San Francisco Travel Demand Forecasting Model Development

Vehicle Availability Model

Final Report

prepared for
San Francisco County Transportation Authority

prepared by
Cambridge Systematics, Inc.

Updated by:
San Francisco County Transportation Authority

October 1, 2002

Table of Contents

Introduction 1
Overview 1
Data Sources 4
  Estimation Data 4
  Validation Data 5
  Candidate Variables 7
  Vehicle Availability Model Estimation Results 8
Model Application 12
APPENDIX A. Final Calibrated Vehicle Availability Model 13
APPENDIX B. Files for Running the Vehicle Availability Models for the Base Year 14

Introduction

This report documents the estimation of models to predict the vehicles available for use in making local travel. Like the rest of the models for the SFCTA project, these models were estimated using the 1990 MTC Bay Area Travel Survey (BATS). Given the location of the household, the characteristics of the household members, and the primary work place location of each of its workers, the model will estimate the probabilities of having none, one, or two or more vehicles available. The location of this model in the San Francisco model stream is shown in Figure 1

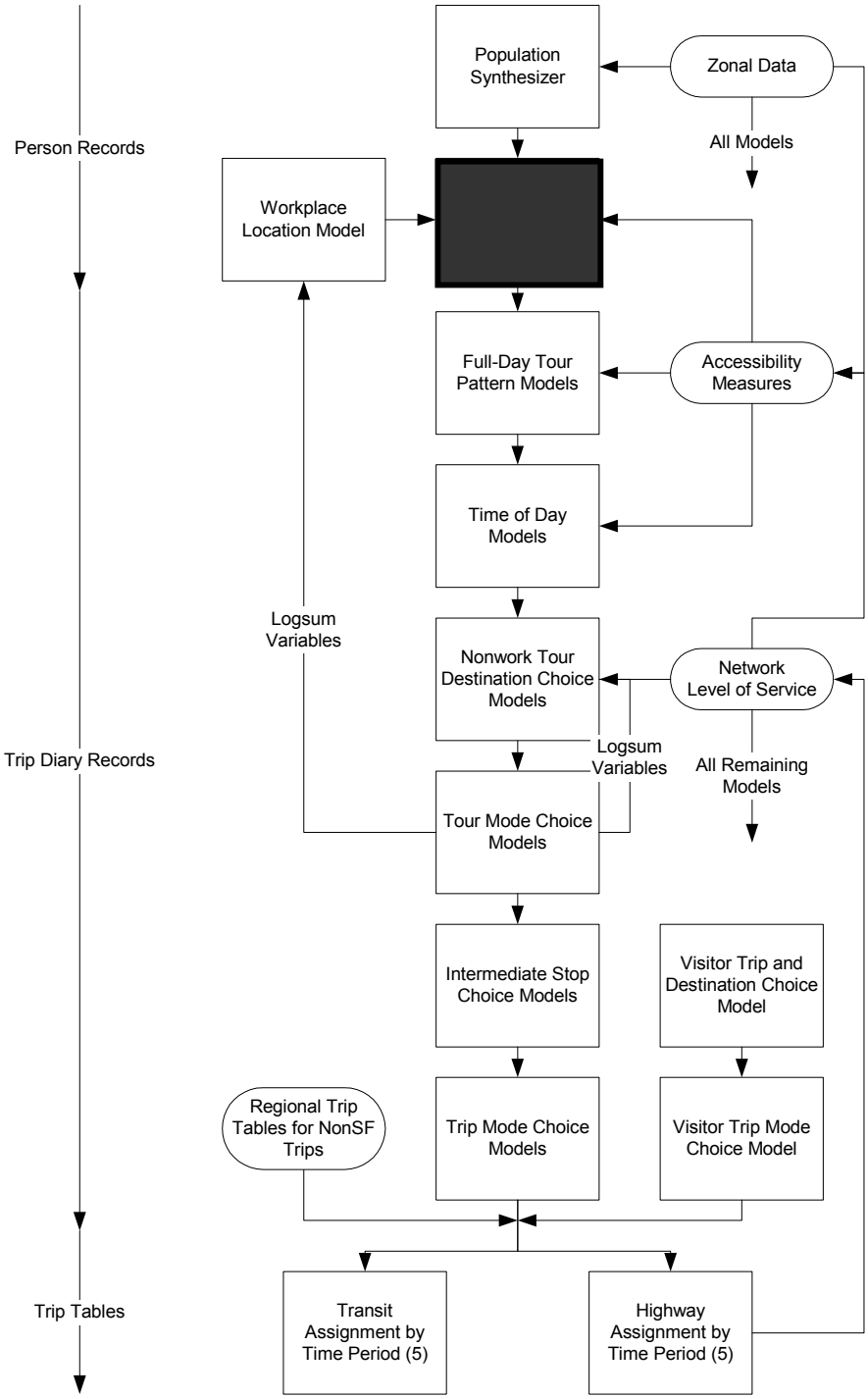
The term "vehicles available" is used to clarify two points. First, the model is dependent on whether a vehicle is "available" to a traveler rather than whether a vehicle is "owned" by a traveler. Second, the model is dependent on whether a "vehicle" is available to a traveler rather than whether an "auto" is available to a traveler. These refinements of travel behavior survey data collection have allowed respondents to be more specific about whether a vehicle was available to them rather than whether they own a vehicle and whether that vehicle was an auto, truck, sport utility vehicle, other motorized four-wheeled vehicle.

Overview

The vehicle availability model was tested in three forms, as a multinomial logit model, an ordered response logit model, or a nested logit model. The three types are shown in Figure 2. These alternative structures correspond to the following assumptions about the underlying behavioral choice mechanism for the households in San Francisco:

- **Multinomial Logit:** Each household makes a one-time choice of the number of vehicles to have available.
- **Ordered Response Logit:** Each household decides first whether to have zero or one-or-more vehicles. Next, if the one-or-more alternative is selected, the household decides to have one or two-or-more vehicles. This process continues until all vehicle availability levels have been considered by the households deciding to have the highest vehicle availability level, three or more vehicles.
- **Nested Logit:** In making its vehicle availability choice, each household considers all alternative levels simultaneously, but groups some alternatives as being more similar than others. Thus, the tradeoff between the two highest levels (two vehicles or three or more vehicles) may be more 'direct' than that between having zero or the highest availability level. In this model form, the degrees of similarity for the various subsets of availability levels can be estimated using the observed household choice data; because of this, the nested logit structure can be termed a generalization of the ordered response logit structure.

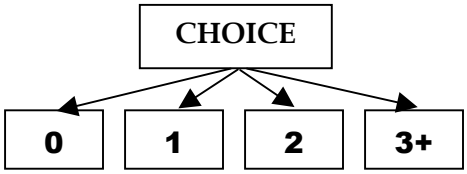
Figure 1. San Francisco Model System



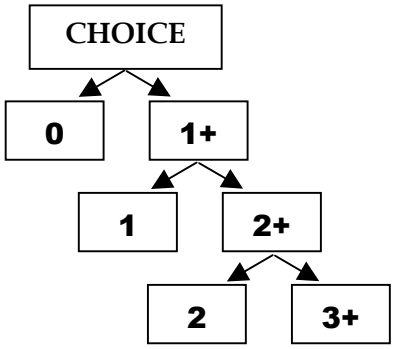
The choice between these three model forms was made following the statistical estimation of preliminary models having each form; based on the ability to obtain reasonable coefficients, and on the goodness of fit, of the models having each structure. The result of this comparison was to select the multinomial logit form.

Figure 2. Potential Model Structures

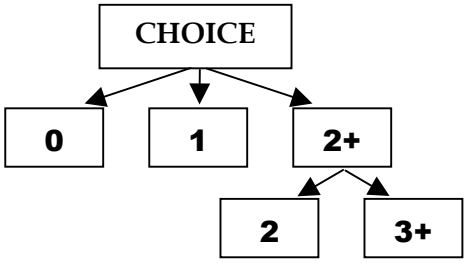
Multinomial Logit (MNL)



Ordered Response Logit (OL)



Nested Logit (NL)



Data Sources

Estimation Data

The number of alternatives to be included in the vehicle availability models was dependent on the numbers of households choosing each reported vehicle availability level. Table 1 provides a tabulation of households in San Francisco by the number of vehicles available, based on the estimation data set. In addition to total surveyed households in San Francisco County, three types of households are included in the table, each with its own totals:

- Households with one or more workers and one or more work tours (A home-based work tour includes the entire chain of trips made between leaving home and arriving back at home);
- Households with one or more workers and no work tours; and
- Households with no workers.

The number of vehicles is defined as automobiles plus trucks; also available in the survey data are the numbers of motorcycles, mopeds and bicycles owned by the household, but these will not be included in the number of vehicles available for household travel.

The model was limited to four alternatives because of the relatively small number of households with four or more vehicles available (1.8%). These four alternatives are:

- Households with no vehicles available;
- Households with one vehicle available;
- Households with two vehicles available; and
- Household with three or more vehicles available.

The average number of vehicles in the fourth alternative (households with three or more vehicles available) is 3.36. This average is used to calculate the number of vehicles per household for this category for summary purposes.

Table 1. Vehicles per Household

Number of Vehicles per Household	1+ workers; 1+ work tours	1+ workers; no work tours	Non working Households	Total Households	Percent of Total
0	158	40	127	325	*25.5%*
1	372	75	99	546	*42.9%*
2	252	35	21	308	*24.2%*
3	60	7	4	71	*5.6%*
4	13	3	0	16	*1.3%*
5	3	1	0	4	*0.3%*
6	1	0	1	2	*0.2%*
7	1	0	0	1	*0.1%*
Total	860	161	252	1273	*100%*

Source: 1990 MTC Survey

Validation Data

Table 1 shows the number of households with vehicles available stratified by the number of workers and work tours in the household. The largest number of households (42.9%) have only one vehicle and the average number of vehicles for all households is 1.16. Table 2 presents the number of vehicles per household from the 1990 Census for comparison and there are also 42% of the households in San Francisco who have one vehicle available. The number of households with no vehicle available in the 1990 MTC Survey (25.5%) is lower than the households with no vehicle available in the 1990 Census (30%). This potentially represents a slight bias in the survey towards households with vehicles available. Table 2 shows the number of households stratified by the number of workers, vehicles available and super-district based on the 1990 Census.

Table 1 is used as the primary input data for model estimation; Table 2 is used primarily for model validation.

Table 2. Households by Superdistrict and Auto Ownership

Total Households

	Number of Households					Percent of Total				
Super-district	0	1	2+	Total HH	Total Veh	0	1	2+	Total HH	Total Veh
1	35,825	17,300	5,411	58,537	29,653	61%	30%	9%	100%	0.51
2	26,502	45,917	23,805	96,226	100,264	28%	48%	25%	100%	1.04
3	23,371	44,066	36,385	103,825	127,133	23%	42%	35%	100%	1.22
4	7,168	20,063	20,171	47,406	66,113	15%	42%	43%	100%	1.39
total SF	92,866	127,346	85,772	305,994	323,163	30%	42%	28%	100%	1.06

Non-Working Households

	Number of Households					Percent of Total				
Super-district	0	1	2+	Total HH	Total Veh	0	1	2+	Total HH	Total Veh
1	16,906	3,208	444	20,559	4,222	82%	16%	2%	100%	0.21
2	12,076	8,792	1,727	22,597	12,735	53%	39%	8%	100%	0.56
3	10,900	10,157	3,135	24,195	17,314	45%	42%	13%	100%	0.72
4	4,211	6,895	1,937	13,047	11,317	32%	53%	15%	100%	0.87
total SF	44,093	29,052	7,243	80,398	45,588	55%	36%	9%	100%	0.57

Single Worker Households

	Number of Households					Percent of Total				
Super-district	0	1	2+	Total HH	Total Veh	0	1	2+	Total HH	Total Veh
1	14,105	9,254	1,598	24,958	12,902	57%	37%	6%	100%	0.52
2	10,712	25,383	6,673	42,770	40,617	25%	59%	16%	100%	0.95
3	8,042	20,131	8,960	37,136	40,587	22%	54%	24%	100%	1.09
4	1,912	8,392	5,207	15,515	20,280	12%	54%	34%	100%	1.31
total SF	34,771	63,160	22,438	120,379	114,386	29%	52%	19%	100%	0.95

Multi Worker Households

	Number of Households					Percent of Total				
Super-district	0	1	2+	Total HH	Total Veh	0	1	2+	Total HH	Total Veh
1	4,814	4,838	3,369	13,022	12,529	37%	37%	26%	100%	0.96
2	3,714	11,742	15,405	30,863	46,912	12%	38%	50%	100%	1.52
3	4,429	13,778	24,290	42,500	69,232	10%	32%	57%	100%	1.63
4	1,045	4,776	13,027	18,852	34,517	6%	25%	69%	100%	1.83
total SF	14,002	35,134	56,091	105,237	163,190	13%	33%	53%	100%	1.55

Source: 1990 Census CTPP

Candidate Variables

The following variables were considered in the estimation data set for the vehicle availability model. Each variable was tested for inclusion in the final model based on whether or not each complies with expectations, how well each explains the observed household choice behavior, and based on statistical considerations.

Household Variables

- Annual income and income transformed to the logarithm of income, to reflect the decreasing importance of income increases for increasing income levels, measuring the resources available to the household for the purchase of vehicles
- Number of persons by age group in the household
- Number of adults in the household
- Number of workers (full time and part time)
- Number of retirees
- Number of children
- Number of licensed drivers (or individuals old enough to be licensed)
- Dwelling unit type (single family or multi family)
- Household ownership (own or rent)

Locational Variables for the Residence Zone

- Residential density, as a potential measure of congestion and the competition for residential parking in a zone
- Employment density, as a measure of the likelihood to be able to walk to both work and non-work destinations rather than drive
- Employed resident density, as a measure of the numbers of workers living in a zone
- Pedestrian environmental factor, based on assessments of the degree to which walk and bike trips are facilitated by zonal characteristics such as building setbacks, sidewalk coverage, and grades, another measure of the substitutability of non-motorized for motorized travel
- Area type, capturing differences in vehicle availability patterns for CBD versus urban versus suburban zones due to factors not reflected in the zonal measures identified above

Accessibility Variables

- Auto and transit travel time (or distance) from the residence zone to the work zone for each worker, based on the work location choice model

- Transit/auto accessibility ratio of transit to auto level of service, measures the likelihood of travel by transit rather than auto to work, reducing the need for higher levels of vehicle availability.
- Auto and transit accessibilities for non-work destinations, in the form of the fractions of regional employment (potentially by employment type) which can be reached within a stated travel time by each mode, possibly combined into ratios of transit to auto accessibility, measures of the likelihood of travel by transit rather than auto to non-work destinations, reducing the need for higher levels of vehicle availability.
- Average parking costs in the residence zone and in the work zone
- Parking availability in the residence zone and in the work zone, measuring the difficulty to find parking space required by the household

These variables were assembled from a number of sources to create the estimation data set. The sources include the 1990 MTC Household Survey, MTC zonal inputs, the San Francisco Planning Departments land use and business databases and a survey of the pedestrian environment in San Francisco. The basis for this data set is a file with one record for each person in the travel survey, obtained by extracting household and work tour data from the tour-based estimation data base. This file was expanded by adding zonal data, level of service data, and accessibility data. The zonal data included population, households, employment by type, area in square miles, area type, pedestrian environment factor, and parking costs. The level of service data included both auto and transit travel times and costs between the residence zone and each household members' workplace. The accessibility data included each of the measures discussed above, representing combinations of employment data and travel times by mode.

Vehicle Availability Model Estimation Results

Table 3 presents the estimation results for the vehicle availability model. A separate constant was estimated for each of the 4 alternatives except for a single "base" alternative. The base alternative was defined as the one for households with no vehicles available. The households with one vehicle available is the most likely alternative and has a positive constant. Households with more than one vehicle available have negative constants, increasing with more vehicles available because this is a less likely alternative.

Coefficients can be compared across alternatives to judge the reasonableness of these values. For example, households with 2 adults are most likely to have 2 vehicles available (coefficient of 1.924) and slightly more likely to have 3 or more vehicles available (coefficient of 0.806) than 1 vehicle available (coefficient of 0.642).

Testing Variables

The following process was used to determine which variables should be retained in the final model:

- Variables that were not reasonable to forecast were dropped from the model - including building type and tenure type (owner/renter)
- Age groups were reconfigured to <18, 18-24, 25-49, 50 and up because there were not enough samples in the 64 and up category to warrant a separate category. The number of people aged 18-24 in the household works better than the other variables.
- We tried specifications making the area type of home zone specific to number of vehicles available.
- Average travel time to work zone (applicable in case of multiple work locations for a household) was replaced by maximum and minimum travel times - maximum was finally retained for best results.
- The number of vehicles available was restricted to a maximum of one vehicle per person and the choice variable was re-coded - this reduced the model's explanatory power and so, was rejected.
- Coefficients for household income were separated by low, medium, and high income ranges - the resulting relative magnitudes of the income coefficients were not in proper order, and hence, this was rejected.
- We tried a number of accessibility variables - amount of total employment and employment by category reachable by auto within 13, 20, 27, and 40 minutes; amount of total employment and employment by category reachable by transit within 20, 30, 45, and 60 minutes; and various ratios of total employment accessibility by transit to auto. Employment within X minutes of highway travel time from home zone did not prove very useful in the models. The transit/auto accessibility ratio was retained in the final model.
- We tried a number of work zone related variables - the only one finally retained in the model is parking cost.
- Most of the pedestrian environment variables were dropped because they were insignificant. The vitality index was retained as the most significant factor. The safety variable was dropped because the coefficient should be positive (and it was not), indicating that a higher level of safety would increase pedestrian activity.
- The parking availability coefficient was used for the home zone rather than parking cost because the variable was prepared with the commuter in mind rather than the resident.
- We tried the employment densities of the home zone using the SF employment categories – but these were not retained because they were not significant.
- The model specifications do not use the variable for number of persons with drivers' licenses in the household. This variable adds a lot to the explanatory power of the model, however, it is not part of the data generated in the synthetic sample, and hence we would need another model that predicts the number of drivers' licenses in the household. The combination of the two models would not work as well as using a single model without the variable.

Testing Model Structure

The comparison of the model structure resulted in choosing the multinomial structure. Both model specifications were pretty well-behaved with coefficients for most variables meeting expectations with respect to relative magnitudes and signs. The nested logit model was rejected because it did not meet reasonable expectations. Some notes on this process are as follows:

- As can be expected, the multinomial and ordinal models differed in their specifications quite significantly. The Ordered Response Logit model estimation started with the final Multinomial Logit specification, however, some of the attributes had to be dropped. Some attributes that were irrelevant in the Multinomial Logit specification proved otherwise in the Ordered Response Logit models.
- For the Multinomial model, the coefficients for the variables capturing the number of people in different age groups in the household were of varying relative magnitudes, though of the proper sign.
- In the Ordered Response Logit models, because we try to differentiate between only two alternatives at a time and because they are relatively closer in characteristics (in Multinomial Logit, we use zero as base), the coefficients are not in general, as significant as those variables n the Multinomial Logit models.

Overall Model Results

Some important results to note are:

- Long term parking costs and parking availability are highly correlated, but the parking cost gives a better fit for the work zone. Generally, in work zones were commute parking costs are highest, residents also have to pay for parking, either directly or indirectly.
- A large number of the available household attributes provided significant explanatory power to the models.
- With respect to travel time to work, the model uses congested highway travel time to the workplace location. For households with multiple work locations, we tested the minimum, maximum and average times for all work locations and retained the maximum travel time. This indicates that the maximum travel time for multiple work locations is driving the decision to obtain more vehicles for travel. For those households with no work place location specified, we used a missing travel time dummy that is multiplied with the travel time.
- The number of people aged 18-24 in the household is a better indicator than the other persons by age group variables. This can be due to the fact that it takes people a while after they become an adult before they are used to having their own car.
- The part time worker variable has slightly more influence on choosing 2 or more vehicles than the full time worker variable, probably because part time workers are more likely to need vehicles for transportation than full time workers.

Table 3. Vehicle Availability Model

Summary Statistics						
Observations	1244					
Final likelihood	-1209.49					
Rho-squared (0)	0.2987					
Rho-squared (c)	0.2200					
Alternative:	0 veh.(base)	1 veh.		2 veh.		3+ veh.	
Chosen obs.:	323	529		303		89	
Variable	**Coefficient**	**T-statistic**	**Coefficient**	**T-statistic**	**Coefficient**	**T-statistic**
Household Variables						
Household income (000)	0.0262	5.8	0.0366	7.5	0.0398	6.8
2 adults in household	0.642	3.7	1.924	7.7	0.806	2.1
3 adults in household			1.874	6.0	1.917	4.5
# adults over 3 in household			0.714	2.9	1.005	2.9
Full time workers in household	0.361	2.6	0.490	2.9	0.946	4.6
Part time workers in household			0.722	3.3	1.293	4.4
# household members age 18-24	-0.317	-2.1	-0.381	-2.2	-0.381	-2.2
Level of Service Variables						
Max. auto time to work (min.)	0.0144	2.3	0.0273	4.0	0.0273	4.0
Transit/auto accessibility ratio	-0.128	-0.5	-0.641	-2.0	-0.641	-2.0
Work zone parking cost ($)	-0.250	-2.0	-0.359	-2.3	-0.832	-3.3
Locational Variables						
Home zone parking availability	-0.469	-1.8	-0.469	-1.8	-0.469	-1.8
Home zone vitality index			-0.218	-1.6	-0.432	-1.9
Households within half mile (000)	-0.145	-5.5	-0.185	-4.9	-0.310	-4.3
Constants						
Residual constant	0.909	1.4	-0.527	-0.7	-1.324	-1.6

Model Application

All of the models in this chapter were incorporated into a C++ program with the following structure:

- Read in and store all zone-based land use data and accessibility measures.
- Loop on the person records in the SF county synthetic sample for the appropriate forecast year, after this sample has been updated by applying the Workplace Location model.
- For each person record:

  - Apply the Vehicle Availability model. Calculate the probability for each vehicle alternative, and use a random Monte Carlo procedure to predict a single alternative (0 vehicle, 1 vehicle, 2 vehicles or 3+ vehicles per household).
  - Write the output as a series of vehicle choices, each with:

    1.the taz number

    2.the number of households with 0 vehicles

    3.the number of households with 1 vehicle

    4.the number of households with 2 vehicles

    5.the number of households with 3+ vehicles

- Print summary information across the sample, to use in validation and calibration.

APPENDIX A. Final Calibrated Vehicle Availability Model

Summary Statistics				
Observations		1244		
Final likelihood		-1209.49		
Rho-squared (0)		0.2987		
Rho-squared (c)		0.2200		
Alternative:	0 veh.(base)	1 veh.	2 veh.	3+ veh.
Chosen obs.:	323	529	303	89

Variable	**Coefficient**	**Coefficient**	**Coefficient**
Household Variables			
Household income (000)	0.0262	0.0366	0.0398
2 adults in household	0.642	1.924	0.806
3 adults in household		1.874	1.917
# adults over 3 in household		0.714	1.005
Full time workers in household	0.361	0.490	0.946
Part time workers in household		0.722	1.293
# household members age 18-24	-0.317	-0.381	-0.381
Level of Service Variables			
Max. auto time to work (min.)	0.0144	0.0273	0.0273
Transit/auto accessibility ratio	-0.128	-0.641	-0.641
Work zone parking cost ($)	-0.250	-0.359	-0.832
Locational Variables			
Home zone parking availability	-0.469	-0.469	-0.469
Home zone vitality index		-0.218	-0.432
Households within half mile (000)	-0.145	-0.185	-0.310
Constants			
Residual constant	0.969	-0.807	-1.641

APPENDIX B. Files for Running the Vehicle Availability Models for the Base Year

TAZDATA.DAT - The zonal data file. It includes1738 records in space-delimited format.

ACCESS.DAT - The accessibility variable data file. In includes 766 records in space-delimited format.

VEHAVL.CPP, VEHAVL.EXE - the vehicle availability code and executable. It uses the two input files above, plus TOURDC.OUT, the output from the work location model application. It creates VEHAVL.OUT, which has the same format as TOURDC.OUT, but has the predicted number of vehicles in place of the PUMS value.

It also creates VEHAVL.SUM, which has a summary record for each of the 766 tazs.

Each record has:

- the taz number
- the number of households with 0 vehicles
- the number of households with 1 vehicle
- the number of households with 2 vehicles
- the number of households with 3+ vehicles