
- 3.1 Advanced Models with Transit and/or Highway Accessibility Variables ..... 16
- 3.2 Advanced Models with Pedestrian Environment Variables ..... 24

**4.0 Innovative Approaches** ..... 31
- 4.1 Vehicle Type Choice Models ..... 31
- 4.2 Dynamic Models of Vehicle Availability ..... 33

**5.0 Report Summary** ..... 37



*PROTECTED UNDER INTERNATIONAL COPYRIGHT*
*ALL RIGHTS RESERVED*
NATIONAL TECHNICAL INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE

Reproduced from best available copy.

# List of Tables


2.1 Specification of the New Hampshire Vehicle Availability Model ..... 13

3.1 Specification of the 1976 MTC Worker Household Auto Ownership Model ..... 18

3.2 Specification of the 1989 Portland Metro Auto Ownership Model ..... 20

3.3 Specification of the Seattle Auto Ownership Model ..... 23

3.4 Specification of the 1992 Portland Vehicle Availability Model ..... 25

3.5 Specification of the Philadelphia Vehicle Availability Model ..... 29

4.1 Specification of Train's Vehicle Quantity Submodel ..... 34


# List of Figures


3.1 Alternative Vehicle Availability Model Structures – Philadelphia Model ..... 27

4.1 Auto Ownership and Use Model – Structure of Train's ..... 32

4.2 Structure of Dynamic Personal Vehicle Submodel ..... 36

# 1.0 Introduction

## 1.1 Background

The number of motor vehicles available to a household has a major impact on the travel behavior of the members of the household. As a result, many metropolitan planning organizations have incorporated models of household vehicle availability or automobile ownership into their travel forecasting model systems. This report summarizes the current state of the practice in modeling vehicle availability. In the past, these models were most frequently labeled auto ownership models, but more recently the term vehicle availability has been used for two reasons. First of all, household vehicle availability, a measure of the total number of motor vehicles available for use by household members (including both passenger cars and trucks owned, leased, and/or provided by employers), is likely to be more closely related with the level of household mobility than the more limited household auto ownership measure. Secondly, data on vehicle availability is collected in the decennial U.S. Census rather than auto ownership.

The term auto ownership has sometimes been used by travel modelers and forecasters in its strict sense to include only a consideration of the number of automobiles actually owned by household members, and sometimes in a more generic sense to include additional motorized vehicles, such as pickup trucks and motorcycles, and to include leased vehicles and vehicles available to household members but owned by others, most often employers. The term vehicle availability is defined to include each of these types of vehicles explicitly. Since these two terms have not always been used precisely, it is sometimes difficult to determine the exact definition used in the model for a particular metropolitan area. For this reason, in our summaries of existing models, we will use the same terminology which has been used in model documentation, and in our general discussion, we will use the currently accepted precise terminology, vehicle availability.

This report provides examples of a number of types of vehicle availability models and specifies their data sources, explanatory variables, and details of their implementation. Each example model is also evaluated, both relative to alternative models and with respect to the types of applications by MPOs and statewide planners for which the model is best suited.

This section of the report provides an introduction to vehicle availability modeling, dealing with its importance and the range of model types. Section 2.0 focuses on examples of basic practice by U.S. transportation planners, including early aggregate and cross-classification models and a recent vehicle availability choice model based completely on Census data. Section 3.0 deals with examples of more advanced practice, including models which are based on household survey data and/or include additional variables such as transit and highway accessibility and pedestrian environment variables. Section 4.0 discusses innovative approaches, typically not yet in production use in regional models, including models which combine vehicle availability and vehicle type choice, and

dynamic models which focus on vehicle acquisition and scrappage rather than solely on vehicle availability. These innovative approaches provide examples of the state of the art rather than state of the practice in vehicle availability modeling and thus help to show how the state of the practice may change in the future. Finally, Section 5.0 provides a concluding summary of the entire report.

## 1.2 The Importance of Vehicle Availability Modeling

Vehicle availability is a particularly critical variable in both trip generation and mode choice models. This household characteristic can also have indirect effects on trip distribution and on household location choices. Because vehicle availability is a factor in each of the major steps of travel forecasting except highway and transit assignment, it should be modeled explicitly as part of the total travel forecasting process. Each of the effects of vehicle availability on travel and location patterns is discussed in this section.

**Trip Generation -** In many urban areas it has been found that, when controlling for the number of persons, households with more vehicles generate more person trips. This is not surprising since the increased availability of a vehicle leads to more tripmaking opportunities. In addition, since auto travel is generally faster than other modes, auto owners may have more time in which to perform activities which require travel. While there is some indication that the greater number of trips for households with higher vehicle availability in some areas may be due to underreporting of walk trips, which are higher in households with fewer vehicles, there still seems to be a correlation between vehicle availability and trip generation. One reason may be that vehicle availability may be a proxy variable for income, and households with higher income levels tend to make more trips.

**Mode Choice -** Vehicle availability variables have proven to be highly significant indicators in mode choice models; there is clearly a relationship between vehicle availability and mode choice. Often, the strongest indicator of whether auto or transit will be used is not relative travel times or costs, but the vehicle availability level of the traveler and/or the traveler's household. This is particularly true for households which have no vehicles available. It is generally accurate to assume that no trips made by these households will be auto driver trips. Since auto drivers are often the most frequently chosen travel mode for the region as a whole, the travel behavior of no-vehicle households differs very significantly from that of the typical household in the region.

Vehicle availability measures such as vehicles per household, person, licensed driver, or worker have all been used in mode choice models to reflect not only the differences in mode choice for no-vehicle households, but also the differences as vehicle availability increases. In addition, nonlinearities are often captured by using dummy or indicator variables based on comparisons of the numbers of vehicles available and the number of persons, drivers, or workers in the household. These variables help to differentiate between households with some level of 'competition' for the use of the limited number of vehicles and other households in which each licensed driver has continuous access to one or more vehicles. Household vehicle availability levels are also sometimes used as a basis

for mode choice model stratification. When this strategy is used, separate models may be developed for different vehicle availability levels, each potentially with its own set of available modes, explanatory variables, and estimated coefficients. Alternatively, a single mode choice model may be applied separately to groups of trips, between given origins and destinations, classified by the vehicle availability level of the tripmakers' households. This approach will provide more accurate predictions than those obtained using average vehicle availability values for total origin/destination travel.

**Trip Distribution** – When gravity models, the most common trip distribution procedures, are used with either highway travel times or composite highway and transit impedances, the traveler's choice of destination is not related to vehicle availability. This ignores, however, the possibility that travelers with vehicles available are more likely to use these vehicles to travel to locations which cannot be reached by transit, walking, or bicycling. This possibility can be included in trip distribution models in a number of ways:

- In the gravity model, using an impedance variable, such as a logsum, derived from a mode choice model which includes vehicle availability variables, so that all variables affecting mode choice will also affect trip destination choice;

- Using a logit destination choice model with logsums from a logit mode choice model as a composite impedance variable, plus additional variables measuring the attractiveness of alternative destinations;

- Using a combined logit model of destination and mode choice; or

- Using separate models of destination choice for households having different levels of vehicle availability.

If any of these strategies are used, then vehicle availability will have an indirect or secondary effect on destination choice.

**Household Location Choices** – Land use allocation models such as DRAM/EMPAL typically locate new households and relocate existing households using a gravity-type procedure to determine residence locations rather than trip attraction locations. As in the case of the simplest type of gravity model for destination choice, these procedures are usually not affected by households' vehicle availability levels. However, if any of the strategies listed above for destination choice models are also used as part of the household allocation model, then vehicle availability will also have an indirect or secondary effect on household location choices. Models formulated in this way will capture the increased likelihood, for example, that households with fewer vehicles available will (re)locate in areas in which the differences between the levels of transit and highway service are more positive (or less negative) than the average difference is.

## 1.3 The Range of Vehicle Availability Model Types

Over the past 30 years, a wide variety of different vehicle availability models have been developed, and many of these types are still in use in one or more metropolitan area model systems. Additional modeling strategies have recently been tested in research environments and show promise of providing more advanced models for practical application in the future. This section identifies the major differences in these models and provides an overview of the classification used in the remainder of this report to group a number of examples of differing vehicle availability models.

### 1.3.1 State of the Practice Versus State of the Art

The first and most general model characteristic is whether a specific vehicle availability model represents the state of the travel forecasting practice or the state of the vehicle availability modeling art. State of the practice models are defined here as all models which have been - or soon will be - incorporated into the 'standard' regional travel forecasting system of a U.S. metropolitan area, or into a statewide travel model. All of the models presented in Sections 2.0 and 3.0 are state of the practice models. The models in Section 4.0 are state of the art models which include extended features such as vehicle type choice and the modeling of vehicle availability dynamics. These models represent research explorations for improved future regional models or more detailed modeling approaches oriented to other forecasting requirements, such as the prediction of future vehicle type mixes in response to clean air and energy conservation measures.

### 1.3.2 Static Versus Dynamic Models

A second general model characteristic is whether a specific vehicle availability model is designed to predict static or dynamic behavior. Static models predict the vehicle availability at a single point in time: the average or expected number of vehicles available to a household at the given time, or the probability that a household will have available a specified number of vehicles at the given time. All models currently in use in regional travel forecasting systems are static models.

Dynamic models predict the change in vehicle availability levels between two points in time: whether a single household or a group of similar households will reduce the number of vehicles available, continue to have available the same number of vehicles, or acquire an additional vehicle. Dynamic models are sometimes labeled *transaction* models; they predict vehicle sales, scrappages, trades, and purchases and provide estimates of the levels of vehicle availability in the forecast year in terms of changes from a base year. Although dynamic models show promise for future application in activity-based travel forecasting systems, none of these models is currently used in an MPO's regional forecasting system.

### 1.3.3 Alternative Data Sources

Vehicle availability models have been developed using a number of different data sources. In most cases, the same local travel survey data used to develop models which predict travel patterns have also been used to develop vehicle availability models. In a number of cases, however, Census data, either the Census Transportation Planning Package (CTPP) or the Public Use Microdata Sample (PUMS), has been used; particularly in regions in which no travel survey data exist or the most recent data has become out of date. Both Census data sources have limitations when compared to local travel survey data. The CTPP data are only reported as aggregations for geographical areas of various sizes, while the PUMS data provides individual household records without detailed locational information. Thus, while the PUMS data have been used in a number of vehicle availability modeling efforts, none of the models based solely on this data source include variables specific to the traffic analysis zones in which the households are located.

The third data source for vehicle availability models is travel survey data collected for a sequence of years – panel surveys in which data are collected from the same set of households at two or more points in time. This type of data is required to develop the transaction models discussed above. Only one panel survey has been completed by a U.S. metropolitan area, in Seattle. However, this data source has not yet been used to estimate any vehicle availability models. Panel survey data have been used, however, in one of the research-oriented vehicle availability modeling efforts discussed in Section 4.2.

A valuable data source which can be used to validate vehicle availability model results is state motor vehicle department files. In most states, vehicle registrations can be obtained by vehicle class, geography (e.g., county), year of registration, and vehicle age. Model results should match these data well, but caution must be exercised as many vehicles are registered at different addresses than homes (e.g., company cars).

### 1.3.4 Data Aggregation Level

Vehicle availability models can either be developed using observations of disaggregate data at the individual household level, or aggregate data at the zonal or district level. Models based on Census data such as the CTPP must be developed at the aggregate level, and aggregate data based on travel surveys in the past provided the basis for all travel model development. The use of disaggregate household-level data is now most common; either from travel surveys or from the Census PUMS data. Models based on disaggregate data are preferred because they retain much more explanatory power than can be determined from zonally aggregated averages. The focus of the models summarized in this report is on models based on disaggregate data, but an example of a model based on aggregate data is presented in Section 2.1.

### 1.3.5 Alternative Model Structures

Over the past 40 years the most predominant mathematical model structures have changed from regression models to cross-classification models to logit and probit choice

models. This sequence is reflected in the order of model presentation in Sections 2.0 through 4.0. The specifications of each of these mathematical structures are also discussed as the models are presented. In general, regression models have been developed with aggregate zonal data and cross-classification models with disaggregate household data summed for various combinations of classification variables. These types of models are presented in Sections 2.1 (Milwaukee) and 2.2 (Detroit).

The Seattle model presented in Section 3.1 combines two model forms - cross-classification and an aggregate logistic regression structure. The cross-classification portion is based on disaggregate PUMS data and the regression portion on aggregate Census data at the zonal level.

Both logit and probit choice models have been developed using disaggregate data. Two types of logit models have been developed. Standard multinomial logit (MNL) models assume that each household considers each available vehicle availability level simultaneously and chooses its own level based on maximizing its expected utility. Models having the MNL structure have become the most common in recent years; six examples are provided in Sections 2.2 through 4.1.

Ordered response logit (ORL) models are based on the assumption that households make separate decisions concerning whether to have available each higher level of vehicles. Thus, households first decide whether they will not have any vehicles or will have at least one vehicle. Next, if they choose to have at least one vehicle, they decide whether to have one vehicle or more than one. This process continues until the decision has been made to have a specific level of vehicle availability. The DVRPC model discussed in Section 3.2 is an example of the ORL model structure. Probit models provide an alternative means of modeling ordered response behavior, or of deciding how the household will change its vehicle availability during a specified time period.

### 1.3.6 Alternative Types of Explanatory Variables

All vehicle availability models use variables describing household characteristics, such as number of persons, number of workers, and annual household income. Many also include demographic and land use characteristics of the zone in which the household is located, such as employment density and area type (CBD, urban, suburban or rural, for example). In recent years, models have been developed in which the zonal variables have been extended to include measures of the quality of the pedestrian environment and the accessibility of employment and/or retail facilities from the household's zone when traveling either by transit or by auto. In general, this progression in the numbers of types of variables used is reflected in the order of presentation in Sections 2.0 through 4.0. Section 3.0 specifically deals with models having pedestrian environment and accessibility variables.

# 2.0 Basic Practice

The models presented as examples of the basic methods of forecasting vehicle availability are either developed using aggregate (zonal-level) data or include only household and zonal demographic and land use data. These models represent a range of sophistication with respect to their structures - from linear-in-parameters regression to cross-classification and multinomial logit. A common characteristic, however, is that each was estimated solely using either travel survey data or Census data. These models therefore reflect minimal data collection requirements for vehicle availability models, and minimal data processing prior to model estimation. These characteristics make the basic practice models advantageous for MPOs with very limited data collection and model estimation budgets. These low costs, however, are offset by significant limitations in the explanatory variables which can be included in the models - differences from zone to zone in accessibility by walk, transit, and auto modes cannot be included in the basic practice models. There are additional advantages and disadvantages which are specific to each of the models; these are discussed in the following subsections.

## 2.1 An Aggregate Model - Milwaukee

### 2.1.1 Description

The Southeastern Wisconsin Regional Planning Commission uses a linear-in-parameters regression model to estimate zonal average automobile availability by zone[1]. The model was developed using a combination of Census data and MPO land use data at the zonal level. The explanatory variables used in the model are average household income (in thousands of dollars), average number of persons in households, and household density, defined as the number of households per developed gross residential acre. Two equations were estimated; one for the central portions of the most urbanized county in the study area, and the other for the remainder of the region. These two equations are:

<u>Central Milwaukee County</u>

Average number of autos available per household =
0.0466 * (Average household income in $1,000) - 0.0622

[1]Southeastern Wisconsin Regional Planning Commission, *Travel Simulation Models for the Milwaukee East-West Corridor Transit Study*, May 1993.

Remainder of Region

Average number of autos available per household =
0.445 * ln(Average household income in $1,000) - 0.163 * ln(Household density) + 0.200 * (Persons per household) - 0.144

These equations could be obtained using standard linear regression procedures in statistical packages after computing each of the variables and, in the case of the income and density variables in the second equation, converting to natural logarithms. Application of the model to obtain future forecasts requires the use of analysis year estimates of the independent variables. Household sizes and densities can be computed from the MPO's zonal forecasts of population, households, and residential land. In Milwaukee's case, forecasts of average household incomes by zone must be made specifically for use in the auto availability model.

In Milwaukee, additional processing of the forecasted average auto availabilities by zone is required to convert to the numbers of households with zero, one, or two or more autos available, as required for input to the region's trip production models. These cross-classification models require estimates of the number of households in each of the twelve cells defined by auto availability level (0, 1 or 2) for four household size categories (1, 2, 3 or 4, or 5 or more persons). Base year travel survey data were used to estimate relationships between the following probabilities and the average household sizes and number of autos available:

- The probability of a specified auto availability level in each zone;
- The probability of a specified household size in each zone; and
- The probability of a specified household size, given a specified auto availability level in the entire region.

These probabilities are then used to subdivide the total households in a zone into the required groups of households.

### 2.1.2 Evaluation

The Milwaukee model has the advantage of using a limited number of data sources, all at the zonal level of aggregation. These sources include the region's most recent travel survey (1991) and the corresponding 1990 Census data. The model is relatively simple to estimate using standard statistical packages. It includes as explanatory variables household and zonal characteristics which intuitively are correctly related to the likelihood of household auto availability. The model can be applied easily in a spreadsheet, database management program, simply programmed stand-alone procedure, or directly in the zonal data manipulation portions of some travel forecasting packages.

The disadvantages of the model are that it does not explicitly include any measures of zonal accessibility or the pedestrian environment. These factors, among others, account for the need for separate equations for the central portion of the study area and all other

areas, but discontinuities are likely to exist in the forecasts for zones on the two sides of the line dividing the areas in which each equation is to be applied. Because aggregate zonal data were used for model estimation, the model parameters are likely to be appropriate for forecasting at the zonal level, but biased toward zero compared to the expected parameters of a corresponding household-based model. As a result, the model's elasticities will underestimate household-level sensitivities of auto availability to the model's explanatory variables. Finally, the statistically estimated model does not provide the variables required for input to the trip generation process. Additional procedures, based on observed base year zonal and regional probability distributions of auto availability and household size shares, must be used to convert the results of the estimated models to the form required for subsequent forecasting steps.

The simplicity of the Milwaukee's core auto availability model lends itself for use in areas with very limited funds available for data collection (since Census data can be used) and model development. However, the use of PUMS data, as discussed in Section 2.2 - or alternatively travel survey data if they are available - to develop cross-classification or logit choice models may be a preferable model development strategy. This alternative would involve basically the same level of data preparation effort with only slightly more complex statistical estimation procedures. The resulting model would reflect the desired household-level sensitivities and could also provide the detailed breakdown of households by household size and auto availability levels required in the trip generation models.

The variables included in the Milwaukee model are limited by their exclusion of walk, highway and transit accessibility measures, but these variables may not be important in smaller regions which have relatively homogeneous urban design characteristics, minimal levels of transit service, and low highway congestion levels. Only as these characteristics are violated will the inclusion of the additional variables be important.

## 2.2 Basic Practice Disaggregate Models

### 2.2.1 A Cross-Classification Model - Detroit

#### *Description*

In a model which is currently being updated, the Southeast Michigan Council of Governments (SEMCOG) estimated vehicle availability using sets of empirical curves for the fraction of households owning zero, one, two and three or more vehicles as a function of household income level (11 categories).[2] The curves are stratified by household size (one, two, and three or more persons) and by residence zone area type (City of Detroit or other). The curves were derived from tabulations provided in the published summaries

[2]Southeast Michigan Council of Governments, *Southeast Michigan Travel Forecasting Process*, Detroit, October 1984.

of the U.S. Census's 1977 Annual Housing Survey (AHS) data for the Detroit metropolitan region. Although these tabulations represent aggregations of the basic household data collected in the housing census, the resulting fractions of households by vehicle availability level are essentially average fractions calculated from summations of the individual household responses in each of the 66 cells defined above.

Vehicle ownership in the SEMCOG model tends to increase with household income, with household size, and with suburban versus urban location. The highest zero-vehicle fraction occurs for one-person households in the lowest income group living in the City of Detroit. Even in suburban zones, however, the four lowest income groups have fractions of households with zero vehicle ownership of more than six percent.

SEMCOG uses its vehicle availability curves to predict the number of households per zone (given zonal household totals) at each vehicle availability level in each of 25 household categories defined by household size (one, two, three, four, and five or more) and income range (income quintiles). The number of households in each category is estimated using conditional probabilities based on 1981 Census data and marginal household totals provided by SEMCOG's small area land use and demographic forecasting process. The results are recombined into 20 categories, based on household size and vehicle availability (i.e., the income dimension is collapsed). These 20 categories are then available as input to trip generation (person trips by purpose) using home-based trip production rates per household cross-classified by household size and vehicle availability. These steps are performed by stand-alone programs which incorporate both vehicle availability and trip generation forecasting.

### *Evaluation*

The SEMCOG model illustrates both the advantages and disadvantages of basic practice. On the positive side, the model is straightforward and requires only a single data source which is readily available for all metropolitan areas; if not from the AHS itself, then from the decennial CTPP or PUMS Census data. The model captures the long-recognized primary relationship between household income and household auto ownership. On the negative side, the model ignores gender of the household head, workforce participation, and age distribution, which have contributed significantly to the recent growth in vehicle ownership. Also, the model relies on the city-suburb distinction as a crude proxy for the combined effects of factors as diverse as socioeconomic status, land use patterns, the quality of transit service and the accessibility it provides to jobs and shopping, and the quality of the pedestrian environment. Considering how each of these aspects of urban social and economic structure are changing, it is likely that such a model would be prone to drift farther from reality as the base year for analysis or the forecast year become more removed from the calibration year. Among other things, this implies that frequent recalibration of the model on up-to-date data would be highly desirable.

As in the case of the Milwaukee model, the Detroit model lends itself for use in areas with very limited funds available for data collection (since Census data can be used) and model development. It provides a method of reflecting the current patterns of vehicle availability as they vary by household income, household size, and area type, and provides directly the shares of households by vehicle availability level frequently required for trip generation. Finally, as for the Milwaukee model, if the missing accessibility variables do

not vary significantly in the study area, then the lack of these variables in the model will not have a significant effect on the accuracy of the vehicle availability forecasting process.

### 2.2.2 A Logit Model Based on PUMS Data - New Hampshire

*Description*

In a paper published in 1994, Purvis explored the usefulness of the PUMS data set as a basis for estimating logit choice models of automobile ownership.[3] He demonstrated the consistency of logit choice models based on PUMS and household survey data, concluding that the PUMS data are useful for metropolitan areas and states which do not have access to recent household travel survey data. He also identifies the major weakness of this approach - an inability to include zonal variables or accessibility measures in the models because the individual households provided in the PUMS data are not identified by their location except at the level of districts including at least 100,000 persons. Thus, the PUMS data source is characterized as a 'second best' data set for automobile ownership model development which cannot completely substitute for data from a comprehensive household travel survey.

The New Hampshire statewide planning study vehicle availability model is an example of the use of the PUMS data set for this purpose.[4] The statewide modeling process included a 2,800-household travel survey conducted in 1995 to obtain data from all regions of the state. However, comparisons of the responses with CTPP data indicated that households with no workers, low incomes, and/or no vehicles available are significantly underrepresented in the survey. Particularly because there are too few households with no vehicles available, it was not possible to estimate satisfactory logit choice models of vehicle availability using the travel survey data. Thus, it was necessary to resort to PUMS data as a 'second best' basis for vehicle availability modeling. However, since the coefficients related to statewide zones, such as population density and employment density, in the models based on the household data were not significantly different from zero, it appears that these measures have little effect on the explanatory power of the models. This suggests that for New Hampshire, PUMS-based models without these zonal variables may be just as accurate as the household survey-based models could have been.

The ALOGIT program was used to estimate a logit choice model using the 20,897 households in the New Hampshire PUMS-A five percent sample. The standard logit mathematical form used is the following:

[3] C.L. Purvis, "Using 1990 Census Public Use Microdata Sample to Estimate Demographic and Automobile Ownership Models," *Transportation Research Record 1443*, 1994.

[4] N. Jonnalagadda and K. Tierney, *New Hampshire Statewide Planning Study Vehicle Availability Model*, Cambridge Systematics, Inc., June 24, 1996.

$$\text{Prob}(n) = \frac{e^{U_n}}{\sum_{i=0}^{n_{max}} e^{U_i}}$$

where:

| | | |
|---|---|---|
| Prob($n$) | = | the probability that a given household will own $n$ vehicles ($n$ = 0, 1, ... $n_{max}$) |
| $e$ | = | the base of Naperian logarithms |
| $n_{max}$ | = | the largest vehicle availability category |
| $U_n$, $U_i$ | = | the utility of owning $n$ or $i$ household-specific vehicles |

The utilities, $u_n$, are defined as:

$$U_n = b_{n0} + \sum_{j=1}^{nv} b_{nj} X_{nj}$$

where:

| | | |
|---|---|---|
| $b_{n0}$ | = | a statistically estimated constant associated with having $n$ vehicles |
| $b_{nj}$ | = | a statistically estimated coefficient indicating the relative importance of variable $X_{nj}$ on the utility of vehicle availability level $n$ |
| $X_{nj}$ | = | a variable specific to the given household or the household's zone of residence |

The New Hampshire model estimates the probabilities of a household having zero, one, two, three, or four or more vehicles. The explanatory variables and estimated coefficients are provided in Table 2.1. The strongest variable is the natural logarithm of household income, but each of the household variables is highly significant. It was possible to include one locational variable, an urban area indicator, because one of the PUMA districts in New Hampshire is completely urbanized.

## *Evaluation*

The New Hampshire model demonstrates the usefulness of the PUMS data set as a supplement to, or substitute for, travel survey data. Disaggregate choice models can be developed using this data source as long as no zonal-based variables are included in the model. The resulting model provides a robust means of estimating vehicle availability at the household or zonal level. In the New Hampshire case, preliminary models based on the insufficient travel survey data suggested that zonal variables would not increase the final model's explanatory power; this indication is consistent with the relatively low population and employment densities and the minimal importance of transit as a travel mode throughout New Hampshire. Where conditions such as these exist, logit models based on the PUMS data represent a desirable vehicle availability modeling strategy due to the availability of the required data and the relative ease of model development. These models have an advantage over cross-classification models based on the same data source in that no artificial segmentation of continuous variables such as household income into categories is required.

# Table 2.1 Specification of the New Hampshire Vehicle Availability Model

| | Vehicle Availability Level | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Zero | | One | | Two | | Three | | Four or More | |
| Variable | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic | Coefficient | t-statistic |
| Alternative-specific constant | 0 | --- | -5.638 | -14.2 | -16.34 | -34.0 | -22.52 | -36.5 | -28.95 | -33.3 |
| Persons per household | 0 | --- | 0.2190 | 4.9 | 0.7280 | 16.0 | 0.8740 | 18.0 | 1.052 | 18.9 |
| Workers per household | 0 | --- | 0.4792 | 7.6 | 1.006 | 15.4 | 1.486 | 21.1 | 2.086 | 26.6 |
| Ln(Household income) | 0 | --- | 0.6585 | 15.2 | 1.564 | 30.8 | 1.912 | 30.6 | 2.243 | 26.6 |
| Single family dwelling dummy | 0 | --- | 0.9992 | 13.8 | 0.9992 | 13.8 | 0.9992 | 13.8 | 0.9992 | 13.8 |
| Urban area dummy | 0 | --- | -0.4569 | -4.7 | -0.9794 | -8.9 | -1.420 | -10.3 | -1.688 | -8.9 |

# 3.0 Advanced Practice

Each of the vehicle availability models selected as an example of current advanced practice by MPOs and state DOTs is based not only on household characteristics and zonal variables such as densities and area type, but also on additional variables which are related to the transportation facilities and services available to the residents of a particular zone. In the advanced models currently in use, these additional variables include either mode-specific measures of accessibility, measures of the quality of the pedestrian environment, or both.

Accessibility measures usually combine information on the distribution of possible trip attractions around a zone and information on the travel times or generalized costs required to reach the full set of available destinations for travelers on a specified mode. The most typical accessibility measure used in vehicle availability modeling is the fraction of total regional employment which can be reached within a specified transit travel time - 30 minutes, for example. In some cases, a similar measure based on highway travel times is also used. An alternative accessibility measure is the summation of the exponentiated logit destination choice model utilities for a specified travel mode and all destinations. Both of these measures increase as more potential trip attractions are located nearer to the residence zone and as levels of service by the specified travel mode are increased. Furthermore, if the accessibility measures for transit increase, the need for automobiles, and thus household vehicle availability, will tend to decrease. Conversely, if accessibility measures for highway trips increase, households may be more likely to obtain additional vehicles to reach the available attractions more conveniently.

Accessibility measures are of necessity complex variables based on a combination of attraction data (usually employment) in all zones of the study area and on mode-specific levels of service between the residence zone and all other zones. Computing these measures requires obtaining skimmed level-of-service matrices under specified conditions (free-flow versus loaded highway networks, or midday versus peak-period transit services, for example), combining these matrices with zonal attraction data, and summing over all attraction zones for each residence zone.

Measures of the quality of the pedestrian environment also require significant effort to obtain. The measures used to date cover a wide range. Some are subjective ratings of pedestrian travel amenities as reflected in building setback distances, sidewalk availability, and the ease of crossing streets. Others are based on an analysis of street network connectivity performed using a geographic information system. All are designed to provide a measure of the degree to which trip attractions can be reached by walking rather than driving. Model estimation results show, as expected, that vehicle availability levels tend to be lower in zones which have higher levels of pedestrian travel amenities.

The subsections which follow describe three models which include accessibility variables as well as household and zonal variables (Section 3.1), and two models which include both accessibility and pedestrian environment variables (Section 3.2). Within each of these subsections, the models are discussed in the order in which they were developed.

# 3.1 Advanced Models with Transit and/or Highway Accessibility Variables

## 3.1.1 The 1976 MTC Logit Models

### *Description*

The earliest logit auto ownership models incorporated into a regional model system were the Metropolitan Transportation Commission's models developed for the San Francisco Bay area in 1976.[5] Two models were developed using a combination of travel survey data, highway and transit network level-of-service data, and zonal land use, population, and employment data. The two models included one to predict shares of households without workers owning zero, one, and two or more autos, and one to predict the same shares for households with workers. Details of the worker-household model are discussed in this section.

The MTC worker-household auto ownership model is applied *after* the work trip attraction zone of the household's primary worker has been predicted and before the mode choice of this trip is predicted. This shift from the normal sequence of estimating auto ownership before trip generation and destination choice allows the model to reflect the characteristics of each available mode for the chosen destination of the trip to work. Thus, if transit is not available for this work trip, the household will be more likely to choose to own more autos; conversely, if good transit service is available and parking costs are high, fewer autos are likely to be chosen.

The independent variables in the model are the following:

- Household size;
- Remaining household income after housing, auto ownership, and commuting expenses for the work trip have been taken into account;
- Whether or not the household lives in a single family detached housing unit;
- Employment density in the residence zone;
- The relative accessibility by transit versus drive alone and shared ride modes for the trip to work; and
- The relative accessibility by transit versus auto for all shopping trips made by the household.

[5]Cambridge Systematics, Inc., *Travel Model Development Project: Phase 2 Final Report, Volume 2: Detailed Model Descriptions,* Cambridge, Massachusetts, June 1980.

The relative accessibility for the chosen work trip is defined more explicitly as a ratio of the exponentiated utility of the transit mode to the sum of the exponentiated utilities of the drive alone and shared ride modes for the chosen work trip between the residence zone and the employment zone. For the shopping accessibility variable, the definition is the ratio of the denominator of a transit-specific shopping destination choice model to the denominator of an auto-specific shopping destination choice model. In both cases, these destination choice denominators are conditional on the auto ownership level of the alternative.

Table 3.1 presents the estimation results for the logit worker household auto ownership model. The variables based on household characteristics - income, household size, and the single-family dummy variable for the two or more auto alternative - are the most significant variables. Although the zonal and accessibility variables are generally less significant, their signs agree with intuition and provide the desired sensitivity of auto ownership to the relative service levels by mode for the chosen work trip and for all potential shopping trip destinations, and also to the distribution of retail and service employment throughout the metropolitan area. Through the accessibility variables, auto ownership is also indirectly related to additional variables such as the proximity of the residence zone to the Central Business District and workers per household.

Due to the complexity of the accessibility variables and the need for data from multiple sources, the data preparation process for the MTC model was extensive. The final estimation data file consisted of 584 households. Once the estimation data file was assembled, predecessor programs to the ALOGIT estimation package could be run efficiently to provide statistical estimates of the model's parameters. Forecasting with the model can be done at the zonal level with households grouped by their number of workers and by the mode of travel to work of the head of household. All of the required variables are provided by prior model outputs, by subsequent model utilities, or by the region's land use and demographic forecasting process.

### *Evaluation*

The MTC model has a number of advantages: it is estimated using disaggregated household data, it predicts probabilities of owning various auto ownership levels, and it includes a wide range of variables which affect auto ownership, including accessibility measures which combine the locations of potential destinations and the auto and transit travel times required to reach these destinations. Furthermore, auto ownership is predicated on the chosen work place of the primary worker. All of these characteristics provide a rich model which is sensitive to a wide range of policy options.

The model does have, however, a number of offsetting costs or limitations. The highest auto ownership level included is two or more vehicles; more modern models instead include two, three, and four or more vehicle alternatives. Compared with models based

# Table 3.1 Specification of the 1976 MTC Worker Household Auto Ownership Model

For zero-auto households:

$$U(0) = 0.7919 * \ln(RINC0) + 0.06814 * RMC0 + 0.5608 * RSHD0$$

For one-auto households:

$$U(1) = 4.989 + 0.7919 * \ln(RINC1) + 0.06814 * RMC1 + 0.5608 * RSHD1 - 0.05419 * RSDENS - 2.689 / PHH + 0.3935 * SFD$$

For two-or-more-auto households:

$$U(2+) = 5.689 + 0.7919 * \ln(RINC2) + 0.06814 * RMC2 + 0.5608 * RSHD2 - 0.05419 * RSDENS - 6.013 / PHH + 1.342 * SFD$$

where:

| | | |
|---|---|---|
| U(n) | = | Utility of auto ownership level n |
| RINCn | = | Remaining income of a household owning n autos, after reductions of household income to account for expected work trip costs, per-person annual costs, and average annual costs of owning n autos (1965 dollars) |
| RMCn | = | Ratio of the exponentiated utilities of the primary worker mode choice model (transit over auto) for a household owning n autos |
| RSHDn | = | Ratio of the denominators of the mode-specific shop trip destination choice model (transit over auto) for a household owning n autos |
| RSDENS | = | Employment density for retail and service employees in the household's residence zone (employees per acre) |
| PHH | = | Household size (persons per household) |
| SFD | = | Single family dwelling dummy variable - one if the household resides in a single family dwelling; zero otherwise |

simply on travel survey or Census data, preparation of an estimation data set for the MTC model is very difficult. Incorporation of the model in the travel forecasting process is also difficult: inputs are required from both a mode choice model and a destination choice model, the model must be preceded by a procedure which divides total households into those with and without workers, and a number of complex variables must be computed.

Although the MTC model was a major pioneering step when it was developed, subsequent efforts have shown that somewhat more simple and 'transparent' vehicle availability models will provide nearly all of its advantages while avoiding many of its disadvantages. In addition, these models can provide additional advantages such as the inclusion of pedestrian environment variables. The MTC model continues to be useful, however, as a guide to strategies for incorporating additional factors into future vehicle availability models if the more current models are found to lack effects such as differences in vehicle ownership due to variations in work locations, to commuting and parking costs, and to the availability of shopping destinations which can be reached by transit.

### 3.1.2 The 1989 Portland Logit Model

#### *Description*

The Metropolitan Service District in Portland, Oregon (Metro) developed its initial logit choice model in 1989.[6] The model has four alternatives: zero, one, two, and three or more vehicles per household. The utility functions for each alternative are shown in Table 3.2. The model is estimated using household survey data supplemented with an accessibility variable which incorporates zonal employment and transit level-of-service data. The explanatory variables obtained from the survey data are household size, household income class (four categories), and workers per household. The accessibility variable is defined as the number of employment opportunities which can be reached within 30 minutes of transit time from the residence zone. After the considerable effort required to assemble the accessibility variables, estimation of the model was straightforward using the ALOGIT estimation package. The resulting model provides the expected positive relationships between each of the household variables and auto ownership levels. Also, auto ownership decreases as transit accessibility increases, mirroring the observed data for Portland, where the fraction of households owning zero cars decreases from 52 percent in the CBD, 10 percent in the remainder of the City of Portland, and less than four percent in the remainder of the study area.

Metro applies its vehicle ownership model at a market segment level in each traffic analysis zone. There are 64 segments per zone based on specific values of each of the household variables included in the model, with four levels for each variable. For application, the accessibility variable is defined using transit level-of-service data which varies as the highway speeds vary. Iteration from highway assignment and transit skimming back to

[6] Metropolitan Service District, *Travel Forecasting Methodology Report, Westside Light Rail Project*, Portland, Oregon, September 1989.

# Table 3.2 Specification of the 1989 Portland Metro Auto Ownership Model

For zero-auto households:

U = 5.125 - 0.918 * HHSIZE - 1.442 * WORKERCL - 1.580 * INCOMECL + 0.0000174 * TOTAL30T

For one-auto households:

U = 5.844 - 0.727 * HHSIZE - 1.076 * WORKERCL - 0.892 * INCOMECL + 0.0000084 * TOTAL30T

For two-auto households:

U = 2.871 - 0.167 * HHSIZE - 0.658 * WORKERCL - 0.215 * INCOMECL + 0.0000041 * TOTAL30T

For three-or-more-auto households:

U = 0

where:

| | |
|---|---|
| U | = Utility |
| HHSIZE | = Number of persons in household |
| WORKERCL | = Number of workers in household |
| INCOMECL | = 1 if household income < $15,000 |
| | 2 if household income > $15,000 and <$25,000 |
| | 3 if household income > $25,000 and <$35,000 |
| | 4 if household income > $35,000 |
| TOTAL30T | = Number of employees within 30 minutes of travel time via the transit mode |

the auto ownership model is required to ensure that consistent transit times and auto ownership levels are used throughout the forecasting process.

### *Evaluation*

The 1989 Portland model represents the state of the art of disaggregate models which include accessibility variables. Compared to the earlier MTC model, its simpler variables greatly facilitate model estimation and application. No inputs from other travel models are required. The transit levels of service are, however, required from subsequent model application steps. The model does reflect the impacts of transit availability and service levels on the need for, and the likelihood of, higher levels of vehicle availability. The model does not include the effects of differences in pedestrian amenities on vehicle availability, but these additional variables have subsequently been added to the more recent Portland model, as discussed in Section 3.2.1. Another disadvantage of the model is the lack of any nonaccessibility zonal variables, such as population and employment density measures.

The 1989 Portland model structure is recommended as a useful starting point for other areas who have travel survey data, wish to begin the development of disaggregate vehicle availability models for the first time, and do not wish to develop or include pedestrian environment variables. The model includes only straightforward household variables and the simplest type of accessibility variable. Then, once this basic logit model including accessibility has been developed, extensions can be explored to include zonal variables, highway as well as transit accessibility variables, pedestrian environment variables, and variables to inhibit the prediction of more vehicles per household than persons per household.

## 3.1.3 The Seattle Combined Cross-Classification/Regression Model

### *Description*

The Puget Sound Regional Council has recently implemented a vehicle availability model which includes accessibility variables and is based on a combination of PUMS data and CTPP data. The model is developed and applied in two stages. In the first stage, PUMS data are used to develop cross-tabulations of households by PUMA district, with the dimensions of household size, income group, number of workers, and vehicles available. These cross-tabulations are then applied to each zone within the PUMA, for which a cross-tab of households by the first three variables is available from the CTPP data, to provide an initial estimate of households by vehicle availability level. In effect, this stage involves using PUMA-average vehicle availability values for each cell of the zone-based tables to estimate households by vehicle availability level in each of these cells.

The second stage of model estimation and application introduces three accessibility variables. After an exhaustive set of tests, the final definitions of the selected variables defined for the zone of residence were the following:

- The number of employees within 10 minutes of walk time (counted at their workplaces);
- The number of employees within 30 minutes of peak transit time; and
- The number of employees within six miles, with distances measured on the highway network.

Table 3.3 provides the equations estimated using linear-in-parameters regression of zonal data to obtain the second stage of the Seattle model. The dependent variables in each equation represent the change in utility due to differences in accessibility levels from zone to zone within any particular PUMA area. Three accessibility measures are used, each is specific to one of the following modes of travel: walk, transit, and highway. Fractions of zero- and one-vehicle households increase as walk and transit accessibilities increase, while those of two- and three-or-more-vehicle households decrease, as expected. The results for the employment intensity variable are mixed, however, with decreases for the lowest and highest vehicle availability levels and increases for the middle levels.

An incremental logit procedure is used to apply the estimated accessibility-based changes in utilities to the first stage estimates of vehicle availability probabilities. The results are revised zone-specific probabilities of household's owning each of the alternative vehicle availability levels.

### *Evaluation*

The Seattle model represents vehicle availability model development at the aggregate level which makes maximum use of the information available in Census data for a region - households by type within PUMAs and households by zone. In addition, it captures accessibility effects at the zonal level. Thus, no survey data are required and maximum use is made of Census, zonal demographic, and transportation system information. As a result, the model is an example of one of the most advanced approaches which can be taken to vehicle availability modeling when no local survey data are available. The model includes a full range of household characteristic and mode-specific accessibility variables.

The model's primary limitation is that it is based on aggregations of household data rather than on information for individual households. As in all aggregate models, this introduces the possibility of aggregation bias which fails to estimate the true effect of changes in the independent variables on households' vehicle availability decisions. Thus, if survey data are available, they ideally would be used to estimate choice models having all of the variables included in the Seattle model. A second, and probably less important, limitation of the Seattle model is the lack of pedestrian environment variables. A walk accessibility variable is included, but the measure used is not sensitive to the quality of the pedestrian environment. The final very minor limitation of the model is that an approximation of the logit model mathematical structure is made to facilitate the estimation of a model using aggregate zonal shares. This approximation is equivalent to assuming that the net effect of the accessibility variables, relative to the first stage estimation results, is no change in the denominator of the logit share model.

# Table 3.3 Specification of the Seattle Auto Ownership Model

For zero-auto households:

$$U(0) = -0.8136 + 0.2474 * \log(\text{Walk Access}) + 0.1275 * \log(\text{Transit Access}) - 0.8136 * \log(\text{Employment Intensity})$$

For one-auto households:

$$U(1) = -0.6594 + 0.0599 * \log(\text{Walk Access}) + 0.0552 * \log(\text{Transit Access}) + 0.0716 * \log(\text{Employment Intensity})$$

For two-auto households:

$$U(2) = 0.2130 - 0.0868 * \log(\text{Walk Access}) - 0.0993 * \log(\text{Transit Access}) + 0.0648 * \log(\text{Employment Intensity})$$

For three-or-more-auto households:

$$U(3+) = 0.4920 - 0.0819 * \log(\text{Walk Access}) - 0.0637 * \log(\text{Transit Access}) - 0.0368 * \log(\text{Employment Intensity})$$

where:

| | |
|---|---|
| U(n) | = The utility of the change from PEn, the 'first cut' estimated percentage of households in a zone with n vehicles available, to POn, the observed percentage of households with n vehicles available, obtained from CTPP data |
| Walk Access | = Number of employees within 10 minutes of walk time from the residence zone |
| Transit Access | = Number of employees within 30 minutes of travel time by transit in the peak period from the residence zone |
| Employment Intensity | = Proportion of the total region's employees within six miles of the residence zone |

# 3.2 Advanced Models with Pedestrian Environment Variables

## 3.2.1 The 1994 Portland Logit Model

### *Description*

As part of the Making the Land Use/Transportation/Air Quality Connection project (LUTRAQ), the Metropolitan Service District in Portland, Oregon (Metro) and the project consultants cooperated to expand the logit choice model discussed in Section 3.1.2 to include revised income variables, a measure of retail activity near the household location, and pedestrian environment variables.[7] The revised model is shown in Table 3.4. The additional variables (beyond those in the model discussed in Section 3.1.2) are the following:

- Number of retail employees working within one mile of the household;
- A revised formulation of the income variable to provide more flexibility in vehicle availability modeling; and
- A pedestrian environment factor, discussed below.

The pedestrian environment factor (PEF) represents a composite measure of the pedestrian friendliness of each analysis zone. It was developed in acknowledgment of the fact that a number of factors at the neighborhood and street level affect individuals' willingness and ability to choose the walk mode for various trip purposes. The PEF is developed by assessing four different parameters for each zone:

- Ease of street crossings;
- Sidewalk continuity;
- Street connectivity; and
- Topography.

Due to time constraints, relatively qualitative approaches were used to assign values to each variable in each zone. The values were integers ranging from one to three, ranging from poor to good for each variable. The final PEF was defined as the unweighted sum of the four zonal values, ranging from 4 to 12.

[7] Metropolitan Service District, *The Phase III Travel Demand Forecasting Model: A Summary of Inputs, Algorithms, and Coefficients*, Portland, Oregon, June 1, 1994.

# Table 3.4 Specification of the 1992 Portland Vehicle Availability Model

For zero-auto households:

U(0) = -1.684 - 0.881 * HHSIZE - 1.452 * WRKRCL + 3.255 * INCOM1 + 1.942 * INCOM2 + 0.000220 * RET1M + 0.00001063 * TOTAL30T + 0.2095 * PEF

For one-auto households:

U(1) = 1.497 - 0.720 * HHSIZE - 1.065 * WRKRCL + 2.259 * INCOM1 + 1.944 * INCOM2 + 1.033 * INCOM3 + 0.000132 * RET1M + 0.00000615 * TOTAL30T + 0.0902 * PEF

For two-auto households:

U(2) = 1.619 - 0.141 * HHSIZE - 0.660 * WRKRCL + 0.377 * INCOM1 + 0.555 * INCOM2 + 0.0478 * INCOM3 + 0.000060 * RET1M + 0.00000334 * TOTAL30T + 0.0337 * PEF

For three-or-more-auto households:

U(3) = 0.0

where:

| | | |
|---|---|---|
| U(n) | = | Utility to a household of owning n autos |
| HHSIZE | = | Number of persons in the household |
| WRKRCL | = | Number of workers in the household |
| INCOMn | = | Dummy variable equal to one if the household income level is n |
| RET1M | = | Number of retail employees located within one mile |
| TOTAL30T | = | Number of employees within 30 minutes of travel time via the transit mode |
| PEF | = | Pedestrian environment factor |

As shown in Table 3.4, the model based on the new variables continues to indicate positive correlations between auto ownership and income, workers, and persons per household. In addition, auto ownership declines as retail intensity increases or the pedestrian environment improves. The revised model performs slightly better for the most pedestrian-friendly and least pedestrian-friendly areas, especially in predicting the number of zero-car households.

### *Evaluation*

Although particularly useful in estimating the impacts of urban design concepts such as neotraditional neighborhoods on vehicle availability and traffic levels, the expanded Portland model also demonstrates how nonmotorized travel behavior can be reflected in a vehicle availability model. This is done without assuming that pedestrian travel is essentially ubiquitous, or that the pedestrian environment has no impact on vehicular trip-making or vehicle availability. In Portland's case, a simplified Delphi approach was used to obtain agreed-upon values of the four components of the PEF with a minimum amount of effort. More detailed qualitative approaches can also be used to measure street widths and observe traffic and pedestrian signal timing, tabulate sidewalk continuity characteristics, evaluate block sizes, and measure land slopes. Some strategy should be used to measure pedestrian conditions and to incorporate these variables into the vehicle availability model in any region where walking is a viable mode of travel, and/or where the pedestrian environment varies significantly due to differences in street widths, sidewalk continuity, local street continuity, and hilliness.

## 3.2.2 The Philadelphia Ordered Response Logit Model

### *Description*

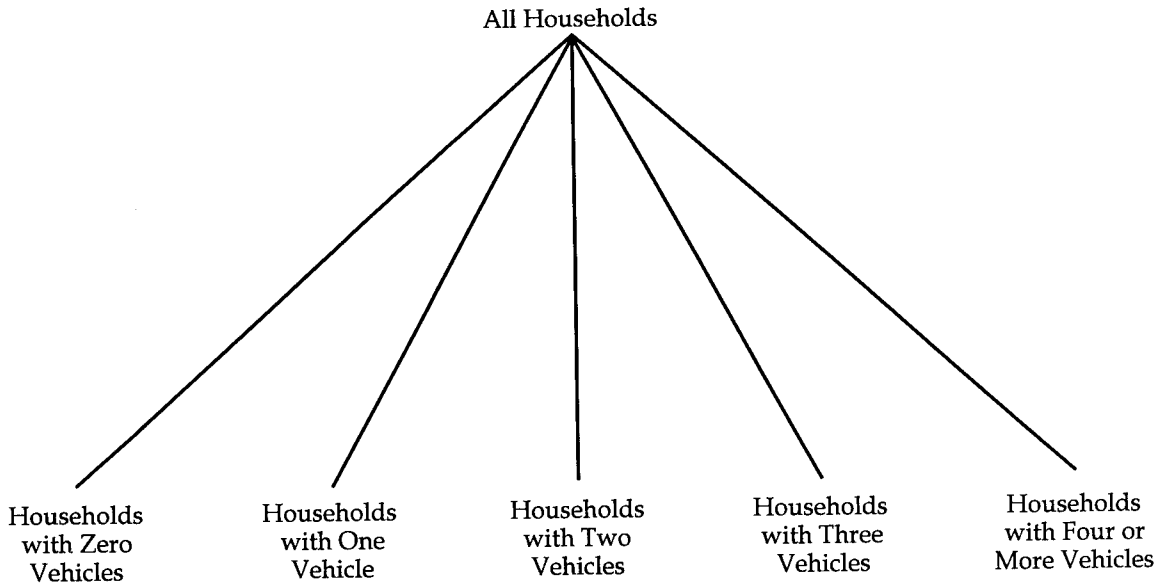
In a recent vehicle availability modeling effort performed for the Delaware Valley Regional Planning Council, alternative model structures were tested.[8] In addition, a combination of data sources were used: household survey data, zonal socioeconomic data, highway and transit accessibility variables based on combinations of zonal and transportation network data, and pedestrian environment factors developed at the zonal level. The resulting model represents an extension of the advanced features of the Seattle and 1994 Portland models discussed in the previous sections and can be presented by describing these extensions.

The first extension of the Portland model involved the estimation of two model structures: multinomial logit (MNL) and ordered response logit (ORL). Figure 3.1 shows these two structures. The MNL structure is consistent with the assumption that each household makes a one-time choice of the number of vehicles to have. The ORL structure, on the other hand, assumes that households arrive at their current vehicle availability level by

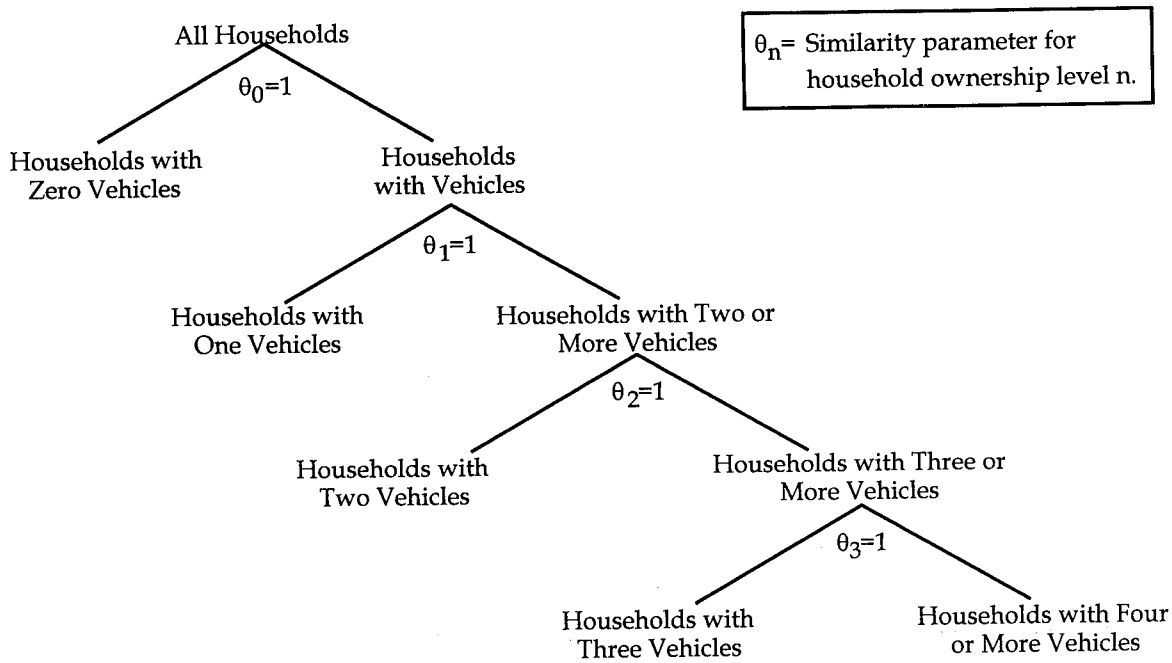
[8]Cambridge Systematics, Inc., *Enhancement of DVRPC's Travel Simulation Models: Task 10, Vehicle Availability Model*, Philadelphia, Pennsylvania, April 1997.

# Figure 3.1 Alternative Vehicle Availability Model Structures - Philadelphia Model

*Multinomial Logit Structure*




*Ordered Response Logit Structure*



making a sequence of decisions: first whether or not to have a vehicle, then whether to have one vehicle or more than one, et cetera. At each stage, if the lower level of vehicle availability is selected, the process is concluded and the household has made its current choice of vehicle availability level. Following the usual practice, the entire MNL model can be estimated in a single statistical estimation step. Estimation of the ORL model, however, involves estimating one less model than the total number of alternatives (four submodels in the case of the Philadelphia model). The first estimation uses all households, because all households choose to have either zero or one or more vehicles. For each subsequent estimation however, the observations are limited to only those households which choose one of the two alternatives available: one or two or more, two or three or more, et cetera.

The second extension involved combining highway and transit accessibility variables similar to those used in the Seattle model with pedestrian environment variables similar to those included in the 1992 Portland model. With respect to the pedestrian variables, the same estimation strategy was used to obtain zonal values of four variables. One variable was changed, however: since hilliness is not a significant factor in Philadelphia, a measure based on the typical building setbacks in a zone replaced the topography measure used in Portland. Another change in the definition of the PEF was the use of a weighted sum of the three components based on nonmotorized mode choice modeling results which indicated that the following relative weights should be those included in the definition of the PEF variable:

PEF = 0.25 * Sidewalk Availability + 0.30 * Ease of Street Crossings +
0.40 * Building Setbacks

The fourth pedestrian variable, street connectivity, was less negatively correlated with vehicle availability than the three remaining variables.

The final extension involved the use of dummy variables to reflect the unlikelihood of households choosing to have more than one vehicle per person in the household. These variables were set to one in a particular alternative if the number of vehicles in the alternative (three, for example) exceeded the number of persons in the household (two, for example).

Because the two model structures could not be compared statistically at the model estimation stage, it was necessary not only to estimate both structures, but also to conduct disaggregate and aggregate validation for both models before selecting the one which best replicates the observed data. In the Philadelphia case, although the two models both performed very well, the ORL structure had a small but consistently higher level of accuracy. Thus, it was selected for inclusion in the updated Philadelphia regional model system. Table 3.5 summarizes each of the four submodels which make up the complete ORL model. As desired, the model includes variables reflecting household, zonal density, pedestrian environment, accessibility, and persons per household dummy variables. The strongest variables are the natural logarithm of household income (found to provide a better statistical fit than income in dollars without any transformation), workers per household, and the vehicles per person dummies. Area type (CBD, urban, suburban and rural) dummy variables were also tested but were not needed. This is a desired outcome; these somewhat arbitrary zonal classifications are replaced in the model with more quantitative density and accessibility variables.

# Table 3.5 Specification of the Philadelphia Vehicle Availability Model

| Variable | Vehicle Availability Decision[1] | | | |
|---|---|---|---|---|
| | 0/1+ | 1/2+ | 2/3+ | 3/4+ |
| Persons per Household | 0.1037 | 0.1930 | - | 0.1064 |
| | 1.2 | 3.2 | - | 0.6 |
| Workers per Household | 0.1239 | 0.6816 | 1.032 | 0.5273 |
| | 0.8 | 7.1 | 9.5 | 2.9 |
| Population Density[2] | -0.03037 | -0.03708 | - | - |
| | 4.0 | 4.3 | - | - |
| Employed Person Density[2] | - | - | -0.02418 | -0.03856 |
| | - | - | 0.9 | 0.6 |
| Ln(Household Income)[3] | 1.454 | 1.383 | 0.4380 | 0.1276 |
| | 10.0 | 10.4 | 2.5 | 0.3 |
| Pedestrian Environment[4] | -0.4433 | -0.2772 | - | - |
| | 1.7 | 1.8 | - | - |
| Transit/Highway Access Ratio[5] | -1.340 | -1.099 | -0.7058 | - |
| | 2.0 | 2.7 | 1.6 | - |
| Persons Less than Vehicles[6] | - | -2.668 | -0.8832 | -0.3987 |
| | - | 8.8 | 4.9 | 0.8 |
| Alternative-specific Constant | -0.2840 | -4.156 | -4.182 | -3.644 |
| | 0.4 | 7.4 | 6.2 | 2.1 |
| Number of Observations | 1,993 | 1,837 | 1,162 | 308 |
| Rho-squared with respect to zero | 0.732 | 0.439 | 0.302 | 0.414 |

Notes:

1 Each column represents a submodel having the two alternatives shown. Except for the first model, each is conditional on at least the smaller number of vehicles being available. In addition, in each submodel the utility of the first alternative equals zero, and the utility of the second is as defined in the column. Each cell of the table contains the estimated coefficient (if any) in the top row, and the estimated t-statistic in the bottom row. T-statistics greater than 1.96 indicate statistically significant coefficients at the 95 percent confidence level.

2 The units are persons per acre and total employed persons per acre.

3 This variable is the natural logarithm of annual household income in thousands of 1989 dollars.

4 This variable is a weighted sum of the four pedestrian environment assessment measures discussed in the text. The value of the sum is in the range 0.95 to 2.85.

5 This variable is the ratio of the percentage of total regional employment which can be reached in 80 minutes by transit from the origin zone to the percentage of total regional employment that can be reached in 60 minutes by highway from the origin zone.

6 These variables equal 1 if the number of persons in the household is less than the minimum number of vehicles in the alternative, and 0 otherwise.

### *Evaluation*

The Philadelphia model contains the full range of desired variables expected to affect vehicle availability. Its structure was selected after testing both ORL and MNL alternatives. For regions which have household travel survey data, a transit system which provides a sufficient level of service to attract noncaptive riders, and significant variations in the pedestrian environment by zone, this type of advanced vehicle availability model should be considered when new models are to be developed.

There are a number of potential limitations of the Philadelphia model, however. In future model development efforts, additional model structures could also be tested - two possibilities are a nested logit structure (which was attempted for the Philadelphia model, but would not converge to a stable set of coefficients) and a probit ordered response structure. Also, more objectively measured pedestrian environment variables could be developed, possibly using a GIS system to determine more detailed versions of the components of the PEF variable. Finally, more inclusive accessibility variables without arbitrary time cutoffs could be based on *logsum* variables obtained from logit mode choice and/or destination choice models.

# 4.0 Innovative Approaches

The models discussed in Sections 2.0 and 3.0 provide a broad range of alternative approaches to vehicle availability modeling for consideration by MPOs and statewide model developers. Depending mainly on data availability and the need to incorporate accessibility and pedestrian environment factors in the forecasts for a particular region, modelers can choose a development strategy which will provide for the current needs for vehicle availability forecasting within any state of the practice travel forecasting process. If, on the other hand, new approaches to regional modeling such as activity-based systems and/or predicting vehicle fleet compositions are foreseen, then more detailed household modeling, including greater detail on vehicle availability, vehicle type choice, and vehicle usage will probably be required. This section discusses examples of innovative approaches which have been taken to address issues such as these. These innovative approaches have not reached the stage of complete implementation within state of the practice model systems; instead, they have been developed in the academic research community to address issues of vehicle fleet composition, usage, and energy requirements, or to explore the potential of using household panel survey data (information collected from the same set of households at two or more points in time) and stated-preference survey data to develop more detailed activity/travel forecasting procedures. Examples of recent work in both of these areas are discussed in the subsections which follow.

## 4.1 Vehicle Type Choice Models

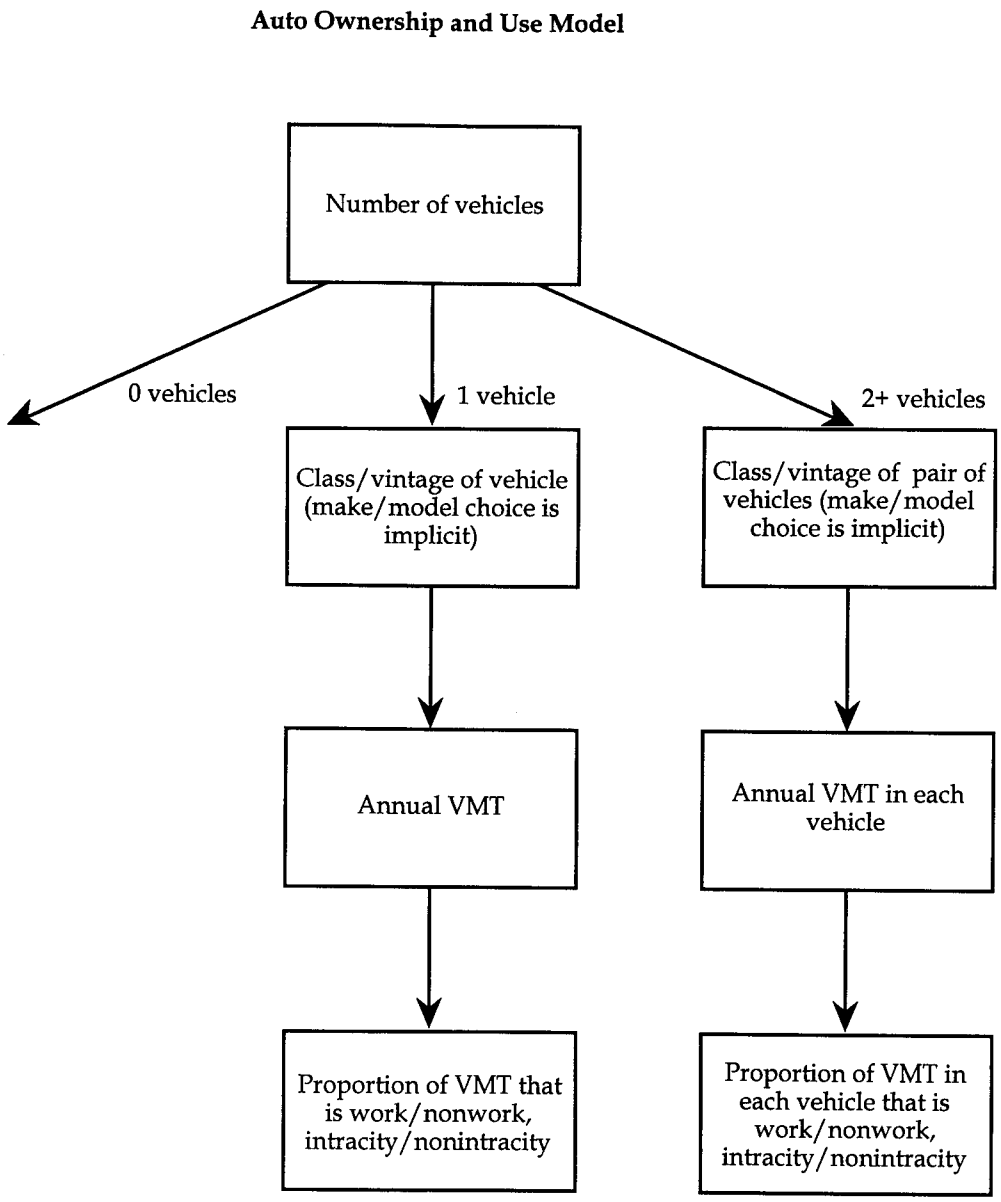
### 4.1.1 Description

In 1986, Kenneth Train published the results of his development of a combined model of household vehicle availability and vehicle type choice.[9] This model is based on a national sample of 1,095 households conducted in 1978, which includes socioeconomic data, detailed vehicle type data, vehicle usage information, vehicle purchase and sale information, and a one-day trip diary for each household member. This survey data was supplemented by information on the characteristics of more than 2,000 makes and models of 1967-1978 vintage vehicles, including physical dimensions, operating characteristics and costs, repair records, prices, and fuel efficiency. Information on the population of the household's metropolitan area and the number of transit trips in the region was also used.

The overall structure of the auto ownership and use model in shown in Figure 4.1. The first component is a vehicle quantity or auto ownership submodel which includes household

[9] K. Train, *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, MIT Press, Cambridge, Massachusetts, 1986.

# Figure 4.1 Structure of Train's Auto Ownership and Use Model




Source: K. Train, *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand,* MIT Press, Cambridge, Massachusetts, 1986.

socioeconomic and regional transit usage variables. The latter serves a similar purpose to the transit accessibility variables used in some of the models discussed in Section 3.0. In addition, the ownership model includes a variable based on the average utility of the household's vehicle class and vintage choice submodel. These submodels are specific to the number of vehicles owned by the household (one or two or more) and include variables reflecting purchase price, operating cost, shoulder room, luggage space, horsepower, vehicle age, and vehicle type (passenger car, pickup, or van). Additional submodels, not relevant in the regional modeling context, deal with annual vehicle miles of travel (VMT) per vehicle and the split of VMT into work and nonwork components.

The specification of the auto ownership submodel is included in Table 4.1. The coefficient of the average utility of the class/vintage submodel, 0.635, is highly significant, indicating that the tradeoffs between different vehicle classes and vintages are much more closely related within a given vehicle ownership level than is the joint tradeoff of number of vehicles and vehicle type.

### 4.1.2 Evaluation

By providing a linkage between auto ownership, vehicle class and vintage, and vehicle usage, the Train model system has the form of a multiply nested logit model in which each of these choices is an interrelated household decision. These interrelationships provide a more accurate means of representing the household's vehicle ownership and usage choice process; one that goes beyond the requirements of current regional models. However, as future regional models which include additional components of household activities are developed, the importance of vehicle type modeling as well as vehicle ownership modeling is expected to increase.

## 4.2 Dynamic Models of Vehicle Availability

### 4.2.1 Description

As an alternative to estimating how many vehicles are available to a household at a given point in time, models can be developed to predict how households will change their vehicle availability as the household changes, its vehicles get older, the transportation system changes, and both vehicular operating costs and purchase costs change. Models of this type require time-series data for estimation, and typically also include vehicle type considerations as in the Train model. Household panel surveys, conducted at two or more points in time, are required for these models. In addition, because dynamic models are often concerned with how auto ownership patterns will change as new vehicle types become available, information may also be required on how households will respond to new vehicle types. Stated-preference surveys are designed to obtain information on hypothetical choices in experimental designs which facilitate the determination of tradeoffs between jointly varying characteristics of these not-yet-available choices.

# Table 4.1 Specification of Train's 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 | 0.181 | 0.43 |
| 6. Log of number of members in household, entering two-vehicle alternative | 0.197 | 0.39 |
| 7. Annual number of transit trips per capita in household's area of residence, entering one-vehicle alternative | -0.0009 | 1.82 |
| 8. Annual number of transit trips per capita in household's area of residence, entering two-vehicle alternative | -0.0021 | 3.42 |
| 9. Average utility in class/vintage choice | 0.635 | 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 |

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 convergence: -475.03.

Source: K. Train, *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, MIT Press, Cambridge, Massachusetts, 1986.

A long-term household activity and transportation panel survey has been performed for some time in the Netherlands and a number of researchers have analyzed its results and explored innovative ways in which the additional information provided in these surveys can be used to develop more detailed models of household's activities and travel patterns. In the United States, the only MPO-based survey in which data have been collected for a number of points in time is the Puget Sound Regional Council panel survey. The data provided by this survey have been analyzed extensively by academic researchers and used for a number of explorations of activity-based regional model systems. Similar panel surveys conducted by non-MPO agencies have also been used to provide the behavioral input to more detailed vehicle availability and usage modeling efforts. The situation is very similar with respect to the collection and usage of stated-preference data. This type of information was recently collected as part of Portland Metro's 1996 household surveying process, but has not yet been used for regional model development.

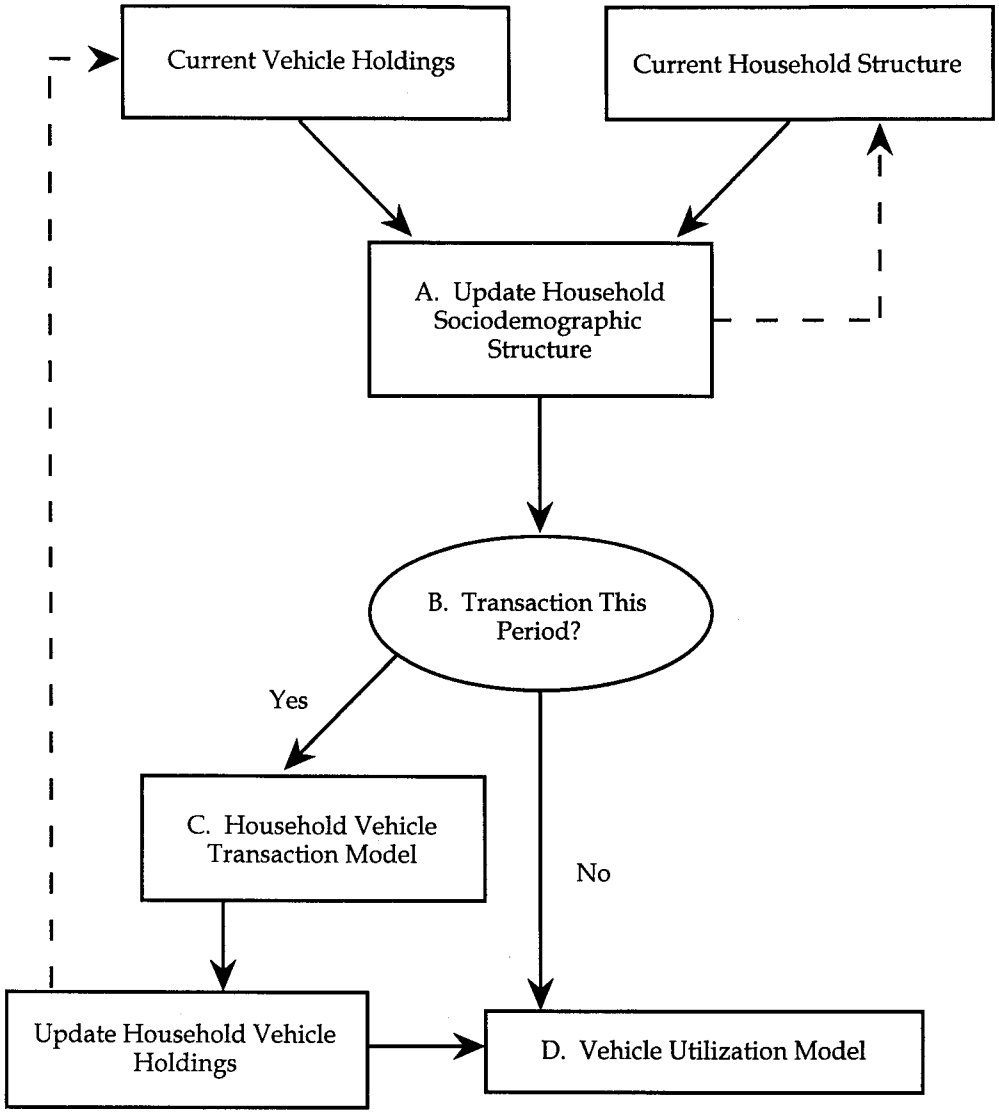
An example of dynamic models of vehicle availability and vehicle type choice is the model development plan presented in a 1994 paper by Brownstone, Bunch, and Golob.[10] They describe the use of a personal vehicle panel survey which includes stated-preference components to develop a vehicle forecasting system made up of a large number of submodels, including a used car and scrappage model and a personal vehicle model which estimates for a number of points in time how changes in the household, in energy costs, in vehicle costs, and in vehicle type availability will affect vehicle purchases and utilization. Figure 4.2 provides an overview of the components of this personal vehicle submodel. The stated-preference survey designed to support this model development process was carried out in two waves in 1993 and 1994. This survey provides information, at both points in time, on household structure, vehicle inventory, housing characteristics, basic employment and commuting for all adults, the next expected vehicle transaction, and a stated-preference component which explores the choice of hypothetical vehicles which include both clean-fuel and gasoline-fueled alternatives.

### 4.2.2 Evaluation

The development of expanded models of vehicle purchases and sales or scrappage, and vehicle utilization, will represent major extensions in data requirements, development time, and development costs for models related to household vehicles. These extensions, however, will be required in the future if fully dynamic activity-based travel models are to be developed. Panel survey data will be required to provide information on changes over time, and stated-preference data will be required if the models are to be made sensitive to new vehicle types such as clean-fuel vehicles. As MPOs and statewide planners project the evolution of their models to incorporate increasing levels of activity forecasting in the future, they must also plan for developing much more detailed models of vehicle availability and usage.

[10] D. Brownstone, D. Bunch and T. Golob, *A Demand Forecasting System for Clean-Fuel Vehicles*, presented at the OECD conference on Fuel Efficient and Clean Motor Vehicles, March 1994.

## Figure 4.2 Structure of Dynamic Personal Vehicle Submodel




Source: D. Brownstone, D. Bunch and T. Golob. *A Demand Forecasting System for Clean-Fuel Vehicles* , presented at the OECD conference on Fuel Efficient and Clean Motor Vehicles, March 1994.

# 5.0 Report Summary

A number of approaches have been taken to vehicle availability modeling by transportation planners and researchers at metropolitan planning organizations, state departments of transportation, and academic research units. Although the number of vehicles available to a household is not always predicted explicitly as part of the regional travel forecasting process by MPOs, when it is, the most common approach is to forecast it as a function of other socioeconomic variables. The most common variables used are household income, size, and location; but additional household characteristics and locational descriptors are also often used. Models of this type represent the *basic practice* adopted by MPOs at the current time; they can typically be developed using either Census or travel survey data using a number of model estimation methods: linear-in-parameters regression using zonally aggregated data; cross-classification analyses of data sets with individual households as the basic unit; and choice models, typically with a logit structure, based on individual household observations. Models of these types are relatively easy to develop; their major limitation is that they provide no explicit representation of differences in transportation services and their impacts on vehicle availability.

The distinguishing characteristic of *advanced practice* vehicle availability models is that they include not only household socioeconomic and locational variables, but also variables which are related to the ease of pedestrian travel, and/or to the transportation facilities and services available to each household. Pedestrian environment variables typically reflect factors such as the ease of street crossing, building setbacks, sidewalk continuity, street connectivity, and topography at the zonal level. Models which include these variables show the negative relationship which exists between the quality of the pedestrian environment and the level of vehicle availability. Variables related to highway and transit system characteristics are typically accessibility measures such as the percentage of regional total employment or of regional retail employment which can be reached by a stated mode of travel within a specified number of minutes. Alternatively, accessibility measures can be derived from mode choice models. Because these additional variables depend on both zonal and transportation network characteristics, they complicate the vehicle availability model development and application process, but they also provide a means of including the observed linkage between transportation levels of service and vehicle availability. These models reflect the increases in vehicle availability with improved highway systems, and the decreases with improved transit systems.

In addition to the basic and advanced practice models presently in use by MPOs, a number of *innovative approaches* to vehicle availability modeling have been explored by transportation researchers. A number of these approaches provide extensions of current practice which are likely to be useful in connection with household microsimulation and activity modeling in the future as these extensions begin to replace current transportation forecasting procedures. Examples of these innovative approaches include vehicle type choice models and dynamic models of vehicle availability. Vehicle type choice models deal not only with the number of vehicles available to a household, but also with the characteristics of these vehicles. They provide a means of forecasting the impacts of

future changes in vehicle technology such as electric-powered autos and smaller, more fuel-efficient vehicles. Dynamic models of vehicle availability explicitly include households' decision-making process with respect to vehicle scrappage and purchase in response not only to the number and age of their existing vehicles, but also to changes in household characteristics such as household size, number of workers, number of licensed drivers, and household income. Both of these types of models present new challenges in terms of their data requirements and model complexity, but also provide the ability to answer new kinds of questions concerning household activity and travel behavior, and the impacts of this behavior on the operation of future transportation systems.

## NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United State Government assumes no liability for its contents or use thereof.

The United States Government does not endorse manufacturers or products. Trade names appear in the document only because they are essential to the content of the report.

This report is being distributed through the U.S. Department of Transportation's Technology Sharing Program.

**FHWA-EP-01-005**

FHWA-EP-01-005

# *TECHNOLOGY SHARING*

**A Program of the U.S. Department of Transportation**