Enhancement of DVRPC'S
Travel Simulation Models

TASK 10

VEHICLE AVAILABILITY MODEL

PREPARED FOR
DELAWARE VALLEY REGIONAL
PLANNING COMMISSION

BY
CAMBRIDGE SYSTEMATICS, INC.

APRIL 1997

Delaware Valley Regional Planning Commission

Enhancement of DVRPC'S
Travel Simulation Models

TASK 10

VEHICLE AVAILABILITY MODEL

PREPARED FOR
DELAWARE VALLEY REGIONAL
PLANNING COMMISSION

BY

CAMBRIDGE SYSTEMATICS, INC.

APRIL 1997

Delaware Valley Regional Planning Commission
The Bourse Building
111 S. Independence Mall East
Philadelphia, PA 19106-2515

This report has been prepared by Cambridge Systematics, Inc., in partial fulfillment of the contract between the Delaware Valley Regional Planning Commission and Cambridge Systematics, Inc. to enhance DVRPC's travel simulation models. The preparation of this report was funded through federal grants from the U.S. Department of Transportation's Federal Highway Administration (FHWA) and the Pennsylvania and New Jersey Departments of Transportation. Cambridge Systematics, Inc. however is solely responsible for its findings and conclusions, which may not represent the official views or policies of the funding agencies.

Created in 1965, the Delaware Valley Regional Planning Commission (DVRPC) is an interstate, intercounty, and intercity agency which provides continuing, comprehensive and coordinated planning for the orderly growth and development of the Delaware Valley region. The region includes Bucks, Chester, Delaware, and Montgomery counties, as well as the City of Philadelphia in Pennsylvania; and Burlington, Camden, Gloucester, and Mercer counties in New Jersey. The Commission is an advisory agency which divides its planning and service functions between the Office of the Executive Director, the Office of Public Affairs, and three line Divisions: Transportation Planning, Regional Planning, and Administration. DVRPC's mission for the 1990s is to emphasize technical assistance and services, and to conduct high priority studies for member state and local governments, while determining and meeting the needs of the private sector.

The DVRPC logo is adapted from the official seal of the Commission and is designed as a stylized image of the Delaware Valley. The outer ring symbolizes the region as a whole while the diagonal bar signifies the Delaware River flowing through it. The two adjoining crescents represent the Commonwealth of Pennsylvania and the State of New Jersey. The logo combines these elements to depict the areas served by DVRPC.

DELAWARE VALLEY REGIONAL PLANNING COMMISSION

Publication Abstract

TITLE	Date Published: April, 1997
Vehicle Availability Model	Publication No. 96017

Geographic Area Covered:

Delaware Valley Region

Key Words:

Vehicle availability, Travel demand models

ABSTRACT

This report presents recommendations for modeling household vehicle availability in the DVRPC travel model system, in which both trip generation and mode choice models use the levels of vehicle availability as an independent variable. The recommended model represents an important advancement of the current demographic projection method by considering not only household characteristics such as household size, income, and number of workers; but also zonal density, pedestrian environment, and accessibility measures. Because the latter measures vary both with the location of employment in the region and with travel times by mode between residences and jobs, the recommended model has a high level of sensitivity to land use, transportation system, and urban design policies as well as to demographic and socioeconomic characteristics by zone.

The report presents the recommended ordered response logit model of vehicle availability, discusses its accuracy in replicating observed base year data, and compares its predictions for 2020 with DVRPC's prior predictions. Procedures for implementing the recommended model are also presented.

For More Information Contact:

Delaware Valley Regional Planning Commission
Regional Information Services Center
The Bourse Building
111 South Independence Mall East
Philadelphia, PA 19106-2515
Telephone: (215) 592-1800 FAX: (215) 592-9125

Table of Contents

List of Tables

List of Tables
(continued)

List of Figures

Executive Summary

This report describes the recommended method for modeling household vehicle availability within the DVRPC travel model system. The method consists of an ordered response logit model which predicts the individual household's sequential process of first choosing whether to have at least one vehicle or not to have any vehicles. Next, a choice between having just one vehicle or having more than one vehicles is assumed; this process continues, if no lower level is selected, until a final choice is made between having three vehicles, or having four or more vehicles. This model structure was selected after developing two alternative models, the other corresponding to the assumption that households make a one-time decision (until its size, income, or location changes) of whether to have zero, one, two, three, four, or more vehicles. Tests of the accuracy of the two alternative model structures showed that the ordered response structure provides a marginally more accurate model than the structure that assumes a one-time decision.

The alternative models were both developed using individual household data obtained from 1,993 respondents to DVRPC's 1987 travel survey. To the socioeconomic and demographic variables provided by the survey were added zonal activity level data (household, population and employed person densities); zonal pedestrian friendliness measures including sidewalk availability, ease of street crossings, and building setbacks; and zonal accessibility measures based on the fraction of the total region's employment which can be reached within specified travel times by both highway and transit.

Both alternative models were estimated using the logit maximum likelihood estimation procedure provided by the ALOGIT program. This estimation process involved testing a large number of alternative model specifications to find the subset of variables for which statistically significant coefficients were found with the signs expected, based on economic and behavioral factors. In both models, the most important variables are the number of workers per household and household income level, but variables of each of the types discussed above proved to be significantly related to vehicle availability levels.

The models based on both alternative model structures were tested extensively. Comparisons of observed and predicted vehicle availability levels for the households in the estimation data set aided in determining that the models replicate the observed differences in behavior for market segments based both on the variables in the model, and on other variables, such as the area type of the residence zone, which did not enter the models. During the early stages of the estimation process, these comparisons also aided in specifying the exact forms of the variables to be used in the models. As an example, these comparisons indicated that a logarithmic function of household income provided a better fit to the observed data than the income variable itself.

Model testing and validation was also performed using individual households from the U.S. Census Public Use Microdata Samples (PUMS) for the DVRPC portions of Pennsylvania and New Jersey, and using aggregations of the full U.S. Census CTPP data

for the DVRPC zone system. Both of these tests required higher levels of aggregation than that used for model estimation. With the PUMS data, it was necessary to aggregate all zonal information to the PUMA districts, of which there are just 39 for the entire DVRPC region. With the CTPP data, all demographic variables were only available as zonal totals or averages. For both tests against Census data, these higher levels of aggregation resulted, as expected, in underestimation of the numbers of households in the lowest and highest vehicle availability categories, and overestimation in the categories closest to average values.

Using the PUMS data, the predicted regionwide average number of vehicles per household is within seven percent of the observed average for both models. With the exception of two highly urban counties, Mercer and Philadelphia, the predicted county-specific averages all have errors of less than five percent of observed averages. The errors in Mercer and Philadelphia counties are nine and 25 percent, respectively, suggesting that their more congested urban characteristics have impacts on vehicle availability which are not captured completely in either of the two alternative models. When the two models are compared, the ordered response model is slightly more accurate — from 0.2 to 1.2 percentage points for all counties except Philadelphia, where it is 0.2 percentage points less accurate.

Using the higher level of aggregation which underlies the zonal averages of Census data, the predicted regionwide average number of vehicles per household is overpredicted by 19 percent of the observed average for both models. Since this test of the aggregate application of the models corresponds to the method which will be used by DVRPC for forecasting, the estimated models were adjusted to ensure that they predict the observed shares of households by vehicle availability level. Based on the results obtained by county with the estimated models, these adjustments to the alternative-specific constants were done separately for three subsets of the entire region: Philadelphia County, Mercer County, and the remaining seven counties taken together. Following these adjustments, the largest absolute county-specific error in the average number of vehicles available per household is 6.5 percent, and the predicted regionwide shares for the individual levels of vehicle availability are all replicated within 0.7 percent for the one-time decision model and within 0.1 percent for the ordered response model.

Following the adjustments required to apply the alternative models at the zonal level, it was possible to compare the errors in predicting average zonal vehicle availability levels of the two models, and based on these errors, to select the final recommended model. These comparisons revealed that the ordered response model results in marginally lower errors, expressed using four measures: average positive, negative, and absolute deviations, and the root mean square error (RMSE) expressed as a percentage of the regionwide average vehicle availability per household. For the most common of these error measures, for example, the ordered response model has a %RSME of 13.2 percent and the alternative one-time decision model's value is 15.3 percent.

Careful analysis of the results obtained when the selected ordered response model was applied to the DVRPC zonal data revealed that at the aggregate level, the initial adjusted model was biased, overpredicting vehicle availability in low density zones and underpredicting in high density zones. This bias was removed by revising the density variables used in the ordered response models and estimating the coefficients of these new variables in an iterative process which focused on removing the density-related biases. At the same time, the model's alternative-specific constants were further revised to maintain the observed vehicle availability shares by county group. During this process, the number of county groups was increased to four by defining Camden County, in addition to Philadelphia and Mercer, as a county for which separate model coefficients and constants were determined.

The final step in model development was to predict vehicle availability levels by zone for the year 2020 and to compare these predictions with DVRPC's current estimates based on demographic projection methods. The inputs to these 2020 estimates were obtained using DVRPC's predictions of 2020 levels of households, population, employed workers by residence zone, and employment by work zone. Also required were projections of future year highway and transit zone-to-zone travel times, pedestrian environment levels by zone, and average household income levels by zone. The 2020 travel times by mode were assumed to remain at the same relative values as in 1990; as a result, the accessibility measures for 2020 only change due to the predicted redistribution of employment. The 2020 pedestrian environment variables by zone were also assumed to remain unchanged from their 1990 values. A projection method based on changes in persons per household and workers per household was used to project 2020 values of zonal average household incomes.

When the results of the zonal vehicle availability levels predicted by the recommended model are compared with DVRPC's current projections, the recommended model is found to provide slightly lower averages by county and for the region as a whole. At the regional level, the difference is -4.4 percent. The differences in the county averages range from minus seven to zero percent. Also, the percent root mean square difference of the individual zonal values is 11.7 percent, nearly as low as the corresponding error when the model results are compared with the observed 1990 zonal data. Thus, the recommended model is highly consistent with DVRPC's current projections of future year vehicle availability levels.

It is recommended that the existing DVRPC modeling process be revised to add the application of the recommended vehicle availability model and household income projection process prior to the trip generation step. In addition to the zonal demographic and other variables also used in trip generation, the required inputs will include zonal pedestrian environment and accessibility variables. The pedestrian environment variables for future years should be obtained by revising the base year values as necessary to reflect expected changes in sidewalk availability, ease of street crossings, and building setbacks as new developments and redevelopments are projected. The future accessibility variables must be based on the future projections of employment by zone and the highway and transit travel times based on the projected unloaded highway and transit networks for the future scenario. Using these inputs, the full set of steps which will be required to estimate future vehicle availability shares by zone will be the following:

- Project future zonal household income levels using the procedure presented in Section 4.2;

- Compute future zonal highway and transit accessibilities using future employment by zone and zone-to-zone travel time skims from the future year unloaded highway and transit networks; and

- Compute zonal numbers of households having 0, 1, 2, and 3+ vehicles using the recommended ORL model, presented in Table 2.4 with the variable, coefficient, and constant revisions shown in Table 3.8 and provide the results as part of a zonal data file for input to the revised trip generation modeling process.

These new vehicle availability prediction steps have been implemented as part of the model development process using Microsoft Access procedures.

1.0 Introduction

As with most travel demand modeling systems, a key descriptive characteristic within the DVRPC model is how many vehicles are available to, or automobiles are owned by, particular households. In this project, the term vehicle availability is used rather than the previously more common term auto ownership 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. Thus, although these two terms are sometimes used synonymously, the more precise term vehicle availability is used throughout this report.

The objective of Task 10 of the *Enhancement of DVRPC's Travel Simulation Models* project is to develop, implement, and test alternative approaches for predicting household vehicle availability levels. This report presents the recommended revised vehicle availability forecasting procedure as well as the results of the model development, implementation, and validation process. In addition, since the recommended model includes household income as a determinant of vehicle availability and DVRPC did not previously forecast average zonal income levels, a simplified procedure to project this variable was developed. This procedure is also documented in the report.

The report is organized as follows. This section discusses the objectives of the vehicle availability modeling process and provides an overview of the technical approach taken to accomplish these objectives. Section 2.0 discusses the model estimation process and the results obtained for the two alternative model forms. Section 3.0 presents model validation results at both the disaggregate and aggregate levels and the choice of a final recommended model based on these results. Section 4.0 documents the results obtained when the recommended model is used to forecast zonal vehicle availability levels for the year 2020 and compares these results with those previously forecast by DVRPC. This section also presents the method used to project future average zonal household income levels. Finally, Section 5.0 summarizes the recommended procedures for incorporating the new vehicle availability forecasting process into DVRPC's modeling process.

■ 1.1 Objectives of the Vehicle Availability Modeling Process

DVRPC currently forecasts the distribution of household vehicle availability within each zone through an analysis of Census trends. This aggregate time-series modeling approach is relatively straightforward and understandable, but it is limited by the small amount of available input data and by the possibility of aggregation errors. Furthermore, the current approach does not explicitly consider each of the following potentially significant determinants of household vehicle availability, except to the extent that they are reflected implicitly in the current vehicle availability levels provided by base year Census data:

- Variations in household characteristics such as household size, number of workers, and household income;
- Variations in zonal activity levels as measured by population, employment, and household densities;
- Variations in levels of 'pedestrian friendliness' by residence zone, as measured by building setbacks, ease of street crossings, local street network connectivity, and sidewalk availability; and
- Variations in levels of mode-specific accessibility by residence zone, as measured by highway congestion and/or various measures of the numbers of jobs surrounding the zone and the travel times by highway and transit required to reach them.

In this task, Cambridge Systematics has developed, implemented, and tested alternative approaches for predicting vehicle availability levels. These approaches are based on disaggregate (household-level) data, which provide the opportunity to improve the level of precision of DVRPC's vehicle availability forecasts. At the household level, vehicle availability levels are influenced by many of the types of variables listed above. In the household level vehicle availability modeling effort, Cambridge Systematics has used DVRPC's base year travel survey and zonal data to develop mathematical models that relate vehicle availability to other household characteristics and to transportation system characteristics that can be forecast.

The model outputs are the predicted shares of households having different numbers of vehicles (e.g., none, one, two, three, and four or more) for each zone. This information will feed into the updated DVRPC trip generation and mode choice models.

■ 1.2 Overview of the Technical Approach

In previous vehicle availability modeling efforts, two general analysis approaches have been used:

- A multiple regression (cross-classification) formulation; and
- A discrete choice (logit, probit, nested logit) formulation.

The two formulations are similar in that they are designed to relate the number of vehicles available by a household to a limited number of explanatory household, person, zonal, and transportation system variables. Their primary structural difference is that the regression formulation typically relates the explanatory variables to average vehicle availability levels, while discrete choice formulations relate the explanatory variables to the fractions of households in each vehicle availability category (e.g., zero, one, two, three, four or more vehicles available). The two disaggregate methods rely on similar base year data for estimation, and if a form of the logit model (described below) is the chosen discrete choice method, both formulations can be developed using a number of different statistical software packages. The regression formulation is more straightforward in terms of model specification and application, but the discrete choice approach has the advantage of more closely replicating actual household vehicle availability decisions.

The discrete choice models provide the opportunity to include more potential variables than cross-classification models can handle easily.[1] In addition, because other elements of the overall model system will be based on discrete choice formulations, using a similar structure for the vehicle availability model enhances the ability to link this model to other components of DVRPC's forecasting process. Most of the recent vehicle availability models have used the discrete choice formulation, including successful efforts in Portland, Oregon and the Bay Area.

There are three possible discrete choice model formulations or structures that can be used for the vehicle availability model. These include the multinomial logit model, the ordered response logit model, and the nested logit model. The general characteristics of each of these structures are the following:

- The **multinomial logit model** (MNL) structure is the most commonly used discrete choice model. This model relates a choice between two or more alternatives to explanatory variables by estimating mathematical expressions that seek to explain how different attributes and variables affect people's decisions. The model seeks to estimate the most likely expressions for each alternative's "utility."

- The **ordered response logit model** (ORL) structure is a mathematical variation of the common multinomial logit model. The ordered response model implies that people make a series of sequential decisions in arriving at the final decision of how many vehicles to have.

- The general **nested logit model** form can be used to obtain a more flexible version of the ordered response model structure in which some subsets of alternatives (whether to have three and four or more vehicles, for example) are more alike than others, such as whether not to have any vehicles or to have one or more vehicles.

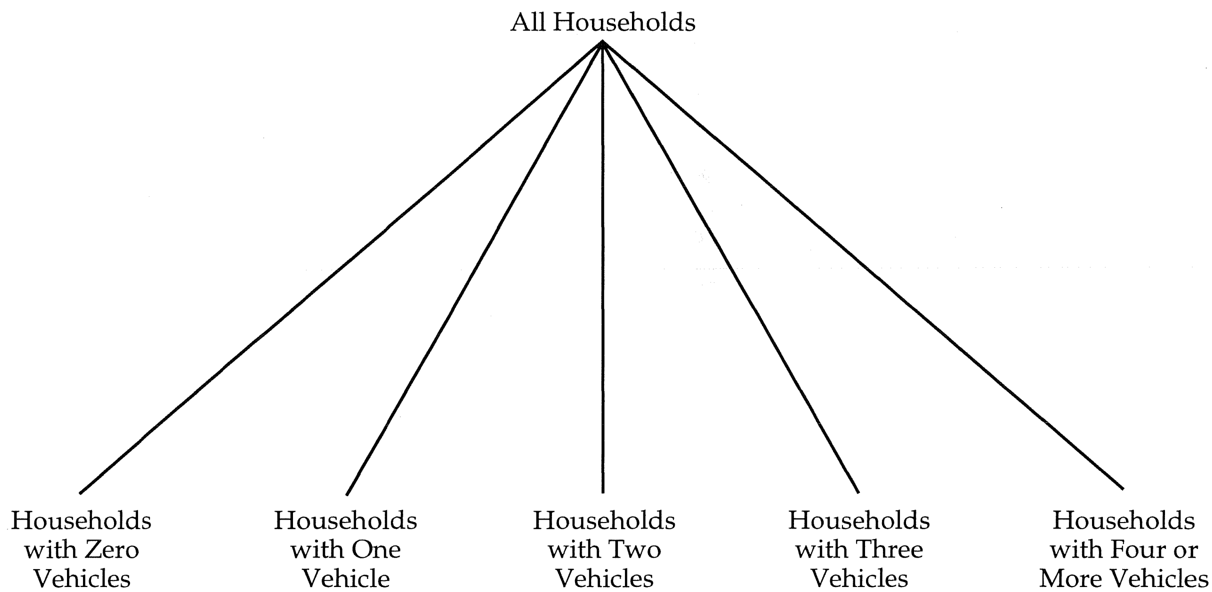
The structural differences between these discrete choice model formulations that have been investigated are shown in Figure 1.1. The first structure shows the more simple multinomial model approach. This choice structure implies a single household choice (subject to revision only when the household's characteristics or location change) of the number of vehicles to have. The second structure assumes a sequential choice by households of first whether to have any vehicles at all, and then if they do choose to have

---

[1]Charles L. Purvis, *Using 1990 Census Public Use Microdata Sample To Estimate Demographic and Automobile Ownership Models*, Transportation Research Record 1443 (National Academy Press, Washington, DC 1994). pp. 21-29.

Figure 1.1 Potential Choice Structures for the Vehicle Availability Model

Multinomial Logit Structure



Ordered Response Logit Structure

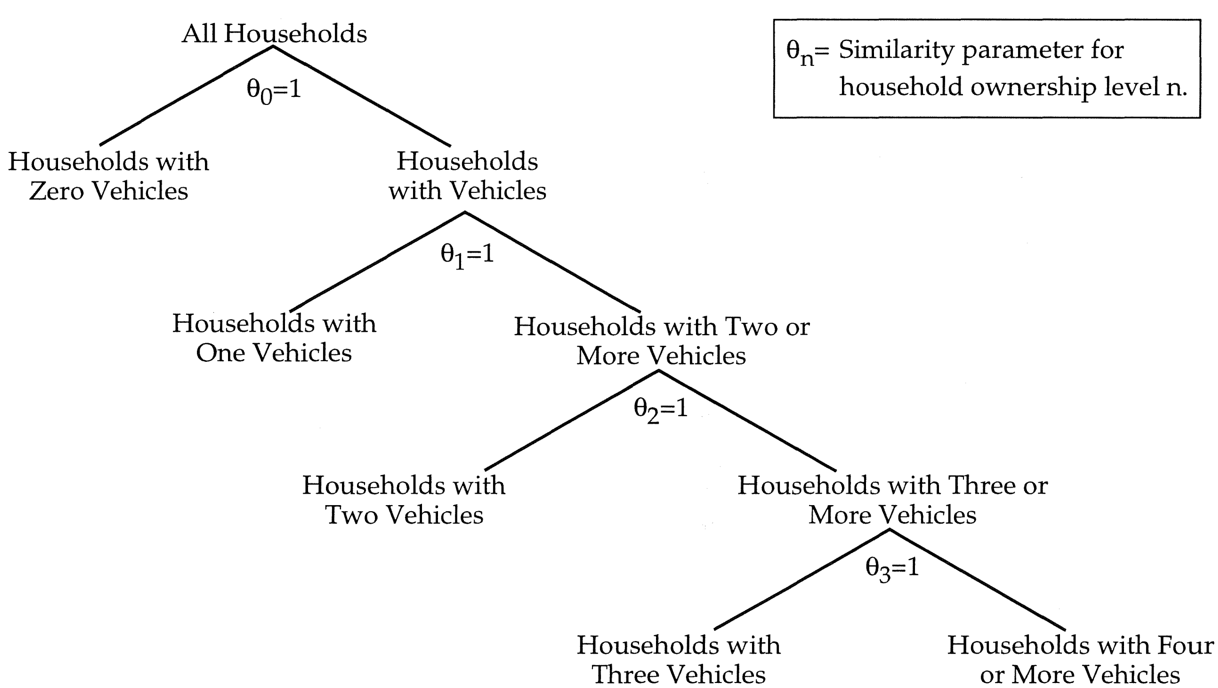
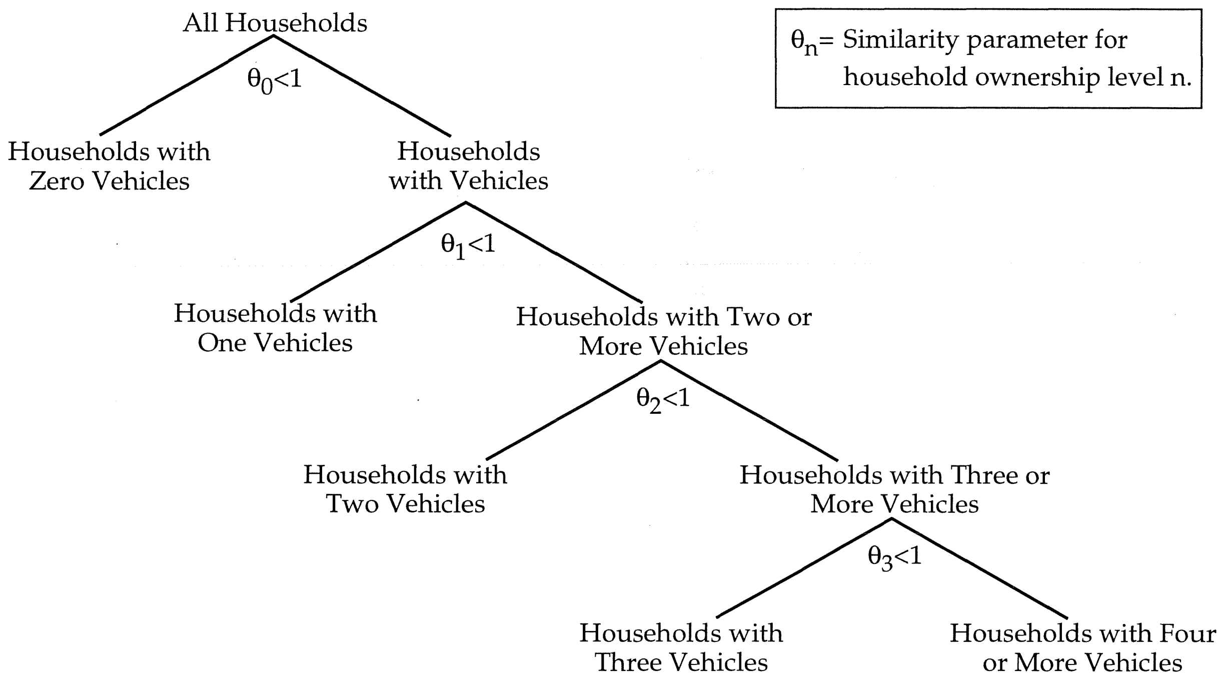


Figure 1.1 Potential Choice Structures for the Vehicle Availability Model *(continued)*

Nested Logit Structure



vehicles whether to have more than one, then whether to have more than two, etc. This structure requires an ordered model approach. The second structure also assumes that the similarity between the two choices available at each level of the choice structure (as reflected in the *theta* coefficient) are equal. The third structure also assumes a sequential choice process, but does not assume that the choices at each level of the structure are considered equally. Instead, the *theta* coefficients of this model structure can vary to provide the best model fit to the available data.

The second and third structures in Figure 1.1 seek to exploit the fact that the potential choices are ordinally scaled (the choices have a logical mathematical order; 0, 1, 2, 3, 4 or more). In most discrete choice applications, like mode choice models, the choices are nominally scaled (the choices do not imply any particular order). A recent shopping trip generation model for the Toronto region was developed using an ordered response logit model.[2] Another recent application of an ordered response modeling approach used an ordered probit formulation to model household income and employment.[3]

Since each of these potential model structures is based on an abstraction of the collective behavior of a large number of households, the final structure must be selected mainly on empirical grounds – which structure best fits the available data? Thus, we have attempted to fit each structure to household data for the DVRPC region. We have had good success in estimating models with each of the first two structures. The third structure could not be estimated, apparently because the non-linear estimation procedure required for nested logit models is not sufficiently robust for use with the combination of the nested structure shown in Figure 1.1 and the DVRPC household data set. Thus, as discussed in Section 3.0, we have selected the recommended model and structure for the DVRPC vehicle availability forecasting process from the multinomial logit and ordered response logit options.

---

[2] K. Agyemang-Duah, W.P. Anderson, and F.L. Hall, *Trip Generation for Shopping Travel*, presented at the 74th Annual Meeting of the Transportation Research Board, January 1995.

[3] Chandra Bhat and Frank Koppelman, *An Endogenous Switching Simultaneous Equation System of Employment, Income, and Car Ownership*, *Transportation Research A*, *Volume 27A(6), 1993.*

2.0 Estimation of Alternative Model Structures

This section presents the results of model estimations for two alternative structures of a DVRPC vehicle availability model. The sections which follow describe the data used for model estimation, the potential variables which affect households' vehicle availability decisions, and the estimation results for both alternative model structures. The models presented here provide two alternative structures from which a final structure has been selected, based on the results of model validation as discussed in Section 3.0, for implementation in the revised DVRPC travel forecasting system.

2.1 Data for Model Estimation

The primary data source for model development is the household file from the 1987 DVRPC travel survey. The data file provided by DVRPC provides a sample of 2,425 households. For model estimation, it was possible to use 1,993 of these households (82 percent). Most of the households not used (429) did not report a household income level. The remaining three households not used were located in zones for which no population data were available. The unweighted distributions of the accepted household records by vehicle availability level and by area type are provided in Tables 2.1 and 2.2. Previous modeling experience has shown that there should be a minimum of 50 observations for each alternative to be modeled. Because this criterion is met in the edited data set, we are able to rely solely on the DVRPC survey for model estimation. If this survey were not sufficient, we would have explored the possibility of using Census data; either from the PUMS data, which provides complete household census records, with the exception of exact location information; or from the CTPP data which includes tables of vehicle availability cross-tabulated with other household variables such as household size, number of workers, household income, housing type, and mode of travel to work. These alternative data sources have been used to validate the estimated models, showing that they accurately replicate the vehicle availability reported in the 1990 Census, as reported in Section 3.0.

The following variables potentially relevant for vehicle availability modeling are available in the household data file:

- Vehicles available per household, the dependent variable;
- Persons per household;
- Workers per household;

Table 2.1 Distribution of Households Used for Model Estimation by Vehicle Availability

Vehicle Availability	Number	Percent
0	156	7.8
1	675	33.9
2	854	42.8
3	258	12.9
4+	50	2.5
Total	**1,993**	**100.0**

Table 2.2 Distribution of Households Used for Model Estimation by Area Type

Area Type	Number	Percent
CBD	16	0.8
Urban	440	22.1
Suburban	1,134	56.9
Rural	403	20.2
Total	**1,993**	**100.0**

- Household income; and
- Area type.

The household file was expanded to form the estimation data set by adding the following data for the residence zone of each household in the file; where necessary, these variables are discussed in greater detail in the remainder of this section:

- Three measures of residential density by zone: population, household, and employed person densities in units per acre, calculated from 1990 Census data and DVRPC geographic data;
- Pedestrian environment variables developed in this project;
- Accessibility variables developed in this project, based on the percentage of total regional employment which can be reached from the residence zone within specified levels of travel time by highway and transit;
- Highway accessibility variables provided by DVRPC, obtained from the denominators of the current DVRPC trip distribution models for HBW and HBNW trip purposes; and
- Average highway volume-capacity ratios for each of DVRPC's 72 county planning areas based on volumes and capacities for all highway facilities, provided by DVRPC.

Four pedestrian environment variables were available; these are discussed in the final report for Task 9, the Non-Motorized Travel Model. Briefly, these variables are:

- Sidewalk availability;
- Ease of street crossings by pedestrians;
- Street connectivity; and
- Building setbacks.

Each is defined as an integer ranking ranging from 1 to 3, with 1 representing the lowest level of the given variable with respect to the resulting pedestrian environment, 2 a medium level, and 3 the highest level.

Eight potential accessibility variables were computed for use in model development; four for accessibility by highway and four for accessibility by transit. Each is defined as the fraction of total regional employment which can be reached within a specified level of travel time by the specified mode from the residence zone. For the highway variables, the travel time is that obtained from travel time skims over the average daily unloaded 1990 highway network developed by DVRPC prior to the start of this project. The link times in this network generally represent posted speed limits. These skimmed times exclude any off-network or terminal times, but do include times on zone centroid links as originally coded by DVRPC. The four time cut-offs used for the highway accessibility variables are 10, 20, 40 and 60 minutes.

For the transit accessibility variables, travel times were obtained from the 1990 PM unloaded transit network developed by DVRPC prior to the start of this project. These times include all components contained in the transit network including walk, wait, transfer and line-haul elements. The four time cut-offs used for the transit accessibility variables are 20, 40, 60 and 80 minutes. These longer-than-highway times were selected because the transit origin/destination times used include walk access and egress components not included in the highway cut-off values.

2.2 Potential Independent Variables

This section discusses the expected relationships between each of the independent variables included in the model estimation data set and the resulting level of household vehicle availability. These expectations provide a basis for determining the reasonableness of the statistical estimation results presented in the next section.

- **Persons per household (PPHH) –** provides a measure of the household's need for transportation for all purposes. Generally therefore PPHH is expected to be positively correlated with vehicle availability. In addition however, we would expect that the likelihood of more vehicles than people in a household would be quite low.

- **Workers per household (WPHH) –** provides a measure of the household's need for transportation to work, the least discretionary portion of total household travel, but also the portion most likely to be served by public transit. Like PPHH, WPHH is expected to be positively correlated with vehicle availability. However, since these two variables are expected to be significantly correlated with each other, it may be necessary to choose just one of them for inclusion in the estimated models.

- **Vehicles per person (VPP) –** provides a means of comparing a potential level of vehicle availability with the number of potential users. As this variable increases from zero to some number near one, it is expected to be positively correlated with household vehicle availability. Values above one, however, are expected to be relatively rare and to be associated with undesirable choices for most households. This expected discontinuous relationship suggests that VPP should be defined for a limited range (from zero to one) and set equal to one for alternatives in which the number of vehicles exceeds the number of persons.

- An alternative means of reflecting the expected relatively uncommon decision of households to have more than one vehicle per person is to define **vehicle/person dummy variables** (PLT2, PLT3 and PLT4). These variables are equal to one for the alternative numbered in the variable name if the corresponding number of vehicles exceeds the number of persons in the household, and equal to zero otherwise. (For example, PLT3, a variable in the utility expression for households choosing to have three vehicles, equals one if the number of persons in a household is one or two.) These variables are each expected to be negatively correlated with their specific level of vehicle availability.

- **Household income (INCOME)** – measures the household's resources available for vehicle purchase, maintenance, and operation. This variable is also expected to be positively correlated with vehicle availability. However, this relationship may not be simply linear; an increase in annual household income from $100,000 to $120,000 is expected to have a smaller effect than an increase from $20,000 to $40,000.

- **Density measures, including population, household, and employment (POPDEN, HHDEN, and EMPDEN)** – are expected to be negatively correlated with vehicle availability. As zonal activity levels increase, the availability of garaging and parking space is likely to decrease, as is the speed and comfort of travel by private auto. Also, transit service is more likely to be available as an alternative travel mode. Finally, more potential destinations are likely to be within walking distance. The correlation between POPDEN and HHDEN will be very high, so we do not expect both of these variables to be required in the estimated models. In most cases, employment density levels are also likely to be significantly related to the other available density variables. Due to these correlations, we do not expect all density variables to be included in the estimated models.

- The four **pedestrian environment variables** are all expected to be negatively correlated with vehicle availability, but not strongly so. Each variable provides a measure of the ease of walking or biking, either from origin to destination, or between home and transit service. Household vehicle requirements are expected to be somewhat lower in zones with higher levels of pedestrian facilities. Because this relationship is not expected to be strong and because the four variables are themselves likely to be highly correlated, the three variables found to be most negatively correlated with vehicle availability were combined to provide a single measure of zonal pedestrian environment (PEV) which could be used as a candidate variable if all four variables could not be used individually. Based on the findings of the non-motorized modeling portion of this project (Task 9), the combination of pedestrian environment variables selected for use in vehicle availability modeling is the following:

$$\begin{aligned} \text{PEV} = \ & 0.25 * \text{Sidewalk Availability} + 0.30 * \text{Ease of Street Crossings} \\ & + 0.40 * \text{Building Setbacks} \end{aligned}$$

- The four **highway accessibility variables** based on numbers of employees within specified minutes of highway travel from the residence zone (HWYEMP10, HWYEMP20, HWYEMP40 and HWYEMP60) are alternative measures of the number of jobs which can be conveniently reached by auto. Four were chosen for testing because the cut-off value is very arbitrary and should be chosen based on which variable is the best predictor of vehicle availability. Each of these variables is expected to be positively correlated with vehicle availability; having more jobs available by auto is expected to be related to higher levels of vehicle availability.

- In a similar but opposite way, the four **transit accessibility variables** (TRNEMP20, TRNEMP40, TRNEMP60 and TRNEMP80) are expected to be negatively correlated with vehicle availability – as transit service improves and more jobs can be readily reached by transit, the need for autos is expected to decrease.

- A number of accessibility variables can be constructed which combine the highway and transit measures discussed in the two previous bulleted items. **Highway/transit**

accessibility differences (HTADij, for highway measure I and transit measure J) can be formed by subtracting a selected transit accessibility measure from a selected highway measure. Consistent with the expectations discussed above, these differences are expected to be positively correlated with vehicle availability.

- Similarly, **transit/highway accessibility ratios** (THARij) can be formed by dividing a selected transit measure by a selected highway measure. These ratios are expected to be negatively correlated with vehicle availability.

- The two **trip distribution denominators** (DHBW and DHBNW) are alternative measures of highway accessibility related to the ease of reaching work and non-work destinations, respectively, by highway. As in the case of the HWYEMP variables, they are expected to be positively correlated with vehicle availability.

- The average **highway volume-capacity variable** (VC) is a measure of highway congestion which is expected to be negatively correlated with the number of vehicles available to households in an area. As highway congestion increases, more households are expected to reduce their usage of auto transportation and thus require fewer vehicles.

- Four **area types** were defined: CBD, urban, suburban, and rural. Dummy (zero or one) variables were defined for three of these (CBD, SUBURB, and RURAL), since the remaining area type is indicated when all of these three variables are zero. The CBD variable equals one for all zones with DVRPC area types of 1 or 2, CBD or CBD fringe. The suburb variable equals one for all zones with the DVRPC area type of 4, suburban. Finally, the rural variable equals one for all zones with the DVRPC area types of 5 or 6, rural or open rural. Ideally, each of these variables would not appear in the final models because all of the features of each area type which affect household vehicle availability would be captured by variables such as the density and pedestrian environment variables. However, because the available variables are quite limited, these dummies were retained as potential variables for inclusion in the models. Assuming that the missing variables would provide other measures of the difficulties of household vehicle availability in the CBD, this variable is expected to be negatively correlated with increasing levels of the dependent variable. Conversely, SUBURB and RURAL are expected to be positively correlated with increasing household vehicle availability.

2.3 Estimation Results

2.3.1 Multinomial Logit Model

A series of runs were made, beginning with one including all relevant available variables. Successive runs were then made after eliminating variables which were highly correlated with other independent variables, and variables which did not have statistically significant estimated coefficients with the expected signs. New forms of variables were also developed in a few cases. The net results of these runs were the following:

- The **persons per household** variable was found to have the expected relationship to vehicle availability in three of the four alternatives: 2, 3 and 4 or more vehicles. No statistically significant relationship could be found between PPHH and having just one vehicle rather than no vehicles, and a single coefficient was estimated for the effect of PPHH on having 2 or 3 vehicles. As expected, the coefficient for the 4+ alternative is larger than that for 2 or 3 vehicles; the number of people in the household has a greater impact on having 4+ vehicles than on having fewer vehicles.

- The **workers per household** variable is a highly significant determinant of having 2, 3 or 4+ vehicles; in each alternative, its coefficient is more significant than the corresponding persons per household coefficient. However, the workers per household variable does not significantly affect the choice of 0 or 1 vehicles. As in the case of PPHH, the magnitude of the WPHH coefficients increases for increasing numbers of vehicles.

- The **vehicles per person** variable was removed completely from the model due to its discontinuous relationship to vehicle availability.

- The **vehicle/person dummy variables** have the expected negative relationship with vehicle availability. The magnitude of the estimated coefficients decreases for increasing numbers of vehicles available, as does the statistical significance of these coefficients. Thus, all other factors being equal, larger households are more likely to have an excess of vehicles over persons than smaller households, but both types of households are more likely to have fewer vehicles than persons.

- **Household income** was replaced with the natural logarithm of household income in thousands of dollars (LNINC) because the income effect on household vehicle availability proved to be increasing, but at a decreasing rate for high incomes. The transformed household income variable is a major determinant of vehicle availability and it affects successive levels at increasing rates.

- All **population density and employment density** variables were removed because **household density** proved to be stronger and the high correlations, as expected, between these three variables made it impossible to include more than one density variable per alternative. Household density has the expected negative impact on vehicle availability which increases in magnitude as vehicle availability increases.

- The **four separate pedestrian environment** variables were replaced by the weighted sum defined above (PEV) because high correlations between them prevented the estimation of separate coefficients with the expected signs and desired significance levels. A single coefficient was estimated for the three highest vehicle availability levels (2, 3, and 4+) to ensure consistent coefficient values and to obtain a statistically significant value.

- A single form of accessibility variable, the **transit/highway accessibility ratio** (THARij), was found to provide the best relationship with vehicle availability. When separate variables for highway accessibility only (HWYEMPn, DHBW and DHBNW) were tried, the estimated coefficients were negative, probably because suburban zones with high vehicle availability are far from the major employment areas. The coefficients for separate variables for transit accessibility (TRNEMPn) had the expected negative sign, but were generally weaker statistically than the selected ratio variables. Accessibility difference variables (HTADij) generally had the expected positive signs, but the coefficients for some alternatives were negative and/or had low significance. Finally, no significant relationship was found between the highway volume-capacity variable (VC) and vehicle availability, probably because this variable was provided as an average value for each of DVRPC's 72 county planning areas rather than as a unique value for each traffic analysis zone. The specific THAR variable found to have the most explanatory power was that involving the use of the largest travel times, 60 minutes for highway and 80 minutes for transit. Separate negative coefficients were obtained for this variable for the 1- and 2-vehicle alternatives, and a combined negative coefficient was found for the 3- and 4+-vehicle alternatives. As expected, these coefficients reveal an increasing contribution of the relative transit/highway accessibility to reducing the likelihood of larger vehicle availability levels.

- Following the inclusion of the full set of variables discussed above, the incremental improvement gained by using **area type variables** was found to be very small. Since these variables are highly correlated with zonal density and accessibility measures, the small size of this improvement is not surprising. Furthermore, these alternative measures provide for continuous variations in utility levels rather than for the step function changes associated with dummy variables. For these reasons, no area type variables were included in the final multinomial model.

Table 2.3 summarizes the estimation results for the final recommended multinomial logit model. Coefficients were estimated and are shown for all alternatives except zero vehicles available, which is the base alternative – the alternative against which each of the others is implicitly compared in the estimated coefficients. Both the estimated coefficients and their t-statistics, which provide a measure of their statistical significance, are provided in the table. T-statistics greater than 1.96 indicate significant coefficient estimates. The rho-squared measure for the complete model is 0.447. Although this measure provides similar information as the r-squared statistic for a linear regression model, it typically has a value in a lower range. The value obtained indicates that the model has an acceptable level of explanatory power.

Each of the expected relationships discussed in the previous section can be used to evaluate the coefficients obtained, all of which are alternative-specific. Generally, this evaluation involves comparing the coefficients for a given variable across each of the alternatives. These evaluations reveal the following:

- The numbers of persons and workers per household do not significantly affect the choice of having zero or one vehicle;

- Multiple vehicle availability is increasingly likely in households with more persons and/or more workers and the effect and significance of more workers is greater than that of more persons;

- Multiple vehicle availability greater than the number of persons decreases at a decreasing level for increasing numbers of vehicles available; thus, all other factors being equal, larger households are more likely to have an excess of vehicles over persons that smaller households, but both types of households are more likely to have fewer vehicles than persons;

Table 2.3 Estimation Results for the Final Multinomial Logit Vehicle Availability Model

	Vehicle Availability Level[1]				
Variable	0	1	2	3	4+
Persons per Household	--	--	0.1164	0.1164	0.2571
	--	--	2.1	2.1	1.7
Workers per Household	--	--	0.4915	1.474	2.139
	--	--	5.2	10.8	10.0
Household Density[2]	--	-0.0458	-0.1327	-0.1717	-0.2549
	--	2.9	5.4	4.4	3.0
Ln(Household Income)[3]	--	1.130	2.497	2.995	3.242
	--	8.7	13.9	12.7	7.6
Pedestrian Environment[4]	--	-0.4277	-0.6959	-0.6959	-0.6959
	--	1.6	2.4	2.4	2.4
Transit/Highway Access. Ratio[5]	--	-1.133	-2.054	-2.742	-2.742
	--	1.7	2.8	3.3	3.3
Persons less than Vehicles[6]	--	--	-2.870	-1.017	-0.5181
	--	--	8.8	5.3	1.1
Alternative-specific Constant	--	0.164	-3.761	-8.229	-12.87
	--	0.2	4.6	8.0	6.8

Notes:

[1] 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 coefficients which are statistically significant at the 95 percent confidence level. Coefficients (for the same variable and for two or more adjacent vehicle availability levels) which are the same were constrained to be equal in the estimation process.

[2] The units are households per acre.

[3] This variable is the natural logarithm of annual household income in thousands of 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 number of vehicles in the alternative, and 0 otherwise. For the 4+ vehicle alternative, the variable is 1 when number of persons is 3 or less.

Increasing household density levels have an increasingly negative effect on having more vehicles per household;

- The relationship between the logarithm of household income and vehicle availability is very strongly positive: the coefficients increase with vehicle availability level and, with the exception of the WPHH coefficient for the 4+ alternative, LNINC has the highest t-statistic in each utility function;

- The pedestrian environment variable is not strong, but does have the expected sign and the expected decreasing effect on higher levels of vehicle availability: Combination of the coefficient for higher levels (2 or more) was required to obtain consistent estimates; and

- The expected negative impact on multiple vehicle availability of increasing levels of the transit/highway accessibility ratio is evident in the results, but the change of this effect decreases for successively increasing numbers of vehicles: no difference could be estimated in the coefficients for the 3 and 4+ vehicle alternatives.

2.3.2 Ordered Response Logit Model

The results obtained for the multinomial (MNL) model discussed above provided the starting point for the estimation of the ordered response model, which is summarized in Table 2.4. Changes were required, however, for the following reasons:

- Coefficients in the ordered response model are more like incremental effects (changes for a given variable between having *N* and *N+1* vehicles), while those in the MNL model are total effects. These incremental effects tend to be harder to estimate than the total effects.

- The numbers of observations available to estimate each successive portion of the ordered response model decrease, and the ability to estimate statistically significant coefficients therefore also decreases for each successive portion.

- Randomness of the observed data sometimes results in a different variable, of a set of variables which is highly correlated, entering the ordered response submodels than a similar variable which entered the MNL model.

The combined effects of these factors are the following changes in the ordered response model:

- The persons per household and workers per household variables appear in the 0/1+ submodel, although they do not appear in the one-vehicle alternative of the MNL model: persons per household, on the other hand, does not appear in the 2/3+ submodel, consistent with the estimation of a single coefficient for the 2- and 3-vehicle alternatives in the MNL model;

Table 2.4 Estimation Results for the Initial Ordered Response Logit Vehicle Availability Model

	Vehicle Availability Decision[1]			
Variable	**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 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.

- Household density, which appears in all alternatives of the MNL model, is replaced by population density in the 0/1+ and 1/2+ submodels, and by employed person density in the remaining submodels;

- The pedestrian environment variable only appears in the 0/1+ and 1/2+ submodels, rather than in all submodels, as in all alternatives of the MNL model; and

- The transit/highway accessibility ratio variable does nor appear in the 3/4+ submodel, consistent with the estimation of a single coefficient for the 3- and 4+-vehicle alternatives in the MNL model.

The rho-squared values of the four ordered response submodels range from 0.302 to 0.732. Due to differences in the number of alternatives, compared with the MNL model, and numbers of observations from submodel to submodel, it is not possible to use these statistics to judge the relative goodness of fit of the complete ordered response and MNL models. Instead, it is necessary to test each against the observed household and zonal data to determine which one predicts the households' choice of vehicle availability level most accurately. The results of these tests are presented in the next section.

3.0 Model Validation and Final Model Selection

This section presents the results obtained when both of the of model structures presented in the previous section are applied to data for the DVRPC region. The sections which follow deal with the application of the models to the households which responded to DVRPC's 1987 travel survey – the same households used to estimate the models; to all households in the 1990 U.S. Census PUMS data – a one percent sample of all households in the region; and to zonal-level averages of all households in the region, as provided by the 1990 U.S. Census CTPP tabulations. The final section discusses the selection of the final model structure based on these model validation procedures.

3.1 Validation Results for Surveyed Households

The first test of the alternative models presented in Section 2.4 was performed as part of the model estimation process. In this test, variations in the MNL model and in each of the ORL submodels were examined by applying them to the estimation data set and comparing predicted and chosen vehicle availability levels. These comparisons were then summed over groups of households defined by variables such as household size and the area types of residence zones. These variables consist both of ones included in the models and others not included. The resulting summations of predicted and chosen alternatives can be compared to identify biases in the model specifications and the need for additional variables.

The procedure used to compare the alternative models' ability to replicate the choices of households in DVRPC's travel survey is the same as the disaggregate validation often used in the discrete choice model development process, with a single exception. Ideally, disaggregate validation is done with a different sample of households than that used for model estimation. However, since the DVRPC travel survey provides too few households to be divided into separate data sets for model estimation and validation, it was necessary to use the same data set for disaggregate model testing as for model estimation. This prevents the use of the survey to obtain an independent validation of the estimated models, but nonetheless provides a useful tool for testing the ability of alternative model specifications to replicate the observed behavior of the survey data at the level of the individual households.

The variables used in the household-level model testing process included all of the non-dummy variables included in either the MNL or ORL model structures, and in addition the area type of each household. The comparisons of predicted and chosen totals by alternative and range of the classification variable show that both the MNL model and each of the ORL submodels replicate the observed data very closely. As shown in Table 3.1, the percent root mean square errors per analysis cell (%RMSE) are all less than 10 percent. The area type variable, which is not included directly in the model specification, has the lowest %RMSE of any analysis variable, 2.3 percent. The other variables not in the model specification, employed person density and population density, have %RMSEs of 8.3 percent and 8.4 percent, respectively. These values are very similar to the error measure of 8.8 percent for the household density variable, which does appear in the model. The variable with the highest %RMSE is workers per household, at 9.7 percent. The relatively low values of all %RMSEs, plus the satisfactory results obtained for variables not in the model, indicate that the MNL model acceptably replicates the observed levels of household vehicle availability.

Table 3.2 displays the percent root mean square errors per analysis cell (%RMSE) for the four ORL submodels. For the 0 versus 1+ and 1 versus 2+ submodels, all %RMSEs are less than those of the MNL model. For the 2 versus 3+ submodel, just two of the nine variables have higher errors than the MNL model. The ORL errors are larger for six of the nine variables in the 3 versus 4+ submodel, apparently due to the small number of observations used to estimate this model. The 2/3+ submodel has a single %RSME greater than 10 percent; the 3/4+ submodel has only two exceeding this level. However, since the ORL submodels must be applied in succession to obtain predictions of the shares of households choosing to have multiple numbers of vehicles, the total errors for these predictions are expected to be the square root of the sum of the squared errors for each submodel used to obtain them. These cumulative total errors, however, are also less than 10 percent for all variables in the first two submodels. When the cumulative effects of the first three models are determined, the only variable with an error greater than 10 percent is workers per household. Only when the full set of ORL submodels are used to predict the 3 versus 4+ alternatives do the errors exceed 10 percent for four of the nine variables: workers per household, population and employed person densities, and the transit/highway accessibility ratio.

The tests of the ability of the final MNL and ORL models to predict the vehicle availability choices of the households in the estimation data set show that the observed variations for a number of variables are very well reflected in the model predictions. These variables include not only all those included in the models, but also a number of variables which do not appear in a specific model or submodel. In particular, predictions stratified by the area type variable have very low error levels even though this variable is not included explicitly in the utility function of the final MNL model or any of the ORL submodels. Unfortunately, these tests cannot be used to select between the alternative MNL and ORL structures because they must be applied individually to the ORL submodels rather than to the complete ORL model. For this reason, it was necessary to continue the model testing process using both model structures.

3.2 Validation Using PUMS Data

The second test of the validity of the final vehicle availability models was performed using the 15,497 households in the DVRPC study area which are included in the one

Table 3.1 Percent Root Mean Square Errors for the MNL Model

Variable	%RMSE
Persons per Household	7.5
Workers per Household	9.7
Population Density	8.4
Employed Person Density	8.3
Household Density	8.8
Household Income	4.7
Pedestrian Environment	3.1
Transit/Highway Accessibility	7.1
Area Type	2.3

Table 3.2 Percent Root Mean Square Errors for the ORL Submodels

	%RMSE by Submodel			
Variable	**0/1+**	**1/2+**	**2/3+**	**3/4+**
Persons per Household	3.2	1.8	5.1	6.0
Workers per Household	1.5	2.5	12.2	1.2
Population Density	2.1	3.4	8.4	15.0
Employed Person Density	2.2	3.8	7.6	8.7
Household Density	0.8	3.0	5.1	6.1
Household Income	3.5	4.1	2.2	7.5
Pedestrian Environment	1.0	1.4	2.5	3.5
Transit/Highway Accessibility	1.5	2.3	5.4	15.9
Area Type	1.4	1.2	3.5	3.0

percent 1990 U.S. Census Public Use Microsamples (PUMS) for Pennsylvania and New Jersey. For each of these households, the PUMS data provides the full set of data collected in the decennial census, with the exception of detailed locational information. This information, not provided to preserve confidentiality, is replaced by each household's public use microsample area (PUMA). There are 39 PUMAs in the DVRPC region, each with a population level in the range from 100,000 to 200,000. In order to use the PUMS data for model testing, it was necessary to attach variables averaged over PUMAs rather than DVRPC zones for each of the following variables: household, population, and employed person densities; pedestrian environment; and transit/highway accessibility.

The chosen and predicted results obtained for the expanded PUMS data set showed both the validity of the estimated models and the problems caused by using broad PUMA-wide averages of the variables not available in the individual household records. Tables 3.3 and 3.4 provide chosen and predicted vehicle availability shares for the two models by county and for the entire region, as well as the resulting average values. For both models, the predicted regionwide average is within seven percent of the chosen average, with the error for the ORL model being very slightly smaller than that for the MNL model. With the exception of two highly urban counties, Mercer and Philadelphia, the predicted county-specific averages all have errors of no more than five percent of observed averages. The errors in Mercer and Philadelphia counties are nine and 25 percent, respectively, suggesting that their more congested urban characteristics have impacts on vehicle availability which are not captured completely in the final estimated models.

The chosen and predicted percentages by vehicle availability level provide greater detail on the ability of both models to replicate the household behavior observed in the PUMS data. The use of highly aggregated and less dispersed PUMA-level values rather than zonal values results, as expected, in less variation in the predicted distribution of vehicle availability levels than in the observed distributions. Thus, the numbers of households predicted to select the zero and 4+ alternatives are both underestimated by 60 to 70 percent, while the predicted selection of the 1- and 2-vehicle alternatives are generally within +6 to +8 percent of the observed values, and the predicted selection of the 3-vehicle alternative is overestimated by 30 percent. Although these differences can be reduced by revising the alternative-specific constants of the models to make them consistent with the changes in data, this was not done for the PUMS data because the models will not be used with PUMS data in the DVRPC forecasting process.

The results obtained when the models are applied to the PUMS data set for the DVRPC region show that both models generally predict average vehicle availability values per household acceptably at the regional level, and also at the county level for all counties but Mercer and Philadelphia. These results suggest that county-specific constants may be warranted for Mercer and Philadelphia counties. The results for the ORL model are very marginally better than for the MNL model. The results also indicate the expected under-prediction of the 'tails' of the vehicle availability distribution, and corresponding over-prediction of the central portion, because it was necessary to use non-household data based on PUMAs rather than zones. These considerations led to the general conclusion that both model forms, possibly with additional county-specific constants, remain as candidates to be tested against the same zonal data which is normally used by DVRPC for travel forecasting, before any adjustments are made and before the final model form is selected.

Table 3.3 1990 PUMS Data versus MNL Vehicle Availability Model – Vehicle Availability Shares and Averages by County

Variable	Data Source	Bucks	Burlington	Camden	County Chester	Delaware	Gloucester	Mercer	Montgomery	Philadelphia	Total Region
% 0 Vehicles	PUMS	5%	5%	13%	5%	12%	7%	12%	7%	37%	17%
	Model	3%	3%	8%	3%	8%	3%	5%	4%	24%	11%
% 1 Vehicle	PUMS	28%	32%	35%	29%	35%	31%	34%	33%	42%	35%
	Model	27%	28%	38%	27%	38%	30%	34%	33%	49%	37%
% 2 Vehicles	PUMS	44%	42%	37%	46%	38%	44%	38%	43%	18%	34%
	Model	47%	47%	39%	48%	39%	45%	43%	43%	21%	36%
% 3 Vehicles	PUMS	16%	13%	10%	14%	12%	12%	11%	12%	3%	9%
	Model	18%	18%	12%	18%	12%	17%	15%	16%	5%	12%
% 4+ Vehicles	PUMS	6%	7%	5%	6%	3%	6%	5%	5%	0%	4%
	Model	4%	4%	3%	4%	3%	4%	3%	4%	0%	3%
Average Vehicle Availability (veh/hhld)	PUMS	1.91	1.89	1.60	1.88	1.60	1.81	1.64	1.77	0.89	1.48
	Model	1.95	1.93	1.65	1.93	1.65	1.90	1.79	1.82	1.11	1.59
% Error		1.7%	2.3%	3.2%	2.9%	3.3%	5.1%	9.3%	2.8%	24.5%	7.3%

Table 3.4 1990 PUMS Data versus ORL Vehicle Availability Model -- Vehicle Availability Shares and Averages by County

Variable	Data Source	Bucks	Burlington	Camden	County Chester	Delaware	Gloucester	Mercer	Montgomery	Philadelphia	Total Region
% 0 Vehicles	PUMS	5%	5%	13%	5%	12%	7%	12%	7%	37%	17%
	Model	3%	3%	8%	3%	8%	3%	5%	4%	24%	11%
% 1 Vehicle	PUMS	28%	32%	35%	29%	35%	31%	34%	33%	42%	35%
	Model	28%	29%	39%	28%	39%	31%	35%	34%	49%	38%
% 2 Vehicles	PUMS	44%	42%	37%	46%	38%	44%	38%	43%	18%	34%
	Model	47%	47%	39%	48%	39%	46%	42%	43%	20%	36%
% 3 Vehicles	PUMS	16%	13%	10%	14%	12%	12%	11%	12%	3%	9%
	Model	18%	17%	12%	17%	12%	16%	15%	15%	6%	12%
% 4+ Vehicles	PUMS	6%	7%	5%	6%	3%	6%	5%	5%	0%	4%
	Model	4%	4%	3%	3%	3%	4%	3%	3%	1%	3%
Average Vehicle Availability (veh/hhld)	PUMS	1.91	1.89	1.60	1.88	1.60	1.81	1.64	1.77	0.89	1.48
	Model	1.93	1.91	1.64	1.92	1.65	1.88	1.78	1.80	1.11	1.58
% Error		1.0%	1.1%	2.5%	2.1%	3.1%	3.9%	8.5%	1.7%	24.7%	6.8%

3.3 Validation Using Zonal Data

The final test performed to validate the vehicle availability models can be termed aggregate validation; in it, only DVRPC's zonal data files, supplemented with additional zonal Census data, were used to predict vehicle availability shares for each zone in the DVRPC region. As in the use of PUMS data, the aggregate data used to obtain these predictions initially resulted in the underprediction of the extremes of the vehicle availability distribution (in the case of zonal data, the 0-, 1-, 3-, and 4+-vehicle levels) and the overprediction of the central value, the 2-vehicle level. In addition, the average regional vehicle availability was overestimated by 19 percent for both models. Clearly, the use of all aggregate data results in larger deviations between chosen and predicted shares and averages than the use of partial aggregate data; as when the PUMS household data is supplemented with PUMA-level densities, accessibilities, and pedestrian variables.

To provide models which can be used to predict accurate vehicle availability levels using zonal data, it was necessary first of all to revise the alternative-specific constants of each model. These revisions were performed using an iterative process which ensures that the predicted and observed household shares are equal in each vehicle availability category, for each of the groups of counties discussed below. Revised values of these constants were initially determined for three groups of counties:

- Philadelphia County, for which the predicted versus observed error in both models was +36 percent prior to adjustments;
- Mercer County, for which the predicted versus observed error in both models was +23 percent prior to adjustments; and
- All other counties, for which the predicted versus observed errors in both models ranged from +11 to +16 percent prior to adjustments.

When the initial revised constants were used, the predicted regional shares for both models matched the observed shares within 0.7 percent, and predicted average vehicle availability levels equaled 1.44 to 1.45 vehicles per household, very close to the observed value of 1.44. Tables 3.5 and 3.6 provide chosen and predicted vehicle availability shares for the two initial adjusted models by county and for the entire region, as well as the resulting average values. Because DVRPC's zonal data set does not distinguish between the three- and four-or-more-vehicle households, the model results for these two counties were added together to create the three-or-more vehicle category provided in Tables 3.5 and 3.6. The absolute predicted county-specific averages all have errors of less than seven percent of the observed averages. Table 3.7 provides error measures of the full set of individual zonal predicted values. The average positive, negative, and total absolute errors are nearly equal for both models, 12 percent of the average value for the MNL model and 10 percent for the ORL model.

Table 3.5 DVRPC 1990 Zonal Data Versus Adjusted MNL Vehicle Availability Model – Vehicle Availability Shares and Averages by County

Variable	Data Source	Bucks	Burlington	Camden	Chester	Delaware	Gloucester	Mercer	Montgomery	Philadelphia	Total Region
		County									
% 0 Vehicles	Observed	5%	5%	13%	6%	12%	7%	13%	7%	38%	18%
	Model	5%	4%	14%	4%	13%	5%	13%	7%	38%	18%
% 1 Vehicle	Observed	29%	31%	35%	28%	37%	31%	34%	33%	41%	35%
	Model	28%	27%	40%	25%	39%	30%	34%	32%	41%	35%
% 2 Vehicles	Observed	46%	45%	37%	46%	38%	44%	38%	44%	18%	34%
	Model	46%	46%	35%	48%	36%	45%	38%	45%	18%	34%
% 3+ Vehicles	Observed	21%	19%	14%	20%	14%	18%	15%	17%	4%	13%
	Model	22%	22%	11%	23%	11%	20%	15%	17%	4%	13%
Average Vehicle Availability (veh/hhld)	Observed	1.86	1.83	1.55	1.84	1.57	1.78	1.58	1.74	0.88	1.44
	Model	1.89	1.91	1.45	1.94	1.48	1.84	1.58	1.75	0.88	1.45
% Error		1.6%	4.4%	-6.5%	5.4%	-5.7%	3.4%	0.0%	0.6%	0.0%	0.7%

Table 3.6 DVRPC 1990 Zonal Data Versus the Initial Adjusted ORL Vehicle Availability Model – Vehicle Availability Shares and Averages by County

	Data	County									Total
Variable	Source	Bucks	Burlington	Camden	Chester	Delaware	Gloucester	Mercer	Montgomery	Philadelphia	Region
% 0 Vehicles	Observed	5%	5%	13%	6%	12%	7%	13%	7%	38%	18%
	Model	5%	5%	14%	5%	13%	6%	13%	7%	38%	18%
% 1 Vehicle	Observed	29%	31%	35%	28%	37%	31%	34%	33%	41%	35%
	Model	28%	27%	40%	25%	39%	29%	34%	32%	41%	35%
% 2 Vehicles	Observed	46%	45%	37%	46%	38%	44%	38%	44%	18%	34%
	Model	45%	46%	35%	47%	36%	45%	38%	45%	18%	34%
% 3+ Vehicles	Observed	21%	19%	14%	20%	14%	18%	15%	17%	4%	13%
	Model	22%	22%	11%	23%	12%	20%	15%	17%	4%	13%
Average Vehicle Availability (veh/hhld)	Observed	1.86	1.83	1.55	1.84	1.57	1.78	1.58	1.74	0.88	1.44
	Model	1.88	1.90	1.46	1.93	1.49	1.83	1.58	1.75	0.88	1.44
% Error		1.0%	3.8%	-5.8%	5.1%	-4.9%	3.0%	0.0%	0.5%	0.0%	0.0%

Table 3.7 Error Measures: Alternative Models Versus 1990 DVRPC Zonal Data

Error Measure/Statistic	1990 Observed Zonal Data	MNL Model	Initial ORL Model
Average Positive Deviation(1)	---	0.165	0.148
Average Negative Deviation(1)	---	0.174	0.148
Average Absolute Deviation(1)	---	0.170	0.148
Percent Root Mean Square Error	---	15.32	13.17
Average Vehicle Availability(1)	1.444	1.457	1.446
Standard Deviation(1)	0.515	0.601	0.569

(1) The units of each of these statistics is vehicles per household.

3.4 Selection of Final Recommended Model

Based on the results summarized in Tables 3.5-3.7, the ordered response vehicle availability model was selected as the model to be recommended for implementation by DVRPC in its enhanced travel forecasting system. Although the two models provide nearly equal levels of accuracy in replicating both PUMS and zonal data for 1990, the ORL model is consistently more accurate by a slight amount. After adjustments are made to both models, the ORL model more accurately replicates average vehicle availability levels at the county and total regional levels, has the lower average deviations and the lower percentage root mean square errors. Also, the standard deviation of the zonal values provided by the ORL model is closer to the corresponding statistic for the observed data. Based on these results, ORL model was selected for use in predicting vehicle availability levels by zone for DVRPC's 2020 forecasts.

Subsequent to this selection, however, client review of the draft report for this task identified a bias in the initial adjusted ORL model. The estimates of average vehicle availability in the zones with the highest density levels were typically less than 50 percent of the observed Census data for these zones. This occurred in spite of the lack of such a bias in the disaggregate validation results for this model, as discussed in Section 3.1. When the 'chained' ORL model is adjusted to predict the correct shares by vehicle availability level, it no longer is unbiased with respect to zonal density variables. Instead, analysis of average values by county and either population or employed person density level revealed a definite trend from overprediction in low density zones to underprediction in high density zones.

Further adjustments of the selected ORL model significantly reduced the density-related biases observed in the initial version of the aggregate model. These adjustments involved, first of all, replacing the density variables in each of the ORL model's submodels with new variables, defined as follows:

- In the 0/1+ and 1/2+ submodels, the population density variable is replaced with the portion of population density which exceeds 12.5 persons per acre, if any; and

- In the 2/3+ and 3/4+ submodels, the employed person density variable is replaced with the portion of employed person density which exceeds 12.5 persons per acre, if any.

Following this change in variables, the ORL model was adjusted further by determining the coefficients of the new density variables which would most closely replicate observed average vehicle availability levels by county and by density level. At the same time, the initial county-level model constants discussed in the previous section were modified further to continue to ensure that vehicle availability shares are matched by county group. As this process was carried out, the three groups of counties used initially was expanded to four; Camden County was removed from the 'All Other' group and treated separately.

Table 3.8 displays the final adjusted constants and density coefficients of the selected aggregate ORL model. The adjustments should be used whenever the recommended model is applied at the zonal level, for both base-year and future-year model runs. The results obtained using this final version of the selected ORL model are provided in Tables 3.9, 3.10 and Figure 3.1 As in the initial adjusted ORL model, the overall shares by vehicle availability level match the observed data. At the county level, the observed average vehicle availability levels are matched more closely than in the initial model, except for very slight variations in Mercer and Philadelphia Counties. The largest variation, for Chester County, is just 2.6 percent. The overall error measures for this final model are provided in Table 3.10. For the final adjusted model, each of the deviation measures is at least 20 percent less than in the initial model, as shown in Table 3.7, and both the regional average and its standard deviation are within one percent of the observed values. Most importantly, the bias in high density zones is reduced significantly.

Table 3.8 Revised ORL Model Constants and Coefficients to Match DVRPC Aggregate 1990 Zonal Data

Constants	1+	2+	3+	4+
Philadelphia County	-2.768	-5.222	-3.751	-3.751
Mercer County	-2.742	-5.634	-3.944	-3.406
Camden County	-2.204	-5.142	-3.611	-3.073
Remaining DVRPC Counties	-2.168	-5.408	-3.830	-3.292
Coefficients of Population Density Greater than 12.5				
Philadelphia County	-0.0059	-0.0077	--	--
Mercer County	0	0	--	--
Camden County	-0.0025	-0.0032	--	--
Remaining DVRPC Counties	-0.0014	-0.0018	--	--
Coefficients of Employed Person Density Greater than 12.5				
Philadelphia County	--	--	-0.0026	-0.0046
Mercer County	--	--	0	0
Camden County	--	--	-0.0002	-0.0001
Remaining DVRPC Counties	--	--	0	0

Table 3.9 DVRPC 1990 Zonal Data Versus the Final Adjusted ORL Vehicle Availability Model – Vehicle Availability Shares and Averages by County

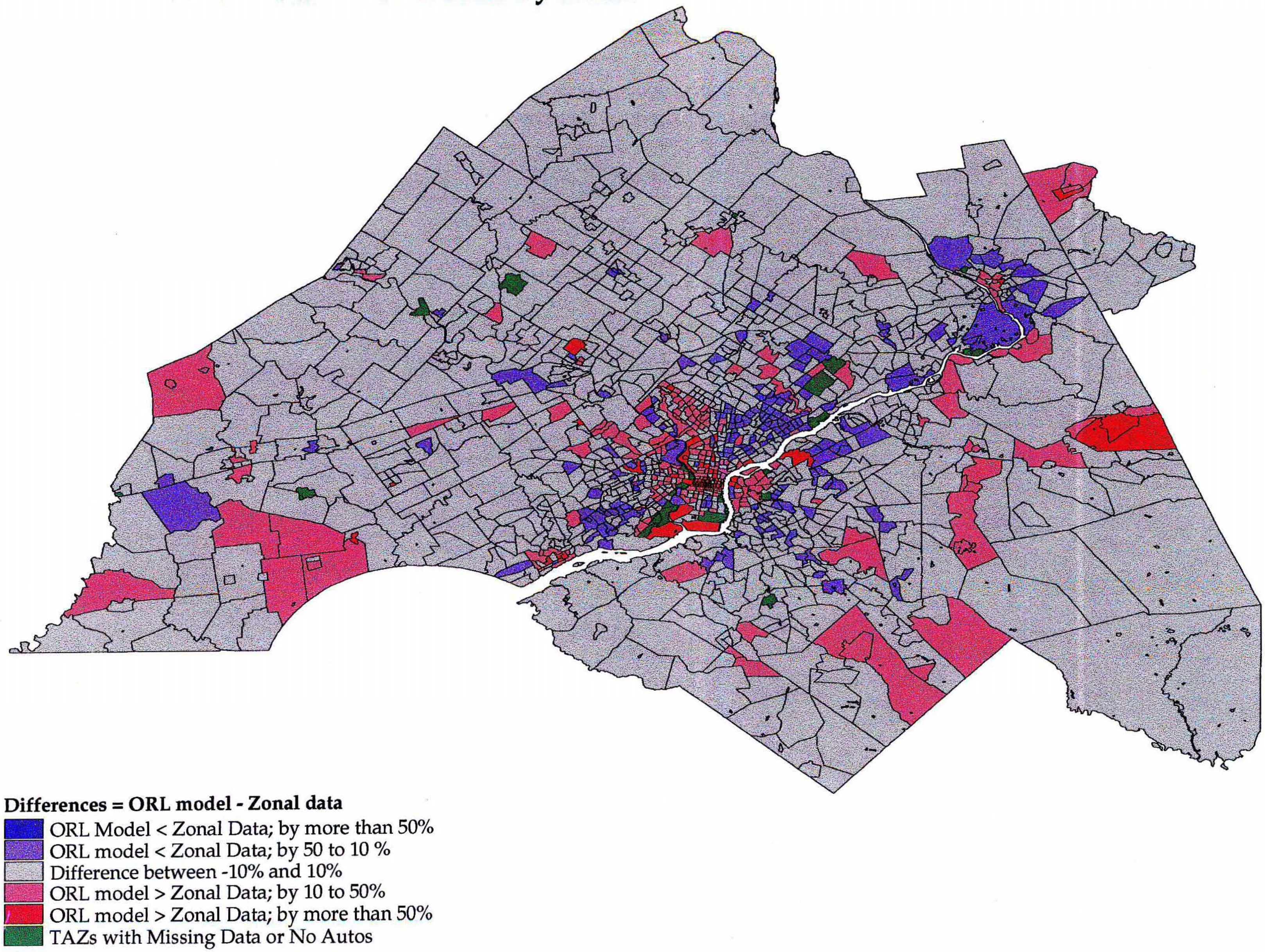
Variable	Data Source	County									Total Region
		Bucks	Burlington	Camden	Chester	Delaware	Gloucester	Mercer	Montgomery	Philadelphia	
% 0 Vehicles	Observed	5%	5%	13%	6%	12%	7%	13%	7%	38%	18%
	Model	6%	6%	13%	5%	11%	6%	13%	7%	38%	18%
% 1 Vehicle	Observed	29%	31%	35%	28%	37%	31%	34%	33%	41%	35%
	Model	29%	29%	35%	27%	38%	32%	34%	32%	41%	35%
% 2 Vehicles	Observed	46%	45%	37%	46%	38%	44%	38%	44%	18%	34%
	Model	44%	44%	37%	46%	38%	43%	38%	44%	18%	34%
% 3+ Vehicles	Observed	21%	19%	14%	20%	14%	18%	15%	17%	4%	13%
	Model	21%	21%	14%	22%	12%	18%	15%	16%	4%	13%
Average Vehicle Availability (veh/hhld)	Observed	1.86	1.83	1.55	1.84	1.57	1.78	1.58	1.74	0.88	1.44
	Model	1.84	1.85	1.55	1.89	1.54	1.78	1.58	1.73	0.88	1.44
% Error		0.9%	1.4%	0.0%	2.6%	-1.8%	0.0%	0.1%	-0.3%	-0.1%	0.0%

Table 3.10 Error Measures: Final ORL Model Versus 1990 DVRPC Zonal Data

Error Measure/Statistic	1990 Observed Zonal Data	Final ORL Model
Average Positive Deviation(1)	--	0.113
Average Negative Deviation(1)	--	0.119
Average Absolute Deviation(1)	--	0.116
Percent Root Mean Square Error	--	10.52
Average Vehicle Availability(1)	1.444	1.442
Standard Deviation(1)	0.515	0.518

(1) The units of each of these statistics is vehicles per household.

Figure 3.1 Final Adjusted ORL Vehicle Availability Model versus DVRPC 1990 Zonal Data – Deviations by Zone



4.0 Prediction Results for 2020

This section describes the application of the selected vehicle availability model to predict future zonal fractions of households having 0, 1, 2, and 3+ vehicles in 2020. Section 4.1 discusses the prediction process, including the data used for prediction and the implementation of the model. Section 4.2 focuses on the method used to project household incomes, the single variable required for prediction which was not predicted by DVRPC prior to the start of this project. Finally, Section 4.3 compares the predictions of the selected model with DVRPC's existing estimates of vehicle availability fractions by zone.

4.1 The Prediction Process

The data required to forecast 2020 zonal vehicle availability for the DVRPC region and its sources are the following:

- Number of households, employed persons, total population, and employment by zone as provided by DVRPC based on the forecasts[1].
- The pedestrian environment variables developed in this project for 1990 were assumed to remain unchanged in the future;
- The relative zone-to-zone transit and highway travel times obtained from base year networks were assumed to remain unchanged in 2020; and
- Household income levels by zone for 2020 were projected from the 1990 values observed in the Census data using the procedure discussed in the next section.

Following the assembly of 2020 zonal data from the sources identified above, each of the independent variables required by the model (as identified in Table 2.4) could be computed for each zone. With the exception of the pedestrian environment variable, each model variable has different zonal values for 2020 than for 1990. This is also true for the transit/highway accessibility variable even though no changes were assumed for the transit and highway travel times, because the distribution of employment levels by zones is predicted to change. Once all zonal variables were computed in the form required by the model, the procedure used to test the ORL model against the observed 1990 zonal data could be used to obtain predictions of household shares by vehicle availability level for

[1] DVRPC, *2020 Zonal Population and Employment Forecasts*, Direction 2020 Report # 25, April 1995.

2020. As a later part of the project, this same procedure will be integrated by Cambridge Systematics into DVRPC's updated travel forecasting process.

4.2 Projecting Average Household Incomes by Zone

This section presents the projection method developed to obtain zonal estimates of household income in 2020. Although household income is an essential variable in the mode choice and vehicle availability models developed in Tasks 3 and 10 of this project, it is not available at the zonal level in the existing DVRPC future year projections of demographic characteristics. This section describes a simplified means of projecting average zonal household income levels to 2020, DVRPC's current forecast year. The projection process is based on two zonal variables which can be computed from demographic data which are projected by DVRPC – persons per household and workers per household – and on the base year zonal average household incomes. These averages implicitly reflect the wide range of zonal characteristics related to income levels which can differ from zone to zone. By using these base year values to project future year values, we assume that the relative levels of these underlying zonal characteristics will not change significantly over the forecasting period. The base year of 1989 is selected because the 1990 Census data provides information on household incomes in that year for all zones in the DVRPC study area. The projection method is presented here, as well as the means by which it has been made operational.

4.2.1 The Income Projection Method

The method used in this project to project future year household income levels by zone can be stated mathematically as follows:

$$Y(f,i) = Gc[pphh(f,i), wphh(f,i)] * F1(i) * F2(f,r)$$

where:

Y(f,i) is the projected average household income in zone i for forecast year f (for base year, f=89).

pphh(f,i) is the average household size for zone i, year f (for base year, f=90).

wphh(f,i) is the average number of workers per household for zone i, year f (for base year, f=90).

Gc[p,w] is a county-specific function or look-up table, based on Census data as discussed below, of average household incomes by household size and workers per household for the county in which zone i is located.

F1(i) is a zonal factor defined as Y(89,i)/Gc[pphh(89,i), wphh(89,i)].

F2(f,r) is a regional factor reflecting the projected average change in real income between the base year (1989) and the future forecast year.

The key to the proposed method is Gc, a set of look-up tables which reflect the variations for each county in household incomes as the household sizes and numbers of workers per household change. Counties were chosen for these tables because sufficient numbers of observations for reliable averages are available at this level. Furthermore, since the projection approach is calibrated to match base-year zonal data, no finer breakdown – by county planning areas, for example – would significantly affect the future-year income projections provided by the procedure. The county-level tables were obtained from the 1990 sample of 15,497 households included in the one percent PUMS data for the DVRPC region. After cross-classifying these records by county, household size, and number of workers, the average household income was computed for each cell. The results are shown in Table 4.1. Interpolation in two dimensions can be used to apply the function with non-integer zonal average values of the independent variables. This interpolation operates on the averages obtained for the integer household-specific values. For future projections, changes in the zonal average independent variables will result in changes in the values obtained from the Gc function for each zone.

Two adjustment factors, F1 and F2, are also included in the projection method. F1 accounts for differences between each specific zone and the county level averages used in the Gc functions. It ensures that when the method is applied to the base year, the observed zonal averages will be predicted. F2 is provided to account for changes, at the regional level, in real incomes over the forecasting period. For use in their current mode choice model, DVRPC assumes that real incomes, on average, are not changing over time. Although our forecasts will be made consistent with this assumption by setting F2 equal to 1.0, we will provide a projection procedure in which the analyst must make an explicit decision concerning trends in average real incomes over time.

4.2.2 Implementing the Method

To provide for ease of transfer of the projection method to DVRPC, it has been implemented as a Microsoft Access database procedure. This procedure has the following inputs for each zone:

- 1989 average household income, Y(89,i);
- 1990 total population, TPOP(90,i);
- 1990 total employed workers (labor force), TLF(90,i);
- 1990 total households, THH(90,i);
- 2020 total population, TPOP(20,i);
- 2020 total employed workers, TLF(20,i);

Table 4.1 Average Income by Persons Per Household and Workers Per Household by County

Workers per Household	Persons per Household 1	2	3	4+
Bucks County				
0	13,982	28,518	26,351	13,730
1	32,374	46,443	48,011	46,283
2		60,047	61,126	58,157
3+			69,940	67,293
Burlington County				
0	14,506	30,222	25,587	23,027
1	31,866	46,421	43,034	48,476
2		55,083	59,454	58,335
3+			77,372	67,225
Camden County				
0	12,593	22,581	19,207	20,666
1	31,899	39,518	41,965	46,006
2		56,532	55,209	59,185
3+			71,629	71,617
*Chester County**				
0	14,369	30,428	36,113	12,568
1	30,584	50,593	46,213	64,061
2		58,615	55,434	62,150
3+			73,193	64,250
*Delaware County**				
0	17,216	33,277	30,301	11,979
1	31,269	50,884	50,583	46,384
2		66,164	71,246	69,322
3+			84,489	83,295
Gloucester County				
0	13,393	22,572	15,598	17,019
1	35,550	36,601	43,639	45,397
2		57,843	49,421	52,260
3+			68,894	59,199

* Contains parts of Montgomery County.

Table 4.1 Average Income by Persons Per Household and Workers Per Household by County (continued)

Workers per Household	Persons per Household 1	2	3	4+
Mercer County				
0	13,285	27,233	39,510	22,924
1	31,418	44,153	40,239	56,809
2		59,111	69,672	70,846
3+			81,829	74,609
Montgomery County				
0	14,605	31,784	31,744	34,717
1	34,294	50,463	49,936	68,829
2		64,107	65,452	62,701
3+			81,242	80,062
Philadelphia County				
0	10,845	18,396	15,440	15,566
1	26,956	29,435	31,355	32,473
2		50,083	49,929	46,991
3+			57,362	54,649

Source: 1990 U.S. Census Public Use Microdata Sample.

- 2020 total households, THH(20,i); and
- County indicator, C(i).

The Gc look-up tables for each county, as in Table 4.1, are also required inputs. The final input variable is F2(20, region), which has been assigned the value of 1.0 for this test of the recommended model.

The following calculations are performed for each zone in the DVRPC study area:

1. Per household demographic data for both base and future years:

   PPHH(y,i) = TPOP(y,i)/THH(y,i)

   IP(y,i) = Integer(PPHH(y,i))

   WPHH(y,i) = TLF(y,i)/THH(y,i)

   IW(y,i) = Integer(WPHH(y,i))

2. Interpolated value of Gc function for base and future years:

   START(y,i) = Gc(IP(y,i), IW(y,i))

   INCP(y,i) = [Gc(IP(y,i) + 1, IW(y,i)) - START(y,i)] if IP(y,i) < 4; 0 otherwise

   DP(y,i) = [PPHH(y,i) - IP(y,i)] * INCP(y,i)

   INCW(y,i) = [Gc(IP(y,i), IW(y,i) + 1) - START(y,i)] if IW(y,i) < IP(y,i) < 3; 0 otherwise

   DW(y,i) = [WPHH(y,i) – IW(y,i)] * INCW(y,i)

   GVAL(y,i) = START(y,i) + INCP(y,i) + INCW(y,i)

3. Value of F1 function:

   F1(i) = Y(89,i)/GVAL(89,i)

4. Projected future year zonal income:

   Y(20,i) = GVAL(20,i) * F1(i) * F2(20, region)

The Access procedure performs these calculations for each zone and retains the new zonal variable for use in computing vehicle availability fractions based on the recommended model.

When this projection procedure is used to obtain estimates of 2020 average household incomes by zone, the regional average changes from the 1989 value of $44,530 to a 2020 value of $46,180. This increase of 3.7 percent is due primarily to the overall increase in labor force participation rate of 2.2 percent – from 1.32 to 1.35 workers per household. The predicted decrease of 5.1 percent in the average household size – from 2.74 to 2.60 persons per household – tends to reduce the predicted average income levels, but this effect is much smaller than the increase due to more workers per household, because average incomes change much more significantly with the workers per household variable than with persons per household. Both county-level and regional average values of persons per household, workers per household, and household incomes are provided in Table 4.2.

4.3 Comparison of the Recommended Model and DVRPC's Prior Predictions

Table 4.3 provides a comparison of the ORL model predictions and DVRPC's previous forecasts for the year 2020. The table displays household fractions by vehicle availability and averages by county and for the region as a whole. Overall, the ORL and previous forecasts differ by just -4.4 percent. The prior predictions reflect an increase in the average of ten percent from 1990 to 2020; for the ORL model, the increase is reduced to six percent. The slight downward bias in the ORL values is reflected in each of the county averages, all of which are less than or equal to the previous DVRPC forecasts. These differences range from minus seven to zero percent. As shown in Table 4.4, the zonal values for which the ORL model provides a lower estimate have an average deviation of just over ten percent; for those with a higher ORL estimate, the average deviation is under eight percent. For all zones, the average absolute deviation is eight percent, compared with an average of ten percent for the 1990 predictions from the final model. The percentage root mean square deviation is just 11.7 percent, nearly as low as that for the 1990 application. The standard deviation of the two sets of estimates are also nearly equal, 0.52 for the prior predictions, and 0.49 for the ORL model results. Finally, Figure 4.1 provides information on the geographical distribution of the differences in the two sets of predicted values by zone.

Table 4.2 DVRPC County-Level Average Household Characteristics for 1989/1990 and 2020

County	Persons per Household[1] 1990	Persons per Household[1] 2020	Workers per Household[1] 1990	Workers per Household[1] 2020	Household Income 1989[2]	Household Income 2020
Bucks	2.84	2.65	1.50	1.53	$51,610	$53,910
Burlington	2.89	2.70	1.53	1.54	50,330	50,680
Camden	2.81	2.68	1.34	1.35	43,840	44,430
Chester	2.82	2.66	1.49	1.52	57,220	57,030
Delaware	2.72	2.58	1.32	1.33	46,100	46,900
Gloucester	2.92	2.74	1.43	4.46	44,020	45,340
Mercer	2.77	2.64	1.43	1.44	51,930	54,070
Montgomery	2.64	2.52	1.41	1.44	56,770	56,750
Philadelphia	2.62	2.48	1.09	1.09	31,210	31,640
Total Region	**2.74**	**2.60**	**1.32**	**1.35**	**$44,530**	**$46,180**

[1] Source: DVRPC zonal data based on 1990 U.S. Census.
[2] Source: Cambridge Systematics, Inc.

Table 4.3 DVRPC 2020 Zonal Data Versus Final Adjusted ORL Vehicle Availability Model – Vehicle Availability Shares and Averages by County

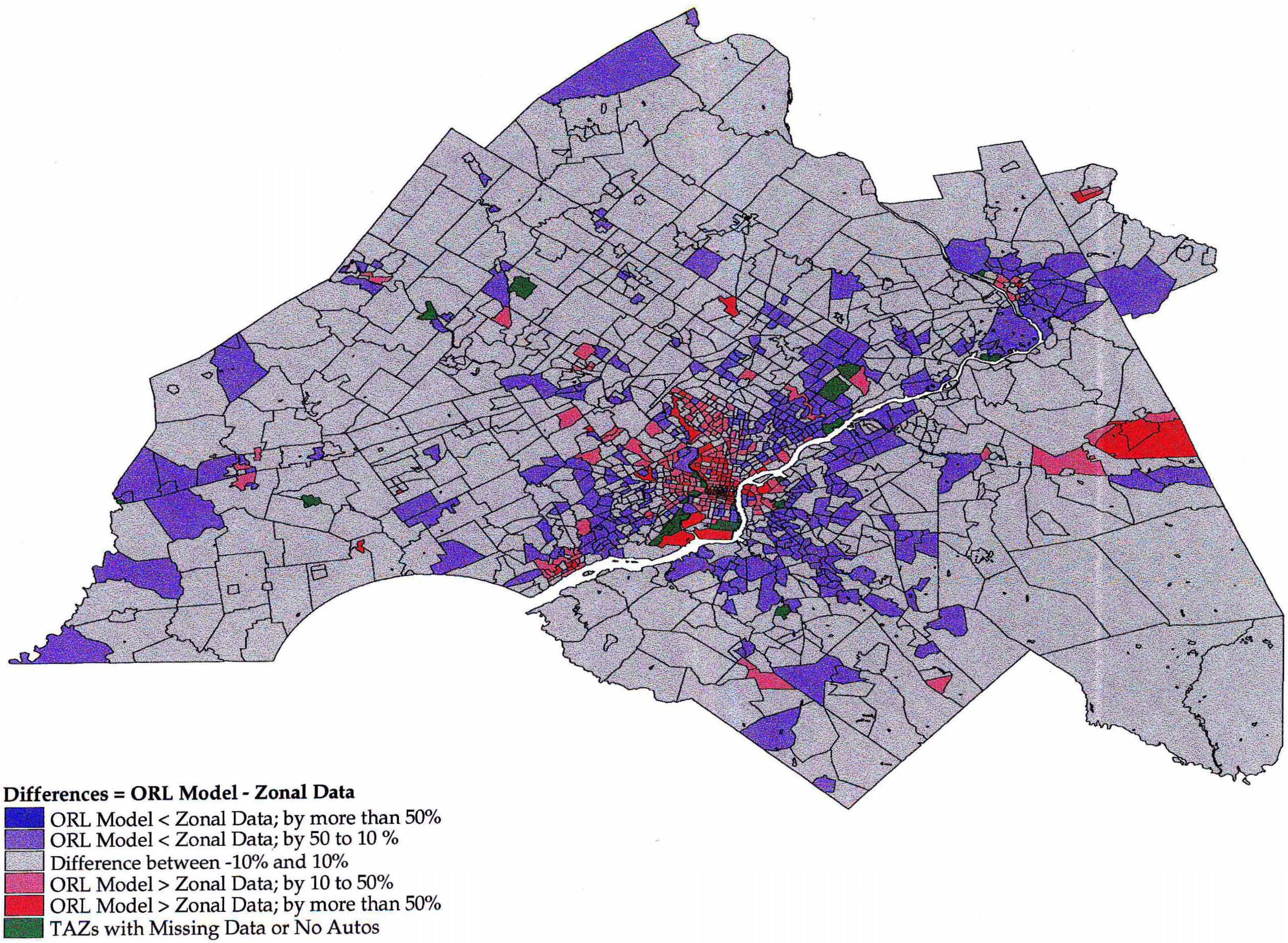
	Data	County									Total
Variable	Source	Bucks	Burlington	Camden	Chester	Delaware	Gloucester	Mercer	Montgomery	Philadelphia	Region
% 0 Vehicles	Observed	4%	4%	11%	5%	10%	5%	10%	6%	37%	15%
	Model	5%	5%	12%	5%	10%	6%	11%	6%	36%	15%
% 1 Vehicle	Observed	23%	25%	30%	22%	33%	24%	29%	29%	39%	30%
	Model	27%	28%	34%	26%	37%	30%	33%	33%	42%	34%
% 2 Vehicles	Observed	49%	47%	41%	49%	41%	50%	42%	47%	20%	39%
	Model	47%	46%	39%	48%	41%	45%	41%	46%	19%	37%
% 3+ Vehicles	Observed	24%	24%	19%	24%	16%	21%	19%	19%	4%	16%
	Model	21%	20%	15%	21%	12%	19%	15%	16%	4%	14%
Average Vehicle Availability (veh/hhld)	Observed	1.97	1.95	1.71	1.98	1.65	1.91	1.75	1.82	0.92	1.59
	Model	1.87	1.85	1.60	1.91	1.57	1.80	1.62	1.74	0.91	1.52
% Error		-4.9%	-5.1%	-6.8%	3.7%	-4.8%	-5.6%	-7.3%	-4.3%	-0.6%	-4.4%

Table 4.4 Variations in Final ORL Model Results Versus Current 2020 DVRPC Vehicle Availability Predictions

Error Measure/Statistic	2020 DVRPC Predictions	Final ORL Model
Average Positive Deviation(1)	---	0.159
Average Negative Deviation(1)	---	0.119
Average Absolute Deviation(1)	---	0.146
Percent Root Mean Square Error	---	11.70
Average Vehicle Availability(1)	1.587	1.519
Standard Deviation(1)	0.536	0.494

(1) The units of each of these statistics is vehicles per household.

Figure 4.1 Final Adjusted ORL Vehicle Availability Model versus Previous 2020 Predictions – Deviations by Zone



5.0 Summary of Recommendations

This report presents a method for estimating future shares of households by zone having up to five levels of vehicle availability: 0, 1, 2, 3, 4, or more vehicles. The method is based on a model estimated using data from 1,993 households included in DVRPC's 1987 travel survey. This model has the following advantages over DVRPC's current method of predicting future vehicle availability shares:

- Because the model is estimated using disaggregate household data, it reflects the full range of variability of households in the DVRPC region and thus can be expected to provide the best estimate of the relationship of household characteristics such as income level and number of workers to these households' vehicle availability decisions;

- The model also considers locational factors affecting vehicle availability as reflected in population and employed person density measures at the zonal level;

- The model includes the effect of the environment available for pedestrian and bicycle travel as an alternative to auto travel; and

- The model reflects the relative accessibility of the households' zones to employment by transit versus highway.

Comparisons of the recommended model with 1990 individual household data (from the U.S. Census Public Use Microdata Set) and with zonal Census data show that its predictions replicate the observed data very well. Comparisons with DVRPC's prior predictions for 2020 show that the two alternative procedures are generally consistent when used to obtain future forecasts.

It is recommended that the existing DVRPC modeling process be revised to add the application of the recommended vehicle availability model and household income projection process prior to the trip generation step. In addition to the zonal demographic and other variables also used in trip generation, the required inputs will include zonal pedestrian environment and accessibility variables. The pedestrian environment variables for future years should be obtained by revising the base year values as necessary to reflect expected changes in sidewalk availability, ease of street crossings, and building setbacks as new developments and redevelopments are projected. The future accessibility variables must be based on the future projections of employment by zone and the highway and transit travel times based on the projected unloaded highway and transit networks for the future scenario. Using these inputs, the full set of steps which will be required to estimate future vehicle availability shares by zone will be the following:

- Project future zonal household income levels using the procedure presented in Section 4.2;

- Compute future zonal highway and transit accessibilities using future employment by zone and zone-to-zone travel time skims from the future year unloaded highway and transit networks; and

- Compute zonal numbers of households having 0, 1, 2, and 3+ (or 3 and 4+) vehicles using the recommended ORL model, presented in Table 2.4 with the variable, coefficient, and constant revisions shown in Table 3.8; and provide the results as part of a zonal data file for input to the revised trip generation modeling process.

These new vehicle availability prediction steps have been implemented as part of the model development process using Microsoft Access procedures.