

| | | | |
|--|--|--|-----------|
| 1. Report No. SWUTC/06/167766-1 | 2. Government Accession No. | 3. Recipient's Catalog No. | |
| 4. Title and Subtitle ESTIMATION OF TRAVELERS' VALUES OF TIME USING A STATED-PREFERENCE SURVEY WITH VARIABLE PRICING OPTIONS | | 5. Report Date December 2006 | |
| | | 6. Performing Organization Code | |
| 7. Author(s) Mark Burris and Sunil Patil | | 8. Performing Organization Report No. Report 167766-1 | |
| 9. Performing Organization Name and Address Texas Transportation Institute Texas A&M University System College Station, Texas 77843-3135 | | 10. Work Unit No. (TRAIS) | |
| | | 11. Contract or Grant No. 10727 | |
| 12. Sponsoring Agency Name and Address Southwest Region University Transportation Center Texas Transportation Institute Texas A&M University System College Station, Texas 77843-3135 | | 13. Type of Report and Period Covered | |
| | | 14. Sponsoring Agency Code | |
| 15. Supplementary Notes Supported by general revenues from the State of Texas. | | | |
| 16. Abstract This study analyzed data from a stated preference survey of Houston travelers faced with numerous mode choices, including value pricing options. The study: <ul style="list-style-type: none"> (a) examined the possibility of using a genetic algorithm to estimate mode choice models while removing the need of making the IIA assumption, (b) estimated nested logit models, (c) attempted to estimate random parameter logit models, and (d) estimated numerous multinomial logit models. After comparing different specifications and optimization techniques (namely the genetic algorithm and Newton-Raphson method in the econometric software, Limited Dependent LIMDEP) the multinomial logit model estimation using LIMDEP was found to be more efficient because of easy estimation and a much lower time requirement for estimation. Hence a multinomial logit model was used for estimating the values of travel time savings (VTTS) and the penalty for changing travel schedule for different groups of travelers. The values estimated for the penalty for changing travel schedule were not statistically significant and were therefore not used. The values estimated for travel time savings were significant and comparable to those obtained in previous studies. It was found that the average VTTS was 39 percent of the wage rate and was higher for females, non-commuters, households with few vehicles, and the wealthy. | | | |
| 17. Key Words Value of Travel Time Savings, Mode Choice Models, Genetic Algorithms | | 18. Distribution Statement No restrictions. This document is available to the public through NTIS: National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161 | |
| 19. Security Classif.(of this report) Unclassified | 20. Security Classif.(of this page) Unclassified | 21. No. of Pages 76 | 22. Price |

Estimation of Travelers' Values of Time Using a Stated-Preference Survey with Variable Pricing Options

by

Mark Burris
Assistant Research Engineer

and

Sunil Patil
Graduate Assistant Research

Report SWUTC/06/167766-1

Sponsored by the
Southwest Region University Transportation Center

December 2006

Texas Transportation Institute
Texas A&M University System
College Station, Texas 77843-3135

ACKNOWLEDGEMENT

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers Program to the Southwest Region University Transportation Center which is funded 50% with general revenue funds from the State of Texas.

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT

This study analyzed data from a stated preference survey of Houston travelers faced with numerous mode choices, including value pricing options. The study:

- (a) examined the possibility of using a genetic algorithm to estimate mode choice models while removing the need of making the IIA assumption,
- (b) estimated nested logit models,
- (c) attempted to estimate random parameter logit models, and
- (d) estimated numerous multinomial logit models.

After comparing different specifications and optimization techniques (namely the genetic algorithm and Newton-Raphson method in the econometric software, Limited Dependent LIMDEP) the multinomial logit model estimation using LIMDEP was found to be more efficient because of easy estimation and a much lower time requirement for estimation. Hence a multinomial logit model was used for estimating the values of travel time savings (VTTS) and the penalty for changing travel schedule for different groups of travelers. The values estimated for the penalty for changing travel schedule were not statistically significant and were therefore not used. The values estimated for travel time savings were significant and comparable to those obtained in previous studies. It was found that the average VTTS was 39 percent of the wage rate and was higher for females, non-commuters, households with few vehicles, and the wealthy.

EXECUTIVE SUMMARY

This study aimed at estimating the value of travel time savings and value of penalty for changing travel schedule for the travelers on Katy Freeway in Houston. Mode choice models were used to estimate the values of travel time savings. The study started with a simple multinomial logit model for the general purpose lane travelers on Katy Freeway. The stated preference surveys used for data collection in this study involved a large number of modes to choose from. The traveler's choice set contained nine travel alternatives, with some of these alternatives being somewhat similar. Hence, attempts were made to check if more complex modeling methods would work better than a multinomial logit model.

To begin, the study attempted to estimate a random parameter logit model for the given data. The estimation results indicated that nothing was gained by forcing parameters to vary across individuals. The standard deviations of the random parameters were not significant to justify the assumption that these parameters vary across individuals. This may be partly due to the nature of the data which contain four stated preference questions per individual; although for modeling they are counted as four separate individuals. Hence the random parameter logit model was not preferred over the MNL model. This study successfully estimated a nested logit model for modeling mode choice of Katy Freeway travelers on general purpose lanes. Different nest specifications based on occupancy, cost of travel and time of departure were tested to choose the final nested logit model satisfying both theoretical and intuitive criteria. The nested logit models were found to have nearly similar log-likelihood values as compared to simple MNL model. Also, the percentage of correct mode choice predictions and value of travel time savings obtained by both MNL and NL were similar, hence the simpler MNL model was used for further analysis.

The study then estimated MNL models using genetic algorithm optimization for solving the likelihood function. It was assumed that the set of parameters in the utility equation could be considered chromosomes, and a function such as the log-likelihood function could be the fitness function. Different trials were carried out to select the best genetic algorithm options such as population size, selection function, stopping criteria, etc. The results from the genetic algorithm

were comparable to MNL model estimation by using LIMDEP. However, with the given optimization settings, the genetic algorithm could not achieve the optima and hence the log-likelihood values of models obtained by using genetic algorithm optimization were slightly lower than those obtained by LIMDEP. Another drawback of using a genetic algorithm for likelihood estimation of MNL model was large computational time. However, the study showed that the genetic algorithm can be successfully used for mode choice modeling. Further, genetic algorithms can prove very useful in case of model specification with complex likelihood function which is difficult to optimize using other techniques. It is also possible to avoid the need for making assumptions such as independence of irrelevant alternatives by correctly specifying the fitness functions. One of the attempts for such a fitness function was made by maximizing the number of correct predictions.

To calculate the value of travel time savings and penalty for changing travel schedule for various groups of travelers, the study estimated MNL models using LIMDEP. In order to account for heterogeneity among the travelers the sample was divided into groups (segments) based on gender, number of vehicles, trip purpose and income. These groups were found to have different utility functions and hence different VTTS and VPCS values. The VPCS values calculated were not significant hence they were not used. The VTTS values different by group and were generally comparable to those estimated by previous research. For example, the average VTTS was found to be 39% of the wage rate, with high income respondents having an average VTTS of \$12.29 per hour and lower income travelers having a VTTS of \$5.84 per hour. Females were found to have higher VTTS than males, non-commuters had a higher VTTS than commuters, and travelers with few vehicles had higher VTTS than those with more than two vehicles.

TABLE OF CONTENTS

| | <u>Page</u> |
|--|-------------|
| 1.0 INTRODUCTION | 1 |
| 2.0 DATA COLLECTION | 3 |
| 3.0 INITIAL MODELING ATTEMPTS..... | 5 |
| 3.1 Multinomial Logit Model | 5 |
| 3.2 Random Parameter/Mixed Logit Model..... | 9 |
| 3.3 Nested Logit Models | 15 |
| 3.4 Model Estimation Using Genetic Algorithms | 32 |
| 3.5 Comparison of Logit Models and GA Estimated Models | 40 |
| 3.6 Improving the Discrete Choice Model by Using a Genetic Algorithm..... | 45 |
| 4.0 VTTS AND VPCS FOR DIFFERENT GROUPS OF TRAVELERS | 51 |
| 4.1 Using GA for Groups of Travelers | 52 |
| 4.2 Models for Different Groups of Travelers and Corresponding VTTS and VPCS | 53 |
| 5.0 CONCLUSIONS | 59 |
| REFERENCES | 61 |

LIST OF FIGURES

| | <u>Page</u> |
|---|-------------|
| FIGURE 1 Typical LIMDEP Command for Random Parameter Modeling | 11 |
| FIGURE 2a Typical Output for RPL Model Estimation in LIMDEP | 12 |
| FIGURE 3 Error Message Returned During the RPL Modeling Trials..... | 14 |
| FIGURE 4 Conceptual Diagram of a Nested Logit Model..... | 17 |
| FIGURE 5 A typical nested logit model command in LIMDEP | 19 |
| FIGURE 6a Different Nest Specifications Used for the Study | 21 |
| FIGURE 7 Commands for Likelihood Ratio Test in LIMDEP | 28 |
| FIGURE 8 An Output for LR Test Using LIMDEP | 29 |
| FIGURE 9 Nesting Details for Nest ID-16 | 30 |
| FIGURE 10 A Typical Genetic Algorithm Tool Window in MATLAB..... | 34 |
| FIGURE 11 Fitness Functions Used for GA estimation of Discrete Choice Models..... | 36 |
| FIGURE 12 Commands for Generating the Cross-tab Matrix Using Estimated Coefficients..... | 44 |
| FIGURE 13 Typical Commands for Rejecting Unnecessary Sample | 52 |

LIST OF TABLES

| | <u>Page</u> |
|--|-------------|
| TABLE 1 Model Specification for Katy Freeway Travelers | 7 |
| TABLE 2 Modeling Results for Katy Freeway Travelers..... | 8 |
| TABLE 3 Summary of Results of RPL Modeling Trials | 14 |
| TABLE 4 Interpretation of Value of τ_k (8) | 30 |
| TABLE 5 LR Test for Katy Freeway Traveler Model | 30 |
| TABLE 6 t-test for Inclusive Value Parameters of Estimated NL Model | 30 |
| TABLE 7 Nested Logit Model for Katy Freeway Travelers (Nest ID-16) | 31 |
| TABLE 8 Settings and Methods Used for the Genetic Algorithm..... | 38 |
| TABLE 9 Models Estimated by GA Using Different Fitness Functions | 41 |
| TABLE 10 Crosstab Matrix for the Model Estimated by Using Fitness Function Predmax (Trips)..... | 42 |
| TABLE 11 Crosstab Matrix for the Model Estimated by Using Fitness Function Lmax (Trips)..... | 42 |
| TABLE 12 Crosstab Matrix for the Model Estimated by Using Fitness Function LLmax (Trips) | 43 |
| TABLE 13 Improved MNL model estimated by LIMDEP and GA | 46 |
| TABLE 14 Crosstab Matrix for the Improved GA Model (Trips)..... | 48 |
| TABLE 15 MNL Model Estimated by LIMDEP and GA during Improvement Process | 48 |
| TABLE 16 Crosstab Matrix for the GA Model Estimated During Improvement Process (Trips)..... | 50 |
| TABLE 17 Market Segmentation Test for Models of Groups Based on Gender | 54 |
| TABLE 18 Models for Groups based on Gender | 54 |
| TABLE 19 Models for Groups based on Trip Purpose | 55 |
| TABLE 20 Models for Groups based on Number of Vehicles | 56 |
| TABLE 21 Models for Groups based on Income..... | 57 |
| TABLE 22 Estimated VTTS for Katy Freeway Travelers | 58 |

1.0 INTRODUCTION

Traveler's mode choice is usually modeled in the theoretical framework of Random Utility Maximization (RUM). According to the RUM principle the traveler is assumed to choose a mode of travel which maximizes his/her utility. Discrete choice models such as Multinomial Logit (MNL) models are often used for modeling mode choice. In order to overcome some of the serious drawbacks of the MNL models, Nested Logit (NL) models are recommended in some cases. This research examined NL models and proposed an alternate methodology using Genetic Algorithms (GA) for estimating mode choice models for the travelers on the Katy Freeway in Houston. The study used these models to examine the effect of variable pricing on mode choice and time of departure.

Congestion pricing, or variable pricing, can reduce travel demand during the peak hours (1). Time of day and occupancy-based pricing can encourage travelers with lower values of time to shift their time of departure to before/after the peak hours to avoid paying higher tolls or they choose carpooling options to take benefit of the lower toll for high occupancy vehicles. Hence, these shifts potentially depend on the Value of Travel Time Savings (VTTS) and Value of Penalty for Changing travel Schedule (VPCS) of the traveler.

The VTTS can be defined as the monetary value that is assigned to unit travel time savings by a traveler. Similarly, traveler's value of penalty for changing travel schedule can be considered to be equivalent to monetary value of changing the travel schedule. Travelers typically have penalties for early departures (e.g. less time for family, recreation or other work) or late departures (e.g. arriving late to work place or loss incurred due to late arrival). Thus travelers generally prefer their current schedule unless an additional monetary cost (toll) is charged to maintain their current choice (2).

The current study examined the VTTS and VPCS for travelers on the Katy Freeway under a variable pricing scenario, in which the tolls vary both by time of day and by vehicle occupancy. The study also determined the VTTS and VPCS values for different groups of travelers and

compared these values in order to draw some potentially useful conclusions for both policy makers and transportation planners.

2.0 DATA COLLECTION

The data used for this report were collected during November 2003 on the Katy Freeway (I-10) and the Northwest Freeway (US-290) in Houston. The data were collected from travelers who were driving on the General Purpose Lanes (GPL), along with the High Occupancy Vehicle (HOV) lane travelers who were traveling by carpooling, slugging or using transit on both the Katy and Northwest Freeway corridors during both the peak and the off-peak hours. This study examined data only from Katy Freeway travelers traveling on general purpose lanes for consistency. Both revealed preference and stated preference questions were asked in the survey. The data include answers to questions regarding the:

- Respondent's most recent trip;
- Respondent's general perceptions and attitudes towards the QuickRide program;
- Respondent's choices among different travel scenarios; and
- Respondent's socioeconomic and demographic characteristics.

The respondent's choices among different travel scenarios were recorded in four stated preference questions with four mode choice options available for each question. Each traveler was asked to choose his/her preferred mode among four hypothesized scenarios marked as A, B, C, and D in each question. Each scenario was characterized by mode, travel time (two or three levels depending on mode) and toll rate (two or three levels depending on mode) factors. In total there were nine potential mode choices with various occupancy criteria (such as single occupancy vehicle (SOV), high occupancy vehicle with two travelers in a vehicle (HOV2) and high occupancy vehicle with three travelers in a vehicle (HOV3)) after combining the data sets for peak and off-peak hours, including:

- A. SOV on the HOV lane in the off-peak period (SOV-HOV-OP);
- B. SOV on the GPLs in the off-peak period (SOV-GPL-OP);
- C. HOV2 on the GPLs in the peak period (HOV2-GPL-P);
- D. SOV on the HOV lane in the peak period (SOV-HOV-P);
- E. Transit, using the park and ride lot (P&R-T);
- F. HOV2 on the HOV lane in the off-peak period (HOV2-HOV-OP);

- G. SOV on the GPLs in the peak period (SOV-GPL-P);
- H. HOV2 on the HOV lane in the peak period (HOV2-HOV-P); and
- I. HOV3 on the HOV lane in the peak period (HOV3-HOV-P).

For the purpose of stated-preference data analysis, each respondent's information was recorded in 16 rows, with each row corresponding to one of four options for the four stated preference questions. The socioeconomic profile data of each respondent were duplicated in all of the 16 rows. Additional details about the survey can be found in Burris and Figueroa, 2006. After this data formatting, a simple multinomial logit model was developed for Katy Freeway travelers. This model was used as the basis for further advanced analysis such as random parameter modeling, nested logit modeling etc., as outlined in the following sections.

3.0 INITIAL MODELING ATTEMPTS

Using the above data, different mode choice models such as a Multinomial Logit Model, Nested Logit Model, and a Random Parameters Logit model were estimated as described below. These models were used for comparison with the models estimated by using genetic algorithm.

3.1 Multinomial Logit Model

Multinomial logit models are typically used to predict the mode choice for an individual and are based on the concept of random utility maximization. The utility of any mode is assumed to consist of two components, the systematic component and the error component. The logit model assumes that the error components are extreme value (or Gumbel) distributed and the choice probability that an individual n chooses mode i ($i=1,2,\dots,J$) from a set of J alternatives is given by Equation 1.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j=1}^J e^{V_{nj}}} \quad (1)$$

Where: P_{ni} is the probability of the individual choosing alternative i and
 V_{nj} is the systematic component of the utility of alternative j .

The equations for the systematic utilities consist of variables which account for the attributes of both the mode and the individual (decision maker). For example, Equation 2 is an example of a systematic utility equation for mode 'i.'

$$V_{ni} = \beta_{i0} + \beta_1 * TravelTime_{ni} + \beta_2 * TravelCost_{ni} + \beta_{i3} * Income_n \quad (2)$$

Where: β_x = the estimated coefficient of for each independent variable x ,

$TravelTime_{ni}$ = the travel time for mode i for individual n ,

$TravelCost_{ni}$ = the cost of travel on mode i for individual n , and

$Income$ = the income of individual n .

This equation can be used to estimate the Value of Travel Time Savings for travelers if the coefficients β_1 and β_2 are included in the utility equations for all modes. The VTTS will then be

given by the partial derivative of the utility equation with respect to time divided by the partial derivative of the utility equation with respect to cost, in this case this results in the ratio β_1/β_2 . The basic MNL model estimated in an earlier study using econometric software LIMDEP for our data is given in Table 1 and Table 2 (2,3).

TABLE 1 Model Specification for Katy Freeway Travelers

| Utility Function for Mode | Variable Name | Description | Coefficient |
|---------------------------|---|--|--------------|
| All | trtime | The travel time savings obtained by using the HOV lane (minutes); the value was 0 for mode A, C, and G, because there were no travel time savings if the trip occurred on the GPLs | β_9 |
| | tollinc | Toll / (annual household income / 2000) | β_{10} |
| A (SOV on GPL Off-peak) | A A | The alternative-specific constant | β_1 |
| | apeak | The dummy variable used to describe if the traveler was driving during peak hours, yes = 1, no = 0 | β_{11} |
| | aeduhs | The dummy variable used to describe if the traveler's education level was high school graduate, yes = 1, no = 0 | β_{12} |
| B (SOV on HOV Off-peak) | A B | The alternative-specific constant | β_2 |
| | btrlth | The total travel time of the trip (minutes) | β_{13} |
| | beduhs | The dummy variable used to describe if the traveler's education level was high school graduate, yes = 1, no = 0 | β_{14} |
| C (HOV2 on GPL Peak) | A C | The alternative-specific constant | β_3 |
| | cage_a | The dummy variable used to describe if the traveler's age was from 25 to 54 years old, yes = 1, no = 0 | β_{15} |
| | ceducv | The dummy variable used to describe if the traveler's education level was some college / vocational, yes = 1, no = 0 | β_{16} |
| | chtpm | The dummy variable used to describe if the traveler's household type was married without children, yes = 1, no = 0 | β_{17} |
| D (SOV on HOV Peak) | A D | The alternative-specific constant | β_4 |
| | dtrlth | The total travel time of the trip (minutes) | β_{18} |
| | dtprec | The dummy variable used to describe if the traveler's trip purpose was recreational, yes = 1, no = 0 | β_{19} |
| E (Park & Ride Transit) | A E | The alternative-specific constant | β_5 |
| | Etrlth | The total travel time of the trip (minutes) | β_{20} |
| | eage_y | The dummy variable used to describe if the traveler's age was from 16 to 24 years old, yes = 1, no = 0 | β_{21} |
| | ehtpm | The dummy variable used to describe if the traveler's household type was married without children, yes = 1, no = 0 | β_{22} |
| | envehs | The number of motor vehicles (including cars, vans, trucks, and motorcycles) available in the traveler's household | β_{23} |
| | eeducv | The dummy variable used to describe if the traveler's education level was some college / vocational, yes = 1, no = 0 | β_{24} |
| | etpcom | The dummy variable used to describe if the traveler's trip purpose was commuting, yes = 1, no = 0 | β_{25} |
| | etprec | The dummy variable used to describe if the traveler's trip purpose was recreational, yes = 1, no = 0 | β_{26} |
| F (HOV2 on HOV Off-peak) | A F | The alternative-specific constant | β_6 |
| | ftprec | The dummy variable used to describe if the traveler's trip purpose was recreational, yes = 1, no = 0 | β_{27} |
| G (SOV on GPL Peak) | A G | The alternative-specific constant | β_7 |
| | goffpk | The dummy variable used to describe if the traveler was driving during off-peak hours, yes = 1, no = 0 | β_{28} |
| | gtpcom | The dummy variable used to describe if the traveler's trip purpose was commuting, yes = 1, no = 0 | β_{29} |
| | Ghtpm | The dummy variable used to describe if the traveler's household type was married without children, yes = 1, no = 0 | β_{30} |
| H (HOV2 on HOV Peak) | A H | The alternative-specific constant | β_8 |
| | htpcom | The dummy variable used to describe if the traveler's trip purpose was commuting, yes = 1, no = 0 | β_{31} |
| I (HOV3 on HOV Peak) | The utility function of mode I only contained the generic variables, trtime and tollinc, because mode I was specified as the reference mode | | |

TABLE 2 Modeling Results for Katy Freeway Travelers

| Variable | Coefficient | | Standard Error | T-stat | P-value |
|------------------------|--------------|--------|-----------------------------------|--------|---------|
| trtime | β_9 | -0.07 | 0.01 | -10.56 | 0.00 |
| tollinc | β_{10} | -10.74 | 0.12 | -9.06 | 0.00 |
| apeak | β_{11} | -0.31 | 0.16 | -1.95 | 0.05 |
| aeduhs | β_{12} | -1.01 | 0.47 | -2.17 | 0.03 |
| btrlth | β_{13} | 0.01 | 0.00 | 2.47 | 0.01 |
| beduhs | β_{14} | -1.66 | 0.62 | -2.67 | 0.01 |
| cage_a | β_{15} | -1.09 | 0.39 | -2.78 | 0.01 |
| ceducv | β_{16} | 2.17 | 0.39 | 5.53 | 0.00 |
| chtpm | β_{17} | 0.85 | 0.41 | 2.10 | 0.04 |
| dtrlth | β_{18} | 0.01 | 0.00 | 4.08 | 0.00 |
| dtprec | β_{19} | -1.33 | 0.45 | -2.95 | 0.00 |
| etrlth | β_{20} | 0.02 | 0.01 | 2.35 | 0.02 |
| eage_y | β_{21} | 2.41 | 0.47 | 5.18 | 0.00 |
| ehtpm | β_{22} | 1.45 | 0.36 | 3.98 | 0.00 |
| envehs | β_{23} | 0.41 | 0.17 | 2.45 | 0.01 |
| eeducv | β_{24} | 0.96 | 0.39 | 2.46 | 0.01 |
| etpcom | β_{25} | 2.66 | 1.06 | 2.52 | 0.01 |
| etprec | β_{26} | 3.11 | 1.18 | 2.64 | 0.01 |
| ftprec | β_{27} | -1.14 | 0.64 | -1.79 | 0.07 |
| goffpk | β_{28} | -0.23 | 0.13 | -1.73 | 0.08 |
| gtpcom | β_{29} | 0.75 | 0.18 | 4.18 | 0.00 |
| ghtpm | β_{30} | 0.39 | 0.15 | 2.52 | 0.01 |
| htpcom | β_{31} | 1.52 | 0.51 | 2.96 | 0.00 |
| A_A | β_1 | 3.04 | 0.27 | 11.12 | 0.00 |
| A_B | β_2 | 1.73 | 0.33 | 5.25 | 0.00 |
| A_C | β_3 | 0.42 | 0.49 | 0.86 | 0.39 |
| A_D | β_4 | 0.84 | 0.35 | 2.44 | 0.01 |
| A_E | β_5 | -5.18 | 1.26 | -4.12 | 0.00 |
| A_F | β_6 | 0.24 | 0.29 | 0.81 | 0.42 |
| A_G | β_7 | 2.41 | 0.31 | 7.87 | 0.00 |
| A_H | β_8 | -0.29 | 0.54 | -0.54 | 0.59 |
| $\rho^2 = 0.584$ | | | Log likelihood function = -1683.3 | | |
| $\bar{\rho}^2 = 0.582$ | | | Number of observations = 1845 | | |

Based on the data shown in Table 1 and Table 2, the utility functions of all the travel mode options for Katy freeway travelers were as follows:

$$U_{\text{SOV-GPL-OP}} = 3.0400 - 10.739\text{tollinc} - 0.3105\text{apeak} - 1.0080\text{aeduhs}$$

$$U_{\text{SOV-HOV-OP}} = 1.7282 - 0.0724\text{trtime} - 10.739\text{tollinc} + 0.0074\text{btrlth} - 1.6600\text{beduhs}$$

$$U_{\text{HOV2-GPL-P}} = 0.4217 - 10.739\text{tollinc} - 1.0881\text{cage_a} + 2.1702\text{ceducv} \\ + 0.8529\text{chtpm}$$

$$U_{\text{SOV-HOV-P}} = 0.8449 - 0.0724\text{trtime} - 10.739\text{tollinc} + 0.0146\text{dtrlth} - 1.3286\text{dtprec}$$

$$U_{\text{P\&R-T}} = -5.1786 - 0.0724\text{trtime} - 10.739\text{tollinc} + 0.0159\text{etrlth} + 2.4100\text{eage_y} \\ + 1.4486\text{ehtpm} + 0.4129\text{envehs} + 0.9550\text{eeducv} + 2.6567\text{etpcom} \\ + 3.1082\text{etprec}$$

$$U_{\text{HOV2-HOV-OP}} = 0.2356 - 0.0724\text{trtime} - 10.739\text{tollinc} - 1.1394\text{ftprec}$$

$$U_{\text{SOV-GPL-P}} = 2.4131 - 10.739\text{tollinc} - 0.2289\text{goffpk} + 0.7474\text{gtpcom} \\ + 0.3889\text{ghtpm}$$

$$U_{\text{HOV2-HOV-P}} = -0.2909 - 0.0724\text{trtime} - 10.739\text{tollinc} + 1.5158\text{htpcom}$$

$$U_{\text{HOV3-HOV-P}} = -0.0724\text{trtime} - 10.739\text{tollinc}$$

The value of travel time savings using the above model was estimated to be around 39% of hourly wage rate (\$/hr) using the coefficient of variables tollinc (toll/approx. hourly wage rate in \$/hr) and trtime (in minutes). The details about the exact formula used are given in section 4.0 of this report.

3.2 Random Parameter/ Mixed Logit Model

Although multinomial logit models are relatively straight forward to estimate, they have certain limitations including their need for Independence of Irrelevant Alternatives (IIA). The IIA property of the MNL restricts the ratio of the choice probabilities for any pair of alternatives to be independent of the existence and characteristics of other alternatives in the choice set. This restriction implies that the introduction of a new mode or improvements to any existing mode will affect all other modes proportionately. That is, the new or improved mode will reduce the probability of existing modes in proportion to their probabilities before the change (4). The presence of highly similar choices in a choice set may cause violation of the IIA assumption. For example, in this study the choices which involve carpooling (such as options involving HOV2, HOV3) are more alike than drive-alone options (involving SOV); hence increase in toll for HOV2 mode will affect the probability of choosing HOV3 options more than that of SOV

options. In such instances it is common to use the nested logit (NL) model or random parameter logit (RPL) model. In case of NL models we have freedom of grouping the similar choices in one nest such that IIA is valid inside it and not across different nests.

The random parameter logit (RPL) model is similar to the standard logit model, except that in case of the RPL models, some individual preference parameters are interpreted as being random. In the case of RPL model the utility of person n from alternative j is specified as shown in Equation 3 (with notation from (1)).

$$U_{nj} = \beta'_n x_{nj} + \varepsilon_{nj} \quad (3)$$

Where: x_{nj} are observed variables that relate to the alternative and decision maker,

β_n is a vector of coefficients of these variables for person n representing that person's tastes, and

ε_{nj} is a random term that is iid extreme value.

The coefficients vary over decision makers in the population with density $f(\beta)$. The choice probability for RPL model can be expressed as given in the Equation 4 [with notation from (1)].

$$P_{ni} = \int \left(\frac{e^{\beta'_i x_{ni}}}{\sum_j e^{\beta'_i x_{nj}}} \right) f(\beta) d\beta \quad (4)$$

Modelers can specify different distributions for the coefficients for estimating the parameters of that distribution. Thus the model is very flexible and can approximate any random utility model (5).

3.2.1 Random Parameter Modeling Attempts

In this study, attempts were made to check the applicability of random parameter modeling to Katy Freeway traveler data by assuming normal and lognormal distributions for several parameters, except the dummy variables. Lognormal distribution can be used only for the parameters which always take positive values. For the parameter TRTIME, which always takes negative values, one can create new variable NRTIME = TRTIME * -1. All the alternative

specific constants (A_A, A_B, A_C, A_D, A_E, A_F, A_G, A_H) and the parameters corresponding TRTIME and TOLLINC were studied as part of these modeling attempts.

A typical LIMDEP command for random parameter modeling is shown in Figure 1, where Pts = n specifies number of random draws to use for estimating the unconditional probabilities by simulation. Very large values of n greatly increase the estimation time, hence one can start from smaller values such as 10 and then increase n if the parameter is found to be distributed across individuals. A_E(N) specifies the assumed distribution is normal for the A_E variable; similarly, A_E(L) specifies the lognormal distribution. A typical output for RPL model estimation using LIMDEP is shown in Figure 2a and Figure 2b.

```
NLOGIT;Lhs=CHDUMMY,NALT,ALTS;
      Choices=A,B,C,D,E,F,G,H,I;

Model:U(A)= A_A +btime* TRTIME + btoll*TOLLINC + bapeak*PEAK +
          baeduhs*EDUCHS/
      U(B)= A_B +btime *TRTIME + btoll*TOLLINC +
          btrlth*TRLENGTH+ bbeduhs*EDUCHS/
      U(C)= A_C + btime*TRTIME + btoll*TOLLINC
          + bcage_A*AGE_YNG + bceducv* EDUCSCV
          +bchtpm*HHTYPEM/
      U(D)= A_D + btime*TRTIME +btoll* TOLLINC
          + bdtrlth*TRLENGTH+
          bdtprec*TPREC/
      U(E)= A_E + btime*TRTIME + btoll*TOLLINC
          +betrlth*TRLENGTH + beage_Y*AGE_YNG
          + behtpm*HHTYPEM +benvehs*NVEHS
          +beeducv*EDUCSCV+betpcom*TPCOMM+betprec*TPREC/

      U(F)= A_F + btime*TRTIME +btoll* TOLLINC
          +bftprec* TPREC/
      U(G)= A_G + btime*TRTIME + btoll*TOLLINC + bgpeak*PEAK+
          bgtpcom*TPCOMM+ bghtpm*HHTYPEM/
      U(H)= A_H + btime*TRTIME + btoll*TOLLINC
          +bhtpcom* TPCOMM/
      U(I)= btime*TRTIME + btoll*TOLLINC;

Rpl;
Fcn= A_B(N) ; Pts=20$
```

FIGURE 1 Typical LIMDEP Command for Random Parameter Modeling

Normal exit from iterations. Exit status=0.

```
+-----+
| Random Parameters Logit Model
| Maximum Likelihood Estimates
| Dependent variable           CHDUMMY
| Weighting variable           ONE
| Number of observations       7388
| Iterations completed        30
| Log likelihood function      -1683.261
| Restricted log likelihood    -4058.274
| Chi-squared                 4750.026
| Degrees of freedom          32
| Significance level           .0000000
| R2=1-LogL/LogL*   Log-L fncn  R-sqrd  RsqAdj
| No coefficients   -4058.2738   .58523   .58282
| Constants only.   Must be computed directly.
|                   Use NLOGIT ;...; RHS=ONE $
| At start values   -1683.2636   .00000   -.00581
| Response data are given as ind. choice.
+-----+
+-----+
| Random Parameters Logit Model
| Replications for simulated probs. = 10
| Number of obs.= 1847, skipped    0 bad obs.
+-----+
```

FIGURE 2a Typical Output for RPL Model Estimation in LIMDEP

| Variable of X | Coefficient | Standard Error | b/St.Er. | P[Z >z] | Mean |
|--|------------------|----------------|----------|----------|------|
| Random parameters in utility functions | | | | | |
| BTOLL | -1.073744436 | .94318066E-01 | -11.384 | .0000 | |
| Nonrandom parameters in utility functions | | | | | |
| A_A | 3.030644695 | .28061941 | 10.800 | .0000 | |
| BTIME | -.7264249638E-01 | .69385355E-02 | -10.469 | .0000 | |
| BAPEAK | -.2868527310 | .16446416 | -1.744 | .0811 | |
| BAEDUHS | -.8861802365 | .46028041 | -1.925 | .0542 | |
| A_B | 1.742322275 | .33540436 | 5.195 | .0000 | |
| BBTRLTH | .7032306897E-02 | .29456074E-02 | 2.387 | .0170 | |
| BBEDUHS | -1.537294272 | .68052863 | -2.259 | .0239 | |
| A_C | -.6776900889 | .44312891 | -1.529 | .1262 | |
| BCAGE_A | 2.574420452 | .58368610 | 4.411 | .0000 | |
| BCEDUCV | 2.327901148 | .40368141 | 5.767 | .0000 | |
| BCHTPM | 1.025878068 | .40784866 | 2.515 | .0119 | |
| A_D | .8405495283 | .34191902 | 2.458 | .0140 | |
| BDTRLTH | .1459505047E-01 | .34150226E-02 | 4.274 | .0000 | |
| BDTPREC | -1.355431066 | .44292611 | -3.060 | .0022 | |
| A_E | -5.206082050 | 1.1260175 | -4.623 | .0000 | |
| BETRLTH | .1456269501E-01 | .96470611E-02 | 1.510 | .1312 | |
| BEAGE_Y | 2.540577790 | .53405274 | 4.757 | .0000 | |
| BEHTPM | 1.512403204 | .41974689 | 3.603 | .0003 | |
| BENVEHS | .3930016674 | .20888388 | 1.881 | .0599 | |
| BEEDUCV | 1.013900886 | .44143698 | 2.297 | .0216 | |
| BETPCOM | 2.766191267 | 1.2020920 | 2.301 | .0214 | |
| BETPREC | 3.215019528 | 1.3522861 | 2.377 | .0174 | |
| A_F | .2281223011 | .30065298 | .759 | .4480 | |
| BFTPREC | -1.157543167 | .65509181 | -1.767 | .0772 | |
| A_G | 2.402799837 | .30708985 | 7.824 | .0000 | |
| BGPEAK | .2226404217 | .13510355 | 1.648 | .0994 | |
| BGTPCOM | .7664511626 | .18599120 | 4.121 | .0000 | |
| BGHTPM | .3874706973 | .15894504 | 2.438 | .0148 | |
| A_H | -.3193807071 | .53577672 | -.596 | .5511 | |
| BHTPCOM | 1.546419382 | .50772579 | 3.046 | .0023 | |
| Derived standard deviations of parameter distributions | | | | | |
| sBTOLL | .4817370300E-02 | .58035363E-01 | .083 | .9338 | |

FIGURE 2b —Typical Output for RPL Model Estimation in LIMDEP

Attempts were made to check the assumption of normal and lognormal distributions for above mentioned parameters. Some of the attempts for our dataset, with better results, are shown in Table 3. Each row in the table corresponds to a model, and the parameter which was made to vary across individuals is listed in the first column. Also the possibility of grouping the parameters or having more than one random parameter at a time was investigated. Further, in order to observe the effect of increase in n (Pts=n) on the model estimation the value of n was

increased from 10 to 1000 for the parameters which showed better results for n=10. The results obtained were not consistent as increasing n did not improve the model's performance, and it returned the error shown in the Figure 3 in most of the cases.

TABLE 3 Summary of Results of RPL Modeling Trials

| Name of the parameter | Assumed Distribution | n (Pts=n) | Without Random Effect | | | With Random Effect | | | | |
|-----------------------|----------------------|-----------|-----------------------|-------------|---------------------|--------------------|-------------|---------------------|----------------------------------|--------------------------|
| | | | Value of parameter | b/std.error | Adj. $\bar{\rho}^2$ | Value of parameter | b/std.error | Adj. $\bar{\rho}^2$ | Value of Std. Dev. for parameter | b/std.error for std dev. |
| A_B | Normal | 20 | 1.73 | 5.25 | 0.58 | 1.73 | 5.14 | 0.58 | 0.03 | 0.35 |
| A_D | Normal | 20 | 0.85 | 2.44 | 0.58 | 0.85 | 2.47 | 0.58 | 0.01 | 0.14 |
| A_F | Normal | 10 | 0.24 | 0.81 | 0.58 | 0.24 | 0.78 | 0.58 | 0.26 | 0.18 |
| A_G | Normal | 10 | 2.41 | 7.87 | 0.58 | 2.41 | 7.87 | 0.58 | 0.02 | 0.35 |
| TOLLINC | Normal | 10 | -1.07 | -9.06 | 0.58 | -1.08 | -11.44 | 0.58 | 0.02 | 0.33 |
| TRTIME | Normal | 10 | -0.07 | -10.56 | 0.58 | -0.07 | -10.46 | 0.58 | 0.01 | 0.39 |

A lognormal distribution was also examined but did not produce good results.

```
Line search does not improve fn. Exit iterations. Status=3
Abnormal exit from iterations. If current results are shown
check convergence values shown below. This may not be a
solution value (especially if initial iterations stopped).
Gradient value: Tolerance= .1000D-05, current value= .8985D-05
Function chg. : Tolerance= .0000D+00, current value= .4547D-12
Parameters chg: Tolerance= .0000D+00, current value= .9372D-03
Smallest abs. parameter change from start value = .8830D-06
Note: At least one parameter did not leave start value.
```

FIGURE 3 Error Message Returned During the RPL Modeling Trials

These attempts to make the parameters vary across individual travelers (decision makers) did not achieve the desired results which would have justified the need to use an RPL model over the MNL model. As can be observed from Table 3, the standard deviations of the estimated random parameters were very small. Additionally, the adjusted $\bar{\rho}^2$ value and log-likelihood value of the estimated RPL model were very similar to that of the base model (standard MNL model). This may be due to the nature of the data used in this study which consists of four stated preference questions per individual traveler, and this may create difficulties in accounting for the observations drawn from the same individual due to correlated choice situations (6).

3.3 Nested Logit Models

The property of the standard logit model to exhibit independence from irrelevant alternatives acts as a restriction to capture all sources of correlation over alternatives. Generalized Extreme Value Models (GEV) overcomes this shortcoming by making the unobserved portions of utility for all alternatives jointly distributed as a generalized extreme value. Hence this class of models includes the logit model and a variety of other models. The most popular of them is the nested logit. Nested logit models are used when the global choice set can be divided into subsets with choices having similar attributes. These subsets are called nests, and the IIA holds within each nest but not across the nests. The ratio of the probabilities of selecting any two alternatives in a nest is independent of the attributes of all other alternatives in that nest, while the ratio of the probability of selecting any two alternatives in different nests can depend on the attributes of other alternatives in those two nests (4).

The nested logit model assumes that the vector of unobserved utility, ε_n , has a cumulative distribution (type of GEV distribution) which is given by Equation 5 [with notation from (1)]:

$$e^{\left(-\sum_{k=1}^K \left(\sum_{j \in B_k} e^{-\varepsilon_{nj} / \lambda_k} \right) \right)} \quad (5)$$

Where: j 's are the sets of alternatives which are partitioned into k non-overlapping subsets denoted by $B_1, B_2 \dots B_k$; and

λ_k is a measure of the degree of independence in an unobserved utility among the alternatives in nest k and is called as inclusive value or logsum parameter.

The inclusive value parameter ($\lambda_k \forall k$) should be between 0 and 1 so as to be consistent with the utility maximization and higher value of λ_k means greater independence and less correlation.

When λ_k is equal to 1 for all k , the nested logit becomes the same as a standard logit model (7,8).

The choice probability for individual n and alternative $i \in B_k$ is given by Equation 6 [with notation from (1)].

$$P_{ni} = \frac{e^{V_{ni}/\lambda_k} \cdot \left(\sum_{j \in B_k} e^{V_{nj}/\lambda_k} \right)^{\lambda_k - 1}}{\sum_{l=1}^K \left(\sum_{j \in B_l} e^{V_{nj}/\lambda_l} \right)^{\lambda_l}} \quad (6)$$

Apart from nested logit models, there are more complex models which can be used to overcome the limitations of the IIA property and are derived by different assumptions about the distribution of the vector of unobserved utility.

3.3.1 Decomposition of the Nested Logit Model

The nested logit probabilities can be expressed as a product of two simple logit probabilities, namely the probability that an individual selects the nest and the probability that individual selects the specific option given that the nest was selected. This concept of decomposition assists the interpretation of the model and allows estimation using standard logit software routines. This concept can be similarly expanded for nests having more than two levels as shown below (4).

3.3.1.1 Various terms used in nests having more than two levels

Trunk, limb, branch and twig are the components of a nested model and these terms are denoted as given below:

Trunk (l = number of trunks)

Limb (i = number of limbs)

Branch (j = number of branches)

Twig (k = number of twigs or the alternatives within a nest)

A conceptual diagram is shown in Figure 4 to explain above terms.

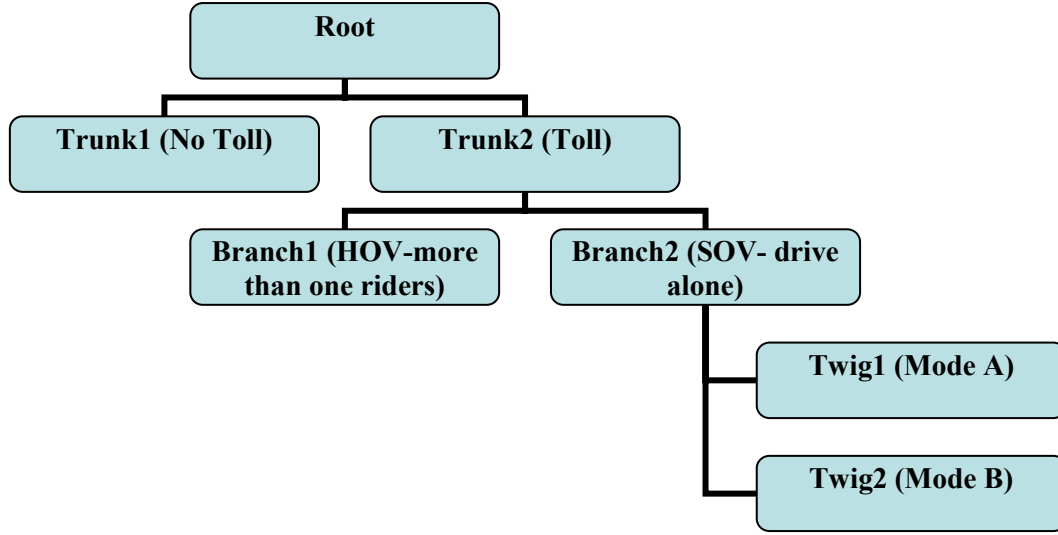


FIGURE 4 Conceptual Diagram of a Nested Logit Model

The (unconditional) probability that an individual selects a mode (alternative) is given by Equation 7:

$$P_{(k,j,i,l)} = P_{(k/j,i,l)} * P_{(j/i,l)} * P_{(i/l)} * P_{(l)} \quad (7)$$

Where $P(k/j,i,l)$ is the probability of alternative k in branch j , limb i , and trunk l which is given by Equation 8:

$$P_{(k/j,i,l)} = \frac{e^{\beta'x_{k/j,i,l}}}{\sum_{n/j,i,l} e^{\beta'x_{n/j,i,l}}} = \frac{e^{\beta'x_{k/j,i,l}}}{e^{J_{j/i,l}}} \quad (8)$$

Where J is Inclusive Value (or log sum) for branch j in limb i , trunk l , which is given by Equation 9:

$$J_{j/i,l} = \log \sum_{n/j,i,l} e^{\beta'x_{n/j,i,l}} \quad (9)$$

Similarly $P(j/i,l)$ is the conditional probability of choosing a particular branch in limb i , trunk l , given by Equation 10:

$$P_{(j/i,l)} = \frac{e^{(\alpha'y_{j/i,l} + \tau_{j/i,l} * J_{j/i,l})}}{\sum_{m/i,l} e^{(\alpha'y_{m/i,l} + \tau_{m/i,l} * J_{m/i,l})}} = \frac{e^{(\alpha'y_{j/i,l} + \tau_{j/i,l} * J_{j/i,l})}}{e^{I_{i/l}}} \quad (10)$$

The above equation $\alpha' y_{j/i,l}$ specifies the utility equation for branch j.

$I_{i/l}$ in the above equation is the inclusive value for limb i in trunk l, given by Equation 11:

$$I_{i/l} = \log \sum_{m/i,l} e^{(\alpha' y_{j/i,l} + \tau_{j/i,l} * J_{j/i,l})} \quad (11)$$

Similarly $P(i/l)$ is the conditional probability of choosing a particular limb i in trunk l, given by Equation 12:

$$P_{(i/l)} = \frac{e^{(\gamma' z_{i/l} + \sigma_{i/l} * I_{i/l})}}{\sum_{m/l} e^{(\gamma' z_{m/l} + \sigma_{m/l} * I_{m/l})}} = \frac{e^{(\gamma' z_{i/l} + \sigma_{i/l} * I_{i/l})}}{e^{H_l}} \quad (12)$$

In the above equation, $\gamma' z_{i/l}$ specifies the utility equation for limb i;

$\sigma_{i/l}$ is the inclusive value parameter for limb i within trunk l (it is equal to one if there is only one limb); and

H_l is the inclusive value for trunk l, given by Equation 13.

The probability of choosing the trunk l is given by Equation 14.

$$H_l = \log \sum_{m/l} e^{(\gamma' z_{m/l} + \sigma_{m/l} * I_{m/l})} \quad (13)$$

$$P(l) = \frac{e^{(\theta' w_l + \phi_l * H_l)}}{\sum_m e^{(\theta' w_m + \phi_m * H_m)}} \quad (14)$$

ϕ_l is the inclusive value parameter for trunk l (it is equal to one if there is only one trunk).

3.3.2 Typical LIMDEP Command for Nested Logit Model

A typical nested logit model command in LIMDEP along with its conceptual diagram is given in Figure 5.

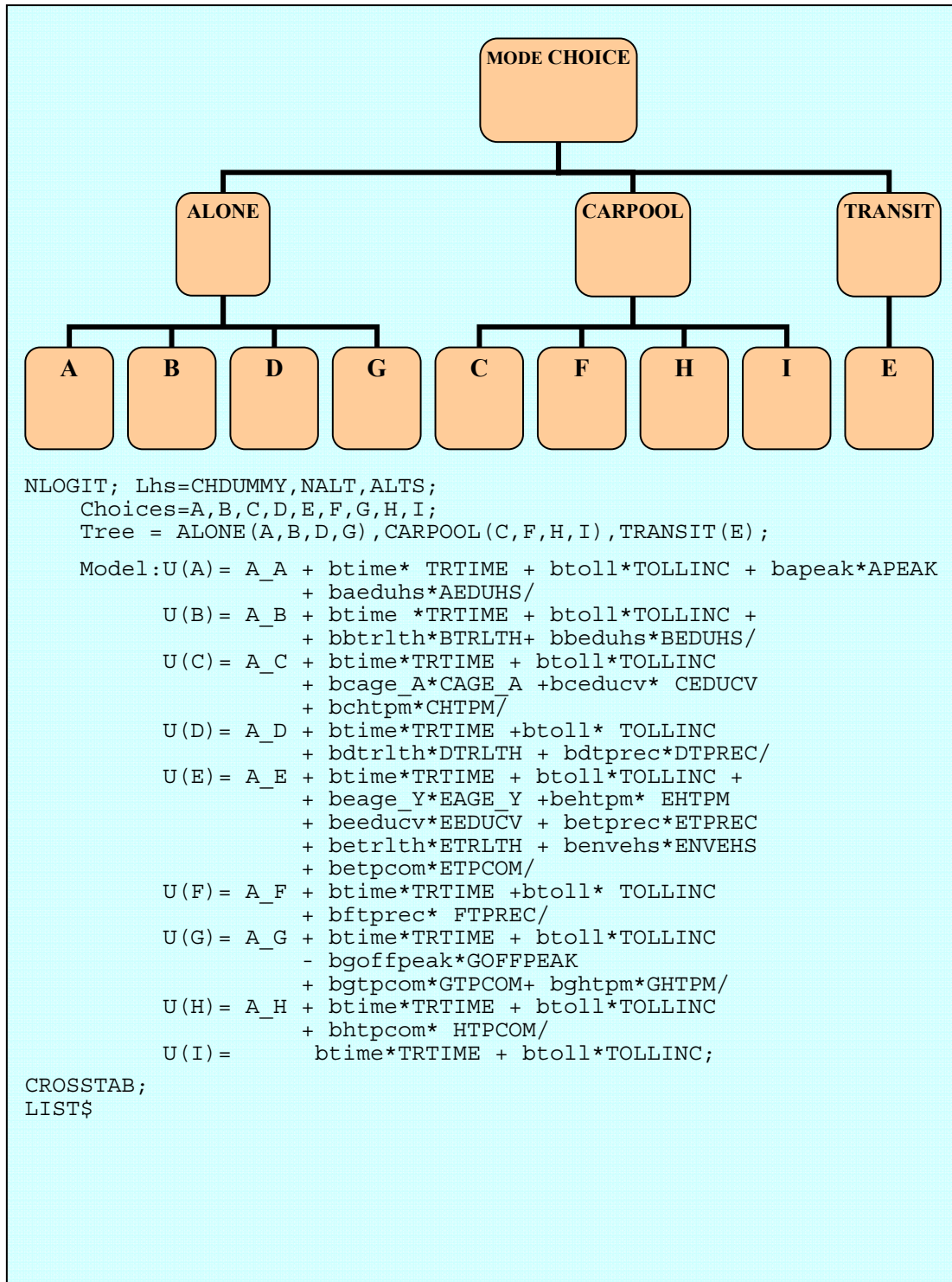


FIGURE 5 A typical nested logit model command in LIMDEP

In the above command, btime, btoll, etc. are the names given to coefficients, and TRTIME, TOLLINC, etc. are the variables in the utility equations.

The command “Tree = SOV(A,B,D,G),HOV(C,F,H,I),TRANSIT(E);” specifies the exact nature of the nest which is a two level nest. The alternatives in a twig are included in round brackets (). While it is optional to give names (SOV, HOV, etc.) to branches, they improve the understanding of the nest. Similarly a command for a four-level nest would be:

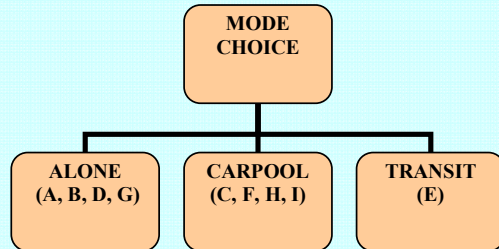
```
Tree = SOV{ SPEAK [SPFREE (G) , SPTOLLED (D) ] , SOPEAK [SOFREE (A) , SOTOLLED (B) ] } ,  
        HOV{ HPEAK [HPFREE (C , I) , HPTOLLED (H) ] , HOPEAK [HOFREE (F) ] , TRANSIT [E] }
```

Branches are included in square brackets [], while limbs are included in { }.

3.3.3 *Different Nest Specifications Tested*

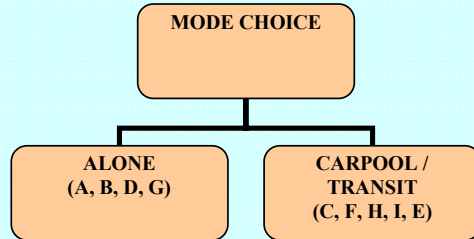
Different nest specifications as shown in Figure 6a through Figure 6g were examined using Katy Freeway travelers’ data. Note the different mode choices are shown inside the lowermost level for simplicity.

NEST-1



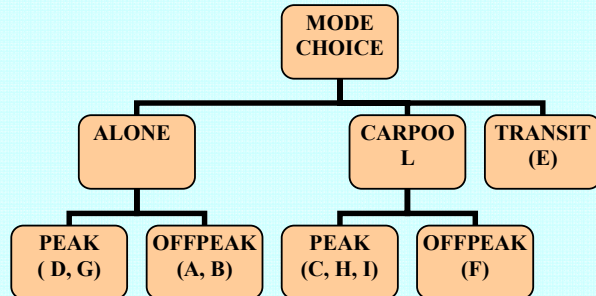
Tree = ALONE (A, B, D, G) , CARPOOL (C, F, H, I) , TRANSIT (E) ;

NEST-2



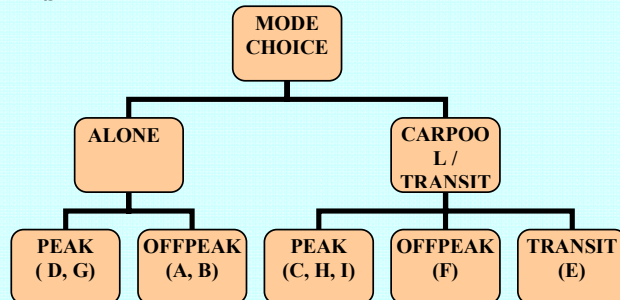
Tree = ALONE (A, B, D, G) , CARPLnTRAN (C, F, H, I, E) ;

NEST-3



Tree = ALONE [ALPEAK (D, G) , ALOFPEAK (A, B)] ,
CARPOOL [CPPEAK (C, H, I) , CPOFPEAK (F)] ,
TRANSIT [E] ;

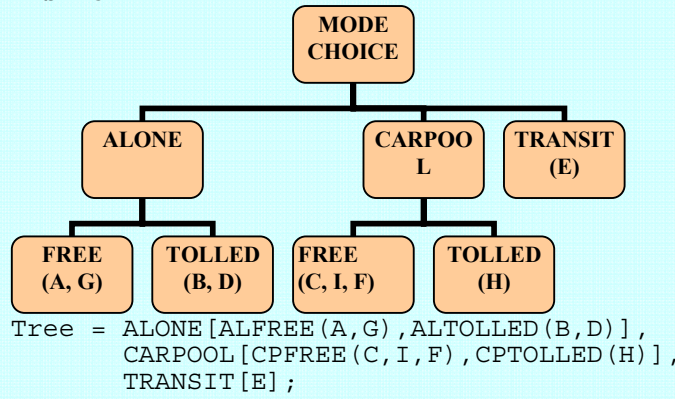
NEST-4



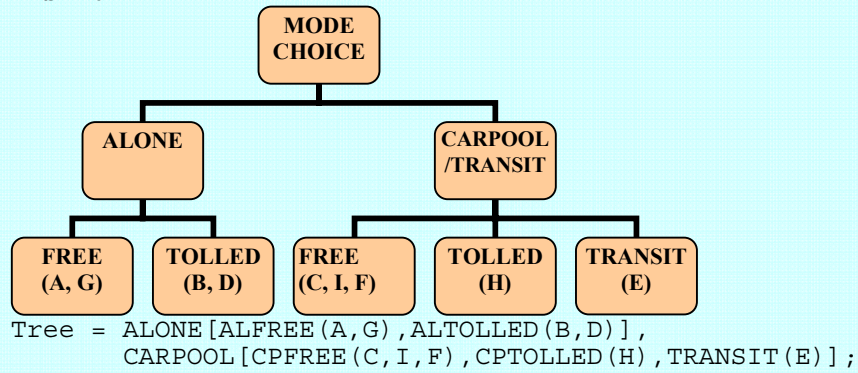
Tree = ALONE [ALPEAK (D, G) , ALOFPEAK (A, B)] ,
CARPOOL [CPPEAK (C, H, I) , CPOFPEAK (F) , TRANSIT (E)] ;

FIGURE 6a Different Nest Specifications Used for the Study

NEST - 5



NEST - 6



NEST - 7

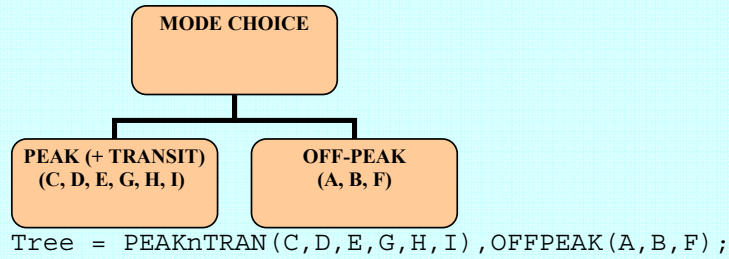
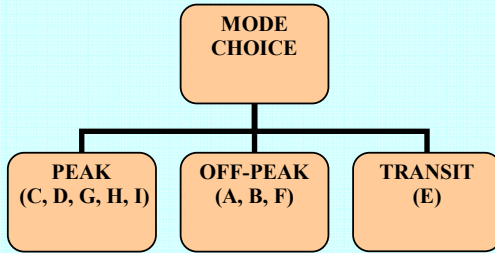


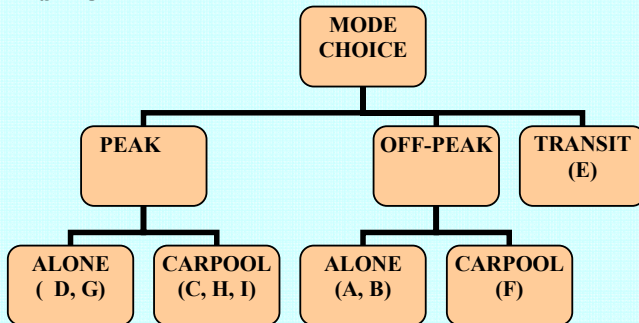
FIGURE 6b Different Nest Specifications Used for the Study

NEST-8



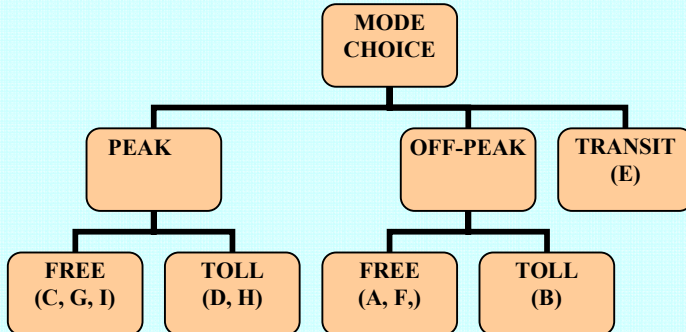
Tree = PEAK (C, D, G, H, I) , OFFPEAK (A, B, F) , TRANSIT (E) ;

NEST-9



Tree = PEAK [PALONE (D, G) , PCARPL (C, H, I)] ,
OFFPEAK [OFALONE (A, B) , OFCARPL (F)] ,
TRANSIT [E] ;

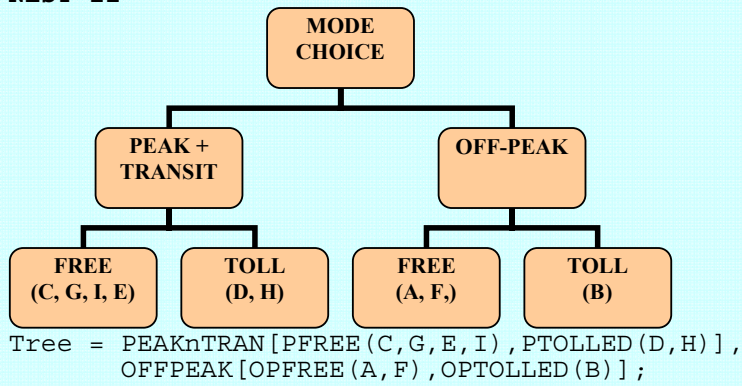
NEST-10



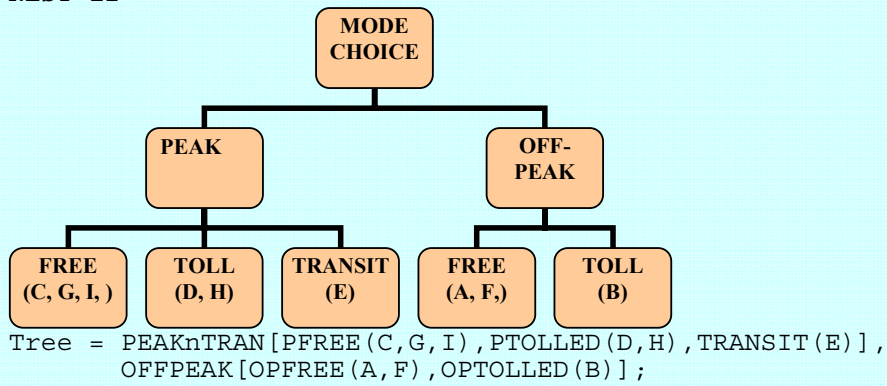
Tree = PEAK [PFREE (C, G, I) , PTOLLED (D, H)] ,
OFFPEAK [OPFREE (A, F) , OPTOLLED (B)] ,
TRANSIT [E] ;

FIGURE 6c Different Nest Specifications Used for the Study

NEST-11



NEST-12



NEST-13

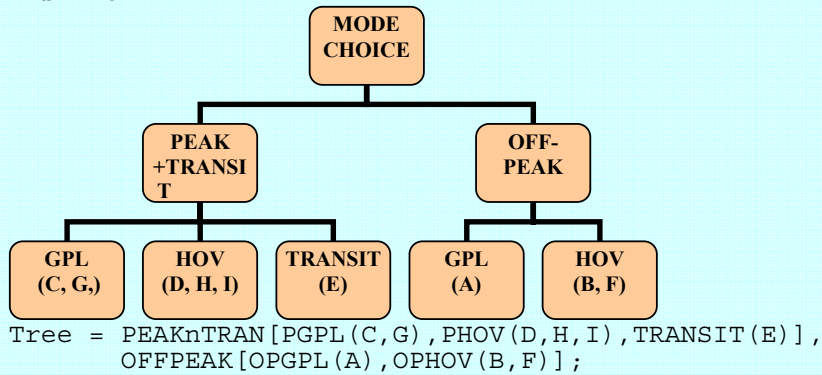
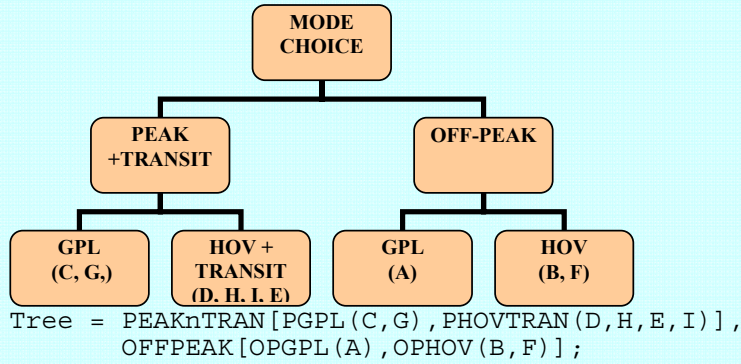
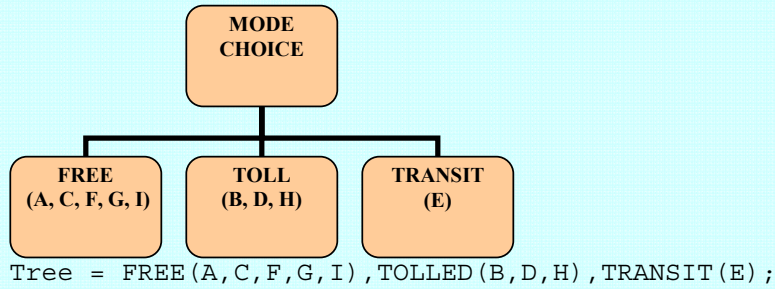


FIGURE 6d Different Nest Specifications Used for the Study

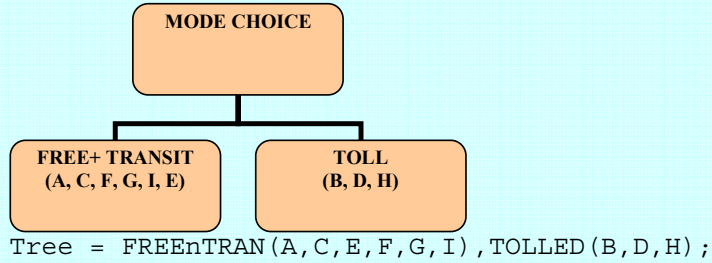
NEST-14



NEST-15



NEST-16



NEST-17

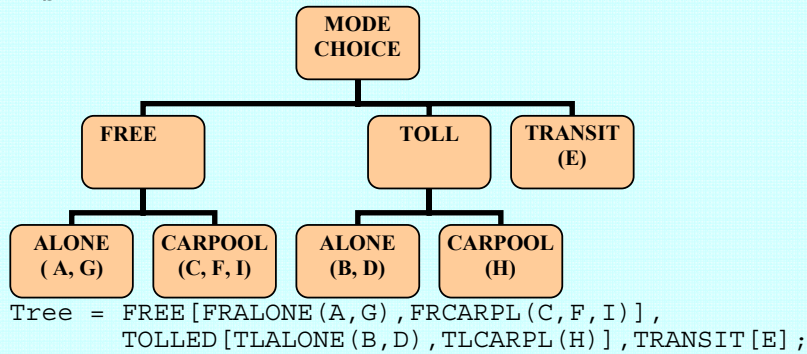
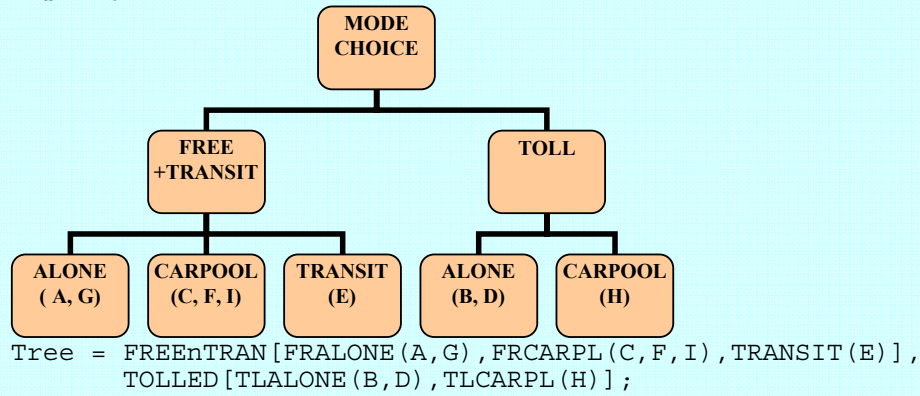
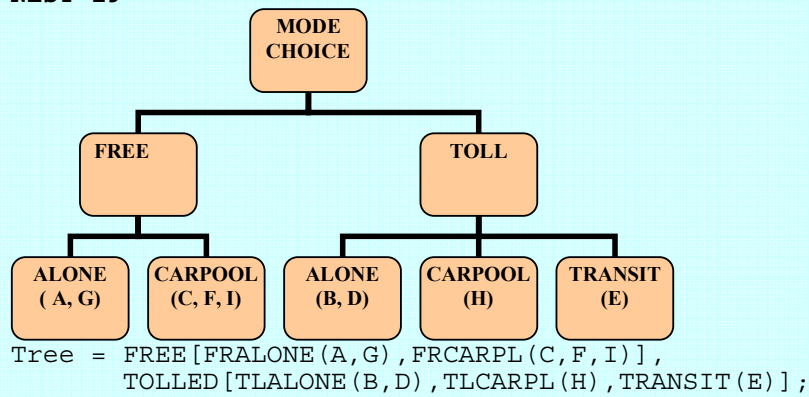


FIGURE 6e Different Nest Specifications Used for the Study

NEST-18



NEST-19



NEST-20

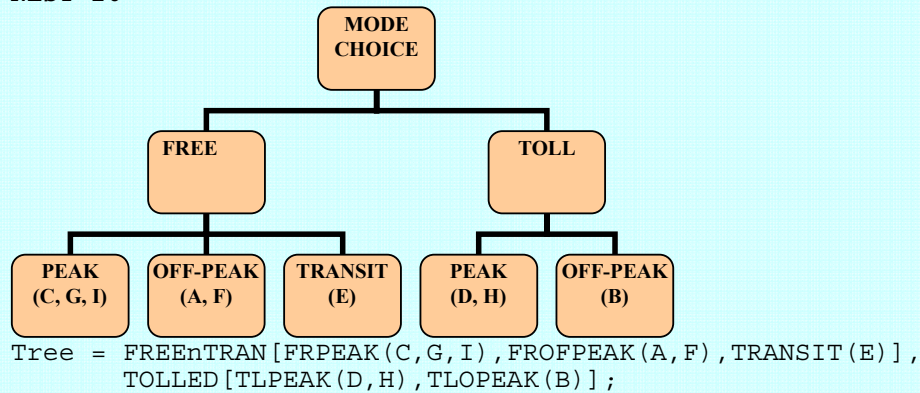


FIGURE 6f Different Nest Specifications Used for the Study

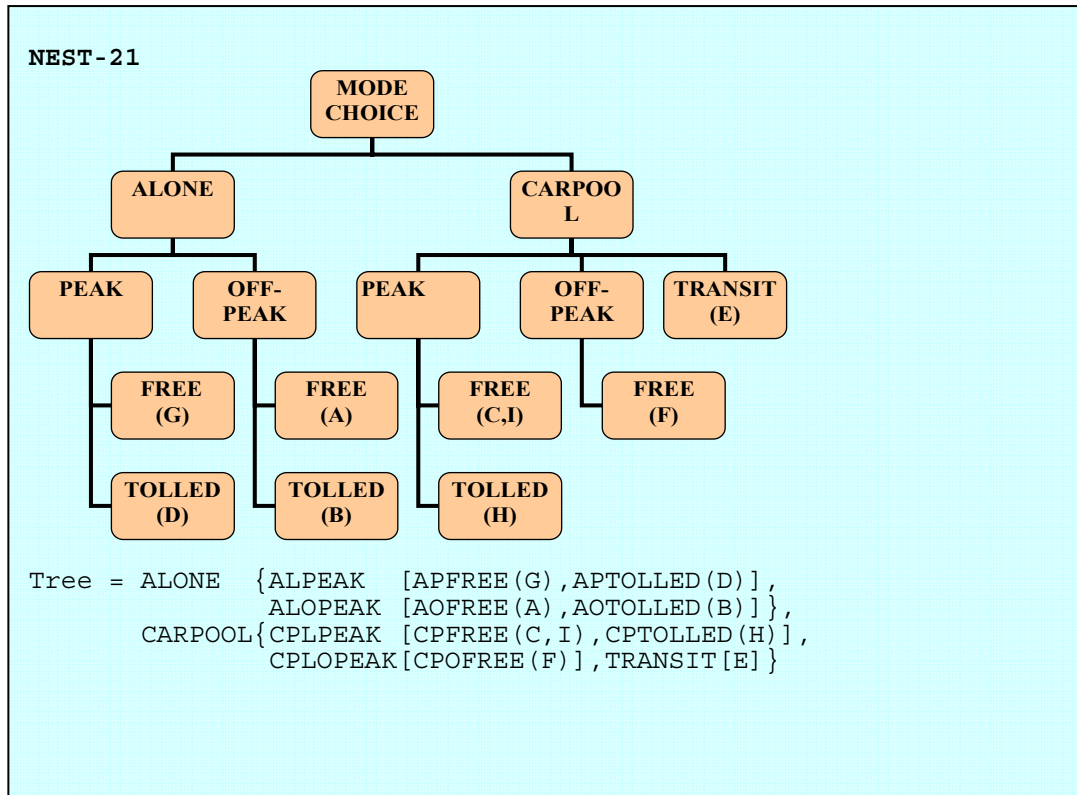


FIGURE 6g Different Nest Specifications Used for the Study

3.3.4 Selection of the Best Nest Specification

The nested logit models specified in Figure 6a through 6g were examined for the log-likelihood value, $\bar{\rho}^2$ value, and values of the inclusive value parameters to select the best models among them. An inclusive value parameter should be between zero and one; it should be strictly equal to one for only nest containing only one alternative. Apart from these preliminary criteria the Likelihood Ratio (LR) test was carried out for the best models to test if they are better than the simple multinomial logit model. Sometimes, the t-test is also carried out to test statistically if the inclusive value parameters are different than one. The likelihood ratio test and t-test are described below followed by the results of the Likelihood Ratio test obtained for the Katy Freeway traveler models.

3.3.4.1 The Likelihood Ratio Test for a Nested Logit Model

When we are adopting a nested logit model we are in a way rejecting the simpler MNL. Hence in order to justify this decision, we can use Likelihood Ratio Test. We start with the hypothesis that the MNL is the true or correct model which is equivalent to stating that all inclusive value

parameters in the nested model are equal to one (8). Hence we are forcing the number of restrictions equal to the number of IV parameters. We reject the null hypothesis that the MNL model is the correct model if the calculated value is greater than the test or critical value for the distribution as:

$$-2*[LL_{MNL} - LL_{NL}] \geq \chi_n^2,$$

Where n is the number of restrictions or number of IV parameters.

A typical command for carrying out the LR test using LIMDEP is given in Figure 7. Where, LR and LU are the names given to the variables which store log-likelihood values for restricted (MNL) and unrestricted (Nested) models. C is the variable which stores test statistics.

```
NLOGIT;Lhs=CHDUMMY,NALT,ALTS;          /* Restricted Model*/
      Choices=A,B,C,D,E,F,G,H,I;
Model:
  U(A)= A_A +btime* TRTIME +.... /
      U(B)= A_B + btime *TRTIME +.. /
      U(C)= A_C + btime*TRTIME + ... /
      U(D)= A_D + btime*TRTIME +... /
      U(E)= A_E + btime*TRTIME + ... /
      U(F)= A_F + btime*TRTIME +.... /
      U(G)= A_G + btime*TRTIME +.... /
      U(H)= A_H + btime*TRTIME +.... /
      U(I)= btime*TRTIME + btoll*TOLLINC$

CALC ; LR = LogL $

NLOGIT;Lhs=CHDUMMY,NALT,ALTS;          /*Unrestricted Model*/
      Choices=A,B,C,D,E,F,G,H,I;

Tree = NoToll (A,C,E,F,G,I) ,Toll (B,D,H) ;
Model:
  U(A)= A_A +btime* TRTIME +.... /
      U(B)= A_B + btime *TRTIME +.. /
      U(C)= A_C + btime*TRTIME + ... /
      U(D)= A_D + btime*TRTIME +... /
      U(E)= A_E + btime*TRTIME + ... /
      U(F)= A_F + btime*TRTIME +.... /
      U(G)= A_G + btime*TRTIME +.... /
      U(H)= A_H + btime*TRTIME +.... /
      U(I)= btime*TRTIME + btoll*TOLLINC$

CALC ; LU = LogL ; C = -2*(LR - LU) $          /* C is test statistic*/
CALC ; CC = CTB(.95,2) $          /*CC is Critical chi-square value*/
```

FIGURE 7 Commands for Likelihood Ratio Test in LIMDEP

The values of C and CC can be obtained in the folder scalars in variables as shown in the Figure 8. In this case we must reject the null hypothesis since the Chi-square value exceeds the critical Chi-square value. Therefore, the nested model is the better model in this case.

```
--> CALC ; LU = LogL ; C = -2*(LR - LU) $
--> CALC ; CC = CTB(.95,2) $
--> CALC;C=9.744223$
--> CALC;CC=5.991465$
```

FIGURE 8 An Output for LR Test Using LIMDEP

3.3.4.2 t-test for the inclusive value parameters

A second test which can also be used is the t-statistic for testing the hypothesis that each of the inclusive value parameters is equal to one. For each case in which this hypothesis is not rejected the corresponding branch of the tree can be eliminated and replaced by the alternatives at the next level. Hence the null hypothesis is:

$$\tau_k = 1$$

Where τ_k is the estimate for the k^{th} inclusive value parameter which can be interpreted as shown in Table 4. Values of τ_k for different nests are estimated along with the coefficients in the utility function and are returned in the LIMDEP output along with the coefficients.

The test statistic is given by Equation 15:

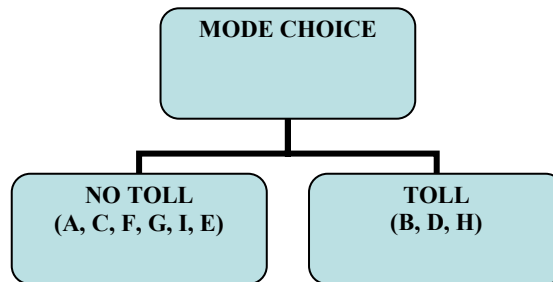
$$t - \text{statistic} = \frac{\tau_k - 1}{S_k} \quad (15)$$

Where S_k is the standard error for the k^{th} inclusive value parameter, and 1 is the hypothesized value of the parameter.

TABLE 4 Interpretation of Value of τ_k (8)

| Value Range for τ_k | Interpretation |
|--------------------------|--|
| $\tau_k > 1$ | Not consistent with the theoretical derivation. Therefore, we reject nested logit model. |
| $\tau_k = 1$ | Implies zero correlation among mode pairs in the nest so the NL model collapses to the MNL model. |
| $0 < \tau_k < 1$ | Implies non-zero correlation among mode pairs. This range of values is appropriate for the nested logit model. |
| $\tau_k = 0$ | Implies perfect correlation between pairs of alternatives in the nest. That is, the choice between the nested alternatives, conditional on the nest, is deterministic. |
| $\tau_k < 0$ | Not consistent with the theoretical derivation. Therefore we reject the nested logit model. |

Based on the likelihood ratio test and t-test for inclusive value parameters, Nest-16 (Figure 9) was found to be the best for our dataset. The summary calculations and results for these tests are given in Table 5 and Table 6.

**FIGURE 9 Nesting Details for Nest ID-16****TABLE 5 LR Test for Katy Freeway Traveler Model**

| Nest ID | LR | LU | DOF | C | Crit. C | Conclusion |
|---------|---------|---------|-----|------|---------|--|
| Nest-16 | -1683.3 | -1678.4 | 3 | 9.74 | 7.81 | Reject null: Accept Nested (at 95% confidence level) |

TABLE 6 t-test for Inclusive Value Parameters of Estimated NL Model

| Nest ID | Branch/ link name | IVP | SE | DOF | tstat | tcrit | Conclusion |
|---------|-------------------------|-------|-------|-----|-------|-------|--------------------------------------|
| Nest-16 | No Toll | 0.665 | 0.108 | 3 | -3.11 | 1.96 | Reject null: Values different from 1 |
| | Toll | 0.737 | 0.103 | 3 | -2.55 | 1.96 | Reject null: Values different from 1 |

The estimated nested logit model is given in Table 7. The variable names were explained in the MNL model section. The table also contains the estimated inclusive value parameters for the nests.

TABLE 7 Nested Logit Model for Katy Freeway Travelers (Nest ID-16)

| | Coeff. | Std.Err. | t-ratio | P-value |
|-----------------------------------|-------------|----------|---------|---------|
| A_A | 2.96 | 0.29 | 10.38 | 0.00 |
| BTIME | -0.09 | 0.01 | -8.62 | 0.00 |
| BTOLL | -13.43 | 0.20 | -6.79 | 0.00 |
| BAPEAK | -0.34 | 0.20 | -1.70 | 0.09 |
| BAEDUHS | -0.70 | 0.53 | -1.34 | 0.18 |
| A_B | 1.17 | 0.39 | 2.97 | 0.00 |
| BBTRLTH | 0.01 | 0.00 | 2.17 | 0.03 |
| BBEDUHS | -1.62 | 0.72 | -2.24 | 0.03 |
| A_C | -0.52 | 0.46 | -1.13 | 0.26 |
| BCAGE_A | 2.87 | 0.57 | 5.00 | 0.00 |
| BCEDUCV | 2.26 | 0.41 | 5.51 | 0.00 |
| BCHTPM | 1.03 | 0.43 | 2.43 | 0.02 |
| A_D | 0.31 | 0.41 | 0.76 | 0.45 |
| BDTRLTH | 0.02 | 0.01 | 3.93 | 0.00 |
| BDTPREC | -1.46 | 0.55 | -2.68 | 0.01 |
| A_E | -5.56 | 1.29 | -4.30 | 0.00 |
| BETRLTH | 0.02 | 0.01 | 2.29 | 0.02 |
| BEAGE_Y | 2.85 | 0.50 | 5.71 | 0.00 |
| BEHTPM | 1.63 | 0.39 | 4.23 | 0.00 |
| BENVEHS | 0.43 | 0.18 | 2.43 | 0.02 |
| BEEDUCV | 1.02 | 0.40 | 2.54 | 0.01 |
| BETPCOM | 2.93 | 1.07 | 2.73 | 0.01 |
| BETPREC | 3.22 | 1.20 | 2.69 | 0.01 |
| A_F | -0.02 | 0.31 | -0.07 | 0.95 |
| BFTPREC | -1.24 | 0.65 | -1.91 | 0.06 |
| A_G | 2.41 | 0.33 | 7.32 | 0.00 |
| BGPEAK | 0.36 | 0.16 | 2.17 | 0.03 |
| BGTPCOM | 0.87 | 0.20 | 4.30 | 0.00 |
| BGHTPM | 0.51 | 0.19 | 2.73 | 0.01 |
| A_H | -0.81 | 0.61 | -1.31 | 0.19 |
| BHTPCOM | 1.71 | 0.58 | 2.94 | 0.00 |
| Inclusive Value Parameters | | | | |
| NO TOLL | 0.67 | 0.11 | 6.22 | 0.00 |
| TOLL | 0.73 | 0.10 | 7.15 | 0.00 |

The value of travel time savings using above model was estimated to be around 40% of hourly wage rate (\$/hr) using the coefficient of variables tollinc (toll/approx. hourly wage rate in \$/hr)

and trtime (in minutes). The details about exact formula used are given in section 4.0 of this report. The nested logit model had similar log-likelihood value and also approximately same percentage correct mode choice prediction (49%) as that of MNL. Also, the VTTS obtained by both NL and MNL were comparable, hence the simple multinomial logit model was preferred for further analysis as this study focuses on VTTS estimation.

3.4 Model Estimation Using Genetic Algorithms

Genetic algorithms are search techniques based on evolution and are frequently used for optimization problems. They are less susceptible to returning sub-optimal solutions (such as local optima) than gradient search methods, but they tend to be computationally more burdensome.

A genetic algorithm initially creates a population of chromosomes which represents the first generation. These chromosomes are solutions to an optimization problem represented as the fitness function. Based on their fitness the chromosomes are selected (better fitness indicates better chances of selection) for reproduction by mixing the genes of two parent chromosomes to create a new child chromosome. The process of reproduction involves mutation and crossover in order to add some diversity in the solutions while maintaining their convergence. The second generation is formed by combining the parent and child chromosomes so that the population size is constant. This process of evolution towards a better solution is continued till the specified stopping criterion is achieved. The stopping criterion can be based on the number of generations, consistency of the fitness function value or time.

It is possible that genetic algorithms may be used to estimate parameters in a mode choice model, but only one example of anything similar to this was found in the literature (9). This research assumed that the set of parameters in the utility equation could be considered chromosomes, and a function such as the log-likelihood function could be the fitness function. Depending on the specification of the fitness function, this method could help to avoid making an assumption regarding the distribution of the error terms in the utility equations and deriving a theoretical equation for choice probabilities unlike the MNL and NL models. In this way it might be possible to overcome some of the limitations such as IIA assumption in MNL models.

This study examined the applicability of GAs in mode choice modeling in a complex mode choice situation and in estimating traveler's value of travel time savings. The MNL model of mode choice on the Katy Freeway described earlier in this study was the base model. This model was compared to the NL model and model estimated by using GA. The best GA model was chosen by comparing the models obtained when varying the options/elements of genetic algorithms such as population size, initial population, selection procedure, and mutation rate.

The variables in GA models (attributes in the utility functions) were kept the same as that of MNL and NL so as to facilitate the comparison of the models (coefficients) estimated. The models were compared based on measures such as log-likelihood values, percent of correct mode choice predictions, number of correct (expected) signs, and even reasonable estimated value of travel time savings.

3.4.1 Genetic Algorithm Estimation

MATLAB's (10) genetic algorithm tool (gatool) was used for the estimation of the model. A typical genetic algorithm tool window in MATLAB is shown in Figure 10. Each trial took approximately 45 minutes on a computer with a processor speed of 1.7 GHz because of the large number of parameters (genes) and the population size. This compares to the nearly instantaneous results (less than 15 seconds) for the different logit models estimated using LIMDEP.

3.4.1.1 Fitness Functions

To estimate the coefficients in the systematic utility equations using a genetic algorithm, a fitness function was developed. This fitness function was optimized based on values of the model coefficients. Hence an individual in the population (a chromosome) corresponds to a solution which was a set of 31 variable values (genes) including the alternative specific constants. One fitness function examined in this research was simply the number of correct mode choice predictions. As defined above, a correct prediction was made when the mode that had the maximum utility for the given set of coefficients was also the mode selected by the respondent. Thus the GA found the set of coefficients which maximized the total number of correct

predictions. As stated earlier, variables used in the utility equations for this genetic algorithm model were the same as those in the logit models. This fitness function was labeled Predmax for convenience of documentation. Use of this fitness function (Predmax) did not require any assumptions regarding the distribution of error terms in the utility equations.

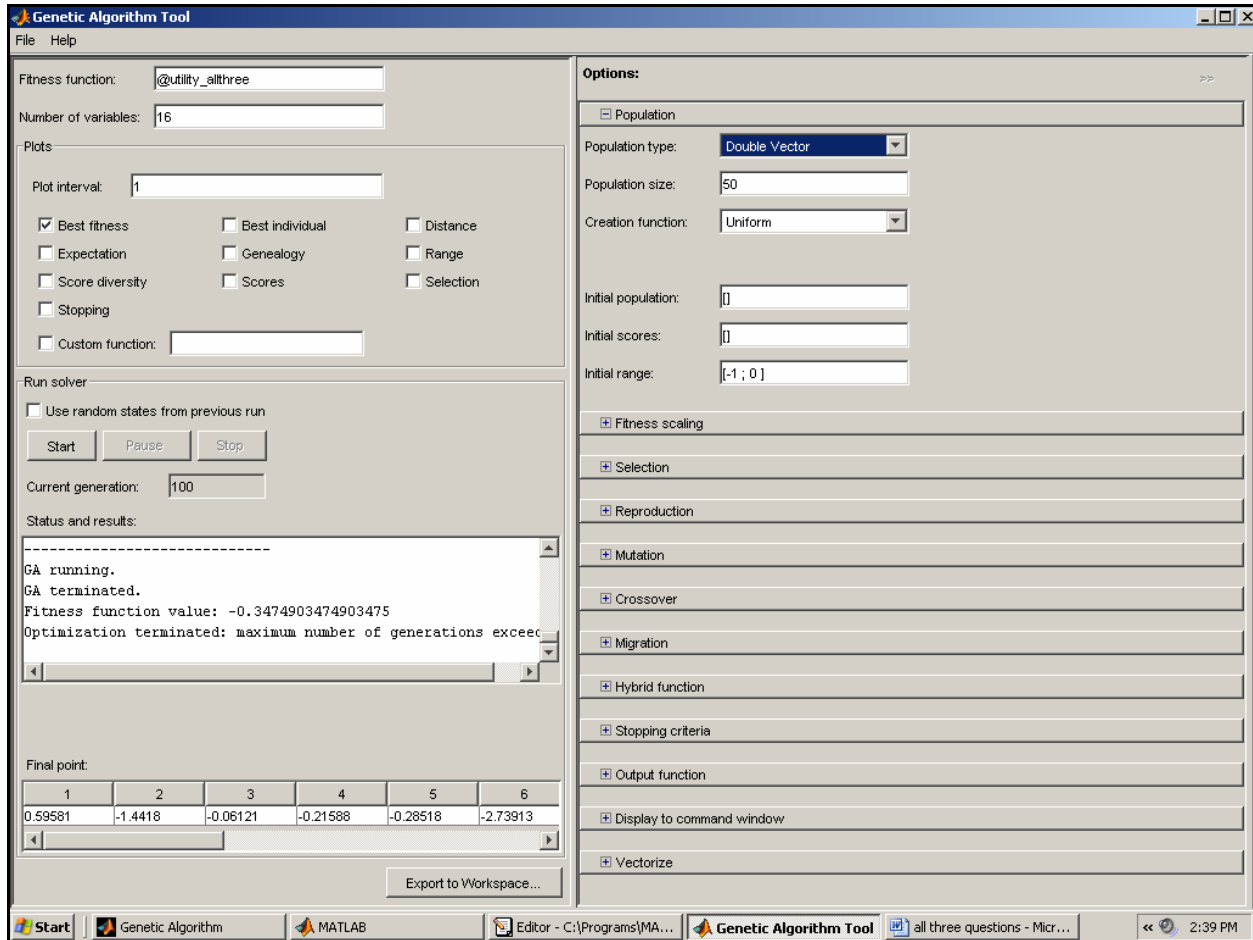


FIGURE 10 A Typical Genetic Algorithm Tool Window in MATLAB

Another fitness function was developed that was equal to the log-likelihood function (shown in Equation 16). The GA then searched for the solution (set of β values) which maximized the log-likelihood value. This fitness function was labeled LLmax for convenience of documentation.

$$LL(\beta) = \sum_{n=1}^N \log(P_n(m^* | \beta)) \quad (16)$$

Where $P_n(m^* | \beta)$ = Probability of person n choosing the observed mode m^* , given the model parameters β and is calculated using Equation 1; and
 LL (β) = Log-likelihood value for the given β values.

The final fitness function examined in this research was the likelihood function (see Equation 17). This fitness function was labeled as Lmax for convenience of documentation.

$$L(\beta) = \sum_{n=1}^N P_n(m^* | \beta) \quad (17)$$

Where L (β) = Likelihood value for given β values.

The Matlab code for these fitness functions is given in Figure 11. The Matlab code as shown in the figure has the option of selecting any one of the above fitness functions by adding “%” before the command lines corresponding to the other two fitness functions.

```
function val_out=utl_max(x)
load k1; % loads the data file%
k1(any(isnan(k1)'), :) = []; % removes rows containing missing values%
len = length(k1); % counts the number of rows/observations%
ntravl =(length(k1))/4; % counts the number of individuals (questions)%
sumA=zeros(ntravl,1);
ut= NaN(ntravl,9);% Initiates the matrix of utilities%
predch= zeros(ntravl,1); % initiates a vector that will store the
                        predicted choice of an individual%

k=1;
i=1;
LL=0;
while (k<len)
    for j=k:(k+3)
        if (k1(j,12))==1
            ut(i,1)=x(1)+x(9)*k1(j,1)+x(10)*k1(j,2)+ x(11)*k1(j,3)
                    +x(12)*k1(j,4);
        else if (k1(j,12))==2
            ut(i,2)=x(2)+x(9)*k1(j,1)+x(10)*k1(j,2)+ x(13)*k1(j,5)
                    + x(14)*k1(j,4) ;
        else if (k1(j,12))==3
            ut(i,3)= x(3)+x(9)*k1(j,1)+x(10)*k1(j,2)
                    + x(15)*k1(j,6)+x(16)*k1(j,7)+x(17)*k1(j,8);
        else if(k1(j,12))==4
            ut(i,4)= x(4)+x(9)*k1(j,1)+ x(10)*k1(j,2)
                    + x(18)*k1(j,5)+x(19)*k1(j,9);
        else if(k1(j,12))==5
            ut(i,5)= x(5)+x(9)*k1(j,1)+x(10)*k1(j,2)
                    +x(20)*k1(j,5)+x(21)*k1(j,6)
                    +x(22)*k1(j,8)+x(23)*k1(j,10)
                    +x(24)*k1(j,7)+x(25)*k1(j,11)
                    +x(26)*k1(j,9);
        else if(k1(j,12))==6
            ut(i,6)= x(6)+x(9)*k1(j,1)+x(10)*k1(j,2)
                    +x(27)*k1(j,9);
        else if(k1(j,12))==7
            ut(i,7)= x(7)+x(9)*k1(j,1)
                    +x(10)*k1(j,2) +x(28)*k1(j,11)
                    +x(29)*k1(j,8)+x(30)*k1(j,3);
        else if(k1(j,12))==8
            ut(i,8)= x(8)+x(9)*k1(j,1)
                    +x(10)*k1(j,2)+x(31)*k1(j,11);
        else if(k1(j,12))==9
            ut(i,9)= x(9)*k1(j,1)+x(10)*k1(j,2);
            end;
        end;
    end;
end;
end;
end;
end;
end;
end;
end;
```

FIGURE 11 Fitness Functions Used for GA estimation of Discrete Choice Models

```

q=zeros(1,9);

q =[ut(i,1),ut(i,2),ut(i,3),ut(i,4),ut(i,5),ut(i,6),ut(i,7),
    ut(i,8),ut(i,9)]; %makes a row in utility matrix corresponding
                        to an individual%

q = q(~isnan(q)); %removes the NaN values and makes q 1x4 vector%

Prb=exp(ut(i,k1(k,13)))/(exp(q(1))+exp(q(2))+exp(q(3))+exp(q(4)));
    %calculates prob.of the mode chosen%

% LL=LL+ Prb; %used only if the fitness function is based on
              likelihood value%

LL=LL+ log(Prb); %Used only if the fitness function is based on the
                 Log-likelihood value %

[mxutl,mxid]=nanmax(q); %finds the maximum and its position%
predch([i])=mxid;
if(predch(i)==k1(k,13))
    sumA(i,1)=1;
end; % Used only if the fitness function is Predmax%

k=k+4;
i=i+1;
end;

val_out=-LL; %for fitness functions LLmax or Lmax%

% val_out=-100*sum(sumA)/ntravl; %for fitness function Predmax%

```

FIGURE 11 Fitness Functions Used for GA estimation of Discrete Choice Models (continued)

Since the fitness functions LLmax and Lmax use Equation 1, one must assume that the error terms in the utility equation are Gumbel-distributed with a mean of zero. In this way the estimated models were the multinomial logit models. The difference between a model estimated using GAs and any other software such as LIMDEP is that the GAs do not use optimization techniques based on the Newton-Raphson method (or Davidon/Fletcher/Powell algorithm) for estimation which estimates the parameters by derivatives. Thus, GAs do not force the predicted and observed mode shares to match through selection of the alternative-specific parameter values (9,10).

3.4.2 Setting the Genetic Algorithm for Mode Choice Modeling

When estimating the models using GAs, several options/settings were available for the component processes of the genetic algorithm. These included the population size, the selection function, mutation rate, crossover function, stopping criteria, etc. Many trials were undertaken to find a combination of options/settings which maximized the fitness function with a good degree of consistency and which converged to a solution in a reasonable time. The same random number states were used during these trials to help with the comparison of the results of the trials and attributing any difference in outcomes primarily to the setting/option altered (the option of using same random number states from the previous run is shown on the gatool interface just above the start tab as shown in Figure 10). Details of these options are listed in Table 8 (10).

TABLE 8 Settings and Methods Used for the Genetic Algorithm

| Genetic Algorithm Element | Chosen Method or Setting | Description of chosen Method or Setting |
|-----------------------------------|---|--|
| Population Size | 65 | Maintains 65 individuals in each generation |
| Method for Fitness Scaling | Rank | Fitness scaling is done based on the rank of the individual so as to convert the raw fitness scores to values in a range that is suitable for the selection function. |
| Selection Method | Remainder | Parents are assigned deterministically from the integer part of each individual's scaled value and then roulette selection is used on the remaining fractional part as a part of parent selection for next generation. |
| Reproduction Option (Elite Count) | 2 | Specifies the number of individuals that are guaranteed to survive to the next generation. |
| Mutation Option | Gaussian | Adds a random number taken from a Gaussian distribution with mean 0 to each entry of the parent vector so as to make small random changes in the individuals in the population to create mutation children. |
| Stopping Criteria | Number of generations with no increase in fitness = 100 | The algorithm stops if there is no improvement in the best fitness value for 100 generations. |

In order to obtain consistent estimation it was necessary to specify an appropriate population size. Too large a population would greatly increase the time required for estimation, while too small a population size may not result in finding the optimum solution. It was necessary to have the population size equal to or larger than the number of parameters (9). As per the available literature, a population size of approximately twice the number of parameters is likely to yield good results (6). In this case a population size of more than 75 was found to increase the optimization time significantly and achieved little benefit over a population size closer to twice the number of parameters; hence the population size of 65 was used.

The selection function specifies how the genetic algorithm chooses parents for the next generation. The options for the selection function were: stochastic uniform, remainder, uniform, roulette, tournament, etc., and the function remainder was found to maximize the fitness function with a good degree of consistency and to converge to a solution in a reasonable time.

Mutation options specify how the genetic algorithm makes small random changes to create children. The options for mutation function were gaussian and uniform, and the mutation function gaussian had maximum fitness function value and converged to a solution in reasonable time.

The combination of two parents to form a crossover child for the next generation is decided by the crossover function for which the options were: scattered, single point, two point, etc. The option of scattered crossover function was found to yield maximum value for the fitness function.

Specifying proper stopping criteria will ensure that the search continued until optimization achieved consistent solutions. Due to more time required for iterations in the optimization, the stopping criteria such as stall time limit (the algorithm stops if the best fitness value is not improved for an interval of time in seconds) are easily achieved and hence the optimization may stop earlier than necessary. Hence the stopping criterion was decided in terms of number of stall generations (the generations for which there is no improvement in the best fitness value). These selected options and settings may not have been optimal, but they were sufficient to maximize

the fitness functions with consistency and speed. The settings and methods chosen are listed in Table 7 as discussed before.

3.5 Comparison of Logit Models and GA Estimated Models

The estimated models obtained by using all of the above fitness functions (LLmax, Lmax, and Predmax) are shown in Table 9. In order to study the prediction success, crosstab matrices were generated for these fitness functions (Table 10 through Table 12). A separate m-file was created in MATLAB for calculating the cross-tab matrix using the estimated coefficients. This file had all the commands from the fitness function with some extra commands for generating the cross-tab matrix. These extra commands are given in Figure 12. In order to generate this crosstab, instead of using an all-or-nothing approach, a probabilistic approach was used. In the case of the all-or-nothing approach an individual is assumed to choose a mode which has the highest logit probability, and it is counted as a correct prediction in the crosstab table. This all-or-nothing approach was replaced by a probabilistic approach. The latter is more realistic considering that if for an individual traveler a mode has a probability of 0.44 (highest among all other alternative modes) then assuming that out of 100 travelers 44 would choose that mode, it is more realistic than assuming the individual will choose that mode (when actually 0.44 individuals choose it). In order to calculate the number of times a mode was selected (as a value in the crosstab) the probabilities of that mode for all the individuals were added and the value was rounded to nearest integer.

TABLE 9 Models Estimated by GA Using Different Fitness Functions

| Variables | Fitness Function Used | | |
|---|------------------------------|----------------|---------------|
| | Predmax | Lmax | LLmax |
| A_A | 2.78 | 43.71 | 2.51 |
| A_B | 2.26 | 1.64 | 1.62 |
| A_C | -2.82 | -12.09 | -1.76 |
| A_D | -2.16 | 6.05 | 1.40 |
| A_E | -1.14 | -1.93 | -3.42 |
| A_F | -3.66 | -26.02 | -0.12 |
| A_G | 2.56 | 6.59 | 1.72 |
| A_H | -1.20 | -24.88 | -1.00 |
| TRTIME | -0.04 | -1.03 | -0.06 |
| TOLLINC | -24.07 | -105.11 | -11.30 |
| APEAK | 0.17 | -19.49 | -0.19 |
| AEDUCHS | -0.28 | -16.70 | -0.90 |
| BTRLENGTH | 0.01 | -9.50 | 0.00 |
| BEDUCHS | -4.71 | 1.65 | -1.68 |
| CAGE_Y | -3.07 | -6.41 | 2.74 |
| CEDUCSCV | 3.03 | 0.14 | 2.69 |
| CHHTPM | 0.55 | -5.04 | 1.14 |
| DTRLENGTH | -2.30 | -8.25 | 0.00 |
| DTPREC | 1.60 | -1.88 | -1.35 |
| ETRLLENGTH | -2.00 | -3.31 | 0.00 |
| EAGE_Y | -1.30 | -1.30 | 2.26 |
| EHHTPM | -0.45 | 1.43 | 1.35 |
| ENVEHS | 0.74 | 3.30 | 0.41 |
| EEDUCSCV | 2.93 | 0.68 | 1.01 |
| ETPCOM | -1.87 | -1.07 | 1.46 |
| ETPREC | -5.50 | 6.04 | 1.80 |
| FTPREC | -2.84 | -4.21 | -1.04 |
| GTPCOM | 0.37 | 21.84 | 0.86 |
| GHHTPM | 0.07 | 16.48 | 0.37 |
| GOFFPEAK | -0.47 | -11.88 | -0.33 |
| HTPCOM | 1.71 | -3.14 | 1.89 |
| | | | |
| VTTS (% of hourly wage rate in \$/hr) | 10 % | 58 % | 31 % |
| Percent correct | 63.2% | 60.5% | 48.4% |
| Number of correct expected signs | 13 | 15 | 14 |

TABLE 10 Crosstab Matrix for the Model Estimated by Using Fitness Function Predmax (Trips)

| Mode | A | B | C | D | E | F | G | H | I | Actual Total |
|-----------------|-----|-----|---|---|---|---|-----|---|---|--------------|
| A | 393 | 49 | 0 | 0 | 0 | 0 | 17 | 0 | 0 | 459 |
| B | 150 | 109 | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 277 |
| C | 16 | 8 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 31 |
| D | 22 | 7 | 0 | 0 | 0 | 0 | 161 | 0 | 0 | 190 |
| E | 9 | 3 | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 42 |
| F | 25 | 3 | 0 | 0 | 0 | 0 | 29 | 0 | 0 | 57 |
| G | 44 | 10 | 0 | 0 | 0 | 0 | 666 | 0 | 0 | 720 |
| H | 0 | 0 | 0 | 0 | 0 | 0 | 53 | 0 | 0 | 53 |
| I | 0 | 0 | 0 | 0 | 0 | 0 | 18 | 0 | 0 | 18 |
| Predicted Total | 659 | 189 | 0 | 0 | 0 | 0 | 999 | 0 | 0 | 1847 |

TABLE 11 Crosstab Matrix for the Model Estimated by Using Fitness Function Lmax (Trips)

| Mode | A | B | C | D | E | F | G | H | I | Actual Total |
|-----------------|-----|---|---|---|---|---|------|---|----|--------------|
| A | 438 | 0 | 0 | 0 | 0 | 0 | 21 | 0 | 0 | 459 |
| B | 238 | 0 | 0 | 0 | 0 | 0 | 39 | 0 | 0 | 277 |
| C | 24 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 31 |
| D | 29 | 0 | 0 | 0 | 0 | 0 | 153 | 0 | 8 | 190 |
| E | 12 | 0 | 0 | 0 | 0 | 0 | 29 | 0 | 1 | 42 |
| F | 27 | 0 | 0 | 0 | 0 | 0 | 30 | 0 | 0 | 57 |
| G | 45 | 0 | 0 | 0 | 0 | 0 | 674 | 0 | 1 | 720 |
| H | 0 | 0 | 0 | 0 | 0 | 0 | 53 | 0 | 0 | 53 |
| I | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 5 | 18 |
| Predicted Total | 814 | 0 | 0 | 0 | 0 | 0 | 1019 | 0 | 14 | 1847 |

TABLE 12 Crosstab Matrix for the Model Estimated by Using Fitness Function LLmax (Trips)

| Mode | A | B | C | D | E | F | G | H | I | Actual Total |
|-----------------|-----|-----|----|-----|----|----|-----|----|----|--------------|
| A | 246 | 130 | 6 | 17 | 8 | 8 | 44 | 0 | 0 | 459 |
| B | 123 | 102 | 3 | 6 | 5 | 6 | 30 | 0 | 0 | 277 |
| C | 11 | 7 | 4 | 2 | 1 | 0 | 5 | 0 | 0 | 31 |
| D | 13 | 8 | 3 | 52 | 3 | 5 | 92 | 7 | 6 | 190 |
| E | 6 | 4 | 0 | 5 | 7 | 0 | 18 | 0 | 1 | 42 |
| F | 16 | 9 | 1 | 6 | 0 | 7 | 17 | 1 | 0 | 57 |
| G | 32 | 19 | 8 | 105 | 17 | 27 | 464 | 32 | 15 | 720 |
| H | 0 | 0 | 1 | 9 | 2 | 1 | 32 | 8 | 0 | 53 |
| I | 0 | 0 | 0 | 3 | 0 | 0 | 12 | 0 | 3 | 18 |
| Predicted Total | 448 | 280 | 26 | 206 | 44 | 55 | 714 | 49 | 25 | 1847 |

```

crtab=zeros(10,10);
x=gareresults.x; %reads the values of the estimated coefficients if
                  %they have transferred to workspace from gatool%
Prm=zeros(ntravl,9); % Probability matrix-probability of each mode

%Following Commands go into main loop which calculates utilities
dr=zeros(1,9); %for PROBABILISTIC CROSSTAB

dr=[exp(q(1)),exp(q(2)),exp(q(3)),exp(q(4)),exp(q(5)),exp(q(6)),exp(q(
(7)),exp(q(8)),exp(q(9))];
    for h=1:9
        if (ut(i,h)~=NaN)
            Prm(i,h)=exp(ut(i,h))/nansum(dr);
        end;
    end;

%Following commands are outside the above mentioned loop.
%for creating crosstab% % PROBABILISTIC CROSSTAB%

i=1; j=1;
while (i< len)
    for n=1:9
        for h=1:9
            if (k1(i,13)==n)
                crtab(n,h)=crtab(n,h)+nansum(Prm(j,h));
            end;
        end;
        i=i+4;j=j+1;
    end;
    for z=1:9
        crtab(z,10)=sum(crtab(z,:));
        crtab(10,z)=sum(crtab(:,z));
    end;
    crtab(10,10)=sum(crtab(:,10));
    crtab;
    betas=x';

```

FIGURE 12 Commands for Generating the Cross-tab Matrix Using Estimated Coefficients

From the comparison between crosstabs generated by using LIMDEP and GA (LLmax), it is observed that these are quite similar and the percentage of correct predictions is comparable. The number of correct expected signs of the parameters is also comparable to that of the logit models estimated using LIMDEP.

However, the mode choice modeling using GA had some shortcomings. The genetic algorithm estimation took more time (40-50 minutes for a computer with processor speed 1.70 GHz) as

compared to estimation using LIMDEP. The coefficients estimated by using fitness functions Predmax and Lmax in GA lack consistency if the states of random numbers were changed during estimation and this leads to significant variation in the VTTS estimated from the systematic utility equations. The results of parameter estimation using GA varied significantly depending on the specification of the fitness function. As the Predmax fitness function works on all-or-nothing approach described above, it predicts no use for some of the modes. This shortcoming of the Predmax fitness function can be understood by an example. Assume that there are two modes A and B, and travelers chose mode A 70% of the time and mode B 30% of the time. When a fitness function such as Predmax is used, it estimates the parameters such that the utility of mode A is always greater than that of B. Hence even though the model never predicts mode B, it still predicts the correct mode 70% of the time (by predicting mode A every time). Similarly this justification can be extended for choice sets with more than two modes (multinomial model). In that case the modes which are selected more frequently have coefficients such that their utilities become higher as compared to the remaining modes and predict no use for these remaining modes. At the same time, the coefficients estimated by GA using fitness functions Predmax were significantly different from those estimated by LIMDEP, hence this approach of using fitness function based on all-or-nothing approach (Predmax) was discarded. Similarly the fitness function based on likelihood value (Lmax) also gave results which predicted no use for some of the modes, hence it was also discarded from further analysis and only the fitness function based on log-likelihood value (LLmax) was used.

3.6 Improving the Discrete Choice Model by Using a Genetic Algorithm

Using the fitness function based on log-likelihood value (LLmax), attempts were made to improve the base (MNL) model by adding and removing some variables. The final model selected after various trials is in Table 13. The options and methods in the genetic algorithm were slightly changed to stretch the optimization process and hence to refine the model further. The starting population was increased to 70, and the number of generations in stopping criteria was increased to 2000. This change increased the time of optimization to four hours (for a computer with processing speed 1.7GHz) but gave better results. This model was found to have lower log-likelihood value than the base model, and the crosstab matrix generated by using the estimated values is given in Table 14. It can be observed from the crosstab matrix that the model

has the same percentage of correct mode choice predictions. Overall the log-likelihood value and predicting ability was found to change only slightly during the model improvement trials. One of the models developed during this improvement process is shown in Table 15 along with the corresponding crosstab matrix given in Table 16.

TABLE 13 Improved MNL model estimated by LIMDEP and GA

| Description | Variables | LIMDEP estimation | GA estimation |
|---|------------------|--------------------------|----------------------|
| The alternative-specific constant | A_A | 3.02 | 2.49 |
| The alternative-specific constant | A_B | 1.72 | 1.65 |
| The alternative-specific constant | A_C | -0.37 | -0.99 |
| The alternative-specific constant | A_D | 0.94 | 0.94 |
| The alternative-specific constant | A_E | -4.44 | -2.88 |
| The alternative-specific constant | A_F | 1.83 | 1.36 |
| The alternative-specific constant | A_G | 2.49 | 1.87 |
| The alternative-specific constant | A_H | -0.31 | -0.85 |
| The travel time savings obtained by using the HOV lane (minutes); the value was 0 for modes A, C, and G because there were no travel time savings if the trip occurred on the GPLs. | TRTIME | -0.07 | -0.06 |
| Toll / (annual household income / 2000) | TOLLINC | -10.7 | -10.5 |
| The dummy variable used to describe if the traveler was driving during peak hours, yes = 1, no = 0 | APEAK | -0.31 | -0.21 |
| The dummy variable used to describe if the traveler's education level was high school graduate, yes = 1, no = 0 | AEDUCHS | -1.00 | -0.93 |
| The total travel time of the trip (minutes) | BTRLENGTH | 0.01 | 0.00 |
| The dummy variable used to describe if the traveler's education level was high school graduate, yes = 1, no = 0 | BEDUCHS | -1.65 | -1.54 |
| The dummy variable used to describe if the traveler's age was from 16 to 24 years old, yes = 1, no = 0 | CAGE_Y | 2.49 | 2.61 |
| The dummy variable used to describe if the traveler's education level was some college / vocational, yes = 1, no = 0 | CEDUCSCV | 2.30 | 2.28 |

TABLE 13 Improved MNL model estimated by LIMDEP and GA (continued)

| Description | Variables | LIMDEP estimation | GA estimation |
|--|------------------|--------------------------|----------------------|
| The total travel time of the trip (minutes) | DTRLENGTH | 0.01 | 0.01 |
| The dummy variable used to describe if the traveler's trip purpose was recreational, yes =1, no = 0 | DTPREC | -1.37 | -1.38 |
| The dummy variable used to describe if the traveler's household type was married without children, yes = 1, no = 0 | DHTPM | -0.42 | -0.40 |
| The total travel time of the trip (minutes) | ETRLLENGTH | 0.02 | 0.01 |
| The dummy variable used to describe if the traveler's age was from 16 to 24 years old, yes = 1, no = 0 | EAGE_Y | 2.74 | 2.62 |
| The number of motor vehicles (including cars, vans, trucks, and motorcycles) available in the traveler's household | ENVEHS | 0.35 | 0.28 |
| The dummy variable used to describe if the traveler's trip purpose was commuting, yes =1, no = 0 | ETPCOM | 2.79 | 1.71 |
| The dummy variable used to describe if the traveler's trip purpose was recreational, yes = 1, no = 0 | ETPREC | 3.00 | 1.82 |
| The number of motor vehicles (including cars, vans, trucks, and motorcycles) available in the traveler's household | FNVEHS | -0.71 | -0.67 |
| The dummy variable used to describe if the traveler's trip purpose was recreational, yes = 1, no = 0 | FTPREC | -1.47 | -1.41 |
| The dummy variable used to describe if the traveler was driving during off-peak hours, yes = 1, no = 0 | GOFFPEAK | -0.21 | -0.25 |
| The dummy variable used to describe if the traveler's trip purpose was commuting, yes =1, no = 0 | GTPCOM | 0.77 | 0.84 |
| The dummy variable used to describe if the traveler's age was from 16 to 24 years old, yes = 1, no = 0 | GAGE_Y | 1.47 | 1.58 |
| The dummy variable used to describe if the traveler's trip purpose was commuting, yes =1, no = 0 | HTPCOM | 1.47 | 1.46 |
| VTTS (% of hourly wage rate in \$/hr) | | 40.3 | 35.9 |
| Percent correct | | 49.16 | 48.4 |
| Number of correct expected signs | | 22 | 22 |
| Log-likelihood value | | -1685.27 | -1692.68 |

TABLE 14 Crosstab Matrix for the Improved GA Model (Trips)

| Mode | A | B | C | D | E | F | G | H | I | Actual Total |
|-----------------|-----|-----|----|-----|----|----|-----|----|----|--------------|
| A | 247 | 131 | 5 | 15 | 8 | 8 | 43 | 0 | 0 | 459 |
| B | 124 | 103 | 4 | 5 | 5 | 6 | 30 | 0 | 0 | 277 |
| C | 11 | 7 | 4 | 2 | 1 | 0 | 5 | 0 | 0 | 31 |
| D | 13 | 9 | 4 | 49 | 5 | 5 | 92 | 7 | 6 | 190 |
| E | 6 | 5 | 0 | 5 | 6 | 0 | 19 | 0 | 1 | 42 |
| F | 15 | 9 | 1 | 6 | 1 | 8 | 16 | 1 | 0 | 57 |
| G | 32 | 20 | 10 | 99 | 18 | 27 | 465 | 33 | 16 | 720 |
| H | 0 | 0 | 1 | 8 | 1 | 1 | 32 | 9 | 0 | 53 |
| I | 0 | 0 | 1 | 3 | 0 | 0 | 12 | 0 | 3 | 18 |
| Predicted Total | 450 | 283 | 30 | 193 | 45 | 56 | 714 | 50 | 26 | 1847 |

TABLE 15 MNL Model Estimated by LIMDEP and GA during Improvement Process

| Description | Variables | LIMDEP estimation | GA estimation |
|---|-----------|-------------------|---------------|
| The alternative-specific constant | A_A | 3.09 | 2.58 |
| The alternative-specific constant | A_B | 1.79 | 1.76 |
| The alternative-specific constant | A_C | -0.33 | -0.88 |
| The alternative-specific constant | A_D | 0.97 | 0.61 |
| The alternative-specific constant | A_E | -4.44 | -2.58 |
| The alternative-specific constant | A_F | 1.94 | 1.06 |
| The alternative-specific constant | A_G | 2.63 | 1.98 |
| The alternative-specific constant | A_H | -0.18 | -0.94 |
| The travel time savings obtained by using the HOV lane (minutes); the value was 0 for mode A, C, and G, because there were no travel time savings if the trip occurred on the GPLs. | TRTIME | -0.07 | -0.07 |
| Toll / (annual household income / 2000) | TOLLINC | -1.09 | -1.07 |
| The dummy variable used to describe if the traveler was driving during peak hours, yes = 1, no = 0 | APEAK | -0.32 | -0.25 |
| The dummy variable used to describe if the traveler's education level was high school graduate, yes = 1, no = 0 | AEDUCHS | -1.12 | -0.99 |
| The total travel time of the trip (minutes) | BTRLENGTH | 0.01 | 0.00 |

**TABLE 15 MNL Model Estimated by LIMDEP and GA during Improvement
Process (continued)**

| | | | |
|--|------------|----------|----------|
| The dummy variable used to describe if the traveler's education level was high school graduate, yes = 1, no = 0 | BEDUCHS | -1.75 | -1.66 |
| The dummy variable used to describe if the traveler's age was from 16 to 24 years old, yes = 1, no = 0 | CAGE_Y | 2.04 | 2.24 |
| The dummy variable used to describe if the traveler's education level was some college / vocational, yes = 1, no = 0 | CEDUCSCV | 2.40 | 2.41 |
| The total travel time of the trip (minutes) | DTRLENGTH | 0.01 | 0.01 |
| The dummy variable used to describe if the traveler's trip purpose was recreational, yes =1, no = 0 | DTPREC | -1.25 | -1.36 |
| The total travel time of the trip (minutes) | ETRLLENGTH | 0.02 | 0.00 |
| The dummy variable used to describe if the traveler's age was from 16 to 24 years old, yes = 1, no = 0 | EAGE_Y | 2.12 | 2.25 |
| The number of motor vehicles (including cars, vans, trucks, and motorcycles) available in the traveler's household | ENVEHS | 0.38 | 0.23 |
| The dummy variable used to describe if the traveler's trip purpose was commuting, yes =1, no = 0 | ETPCOM | 2.75 | 1.86 |
| The dummy variable used to describe if the traveler's trip purpose was recreational, yes = 1, no = 0 | ETPREC | 3.07 | 2.24 |
| The number of motor vehicles (including cars, vans, trucks, and motorcycles) available in the traveler's household | FNVEHS | -0.73 | -0.56 |
| The dummy variable used to describe if the traveler's trip purpose was recreational, yes = 1, no = 0 | FTPREC | -1.44 | -1.39 |
| The dummy variable used to describe if the traveler was driving during off-peak hours, yes = 1, no = 0 | GOFFPEAK | -0.23 | -0.30 |
| The dummy variable used to describe if the traveler's trip purpose was commuting, yes =1, no = 0 | GTPCOM | 0.73 | 0.84 |
| The dummy variable used to describe if the traveler's age was from 16 to 24 years old, yes = 1, no = 0 | GAGE_Y | -1.28 | -1.07 |
| The dummy variable used to describe if the traveler's trip purpose was commuting, yes =1, no = 0 | HTPCOM | 1.47 | 1.70 |
| VTTS (% of hourly wage rate in \$/hr) | | 40 | 38 |
| Percent correct | | 49.3 | 48.8 |
| Number of correct expected signs | | 16 | 16 |
| Log-likelihood value | | -1682.95 | -1692.44 |

TABLE 16 Crosstab Matrix for the GA Model Estimated During Improvement Process (Trips)

| Mode | A | B | C | D | E | F | G | H | I | Actual Total |
|-----------------|-----|-----|----|-----|----|----|-----|----|----|--------------|
| A | 249 | 132 | 6 | 15 | 8 | 8 | 42 | 0 | 0 | 459 |
| B | 125 | 104 | 4 | 5 | 5 | 5 | 29 | 0 | 0 | 277 |
| C | 12 | 8 | 4 | 2 | 1 | 0 | 5 | 0 | 0 | 31 |
| D | 13 | 8 | 4 | 49 | 5 | 5 | 93 | 7 | 6 | 190 |
| E | 6 | 5 | 0 | 6 | 6 | 0 | 17 | 0 | 1 | 42 |
| F | 15 | 9 | 1 | 5 | 1 | 8 | 18 | 1 | 0 | 57 |
| G | 32 | 20 | 10 | 94 | 20 | 26 | 471 | 31 | 16 | 720 |
| H | 0 | 0 | 2 | 8 | 2 | 1 | 32 | 8 | 0 | 53 |
| I | 0 | 0 | 1 | 2 | 0 | 0 | 12 | 0 | 3 | 18 |
| Predicted Total | 452 | 285 | 31 | 186 | 47 | 52 | 719 | 48 | 26 | 1847 |

From the tables it can be observed that the GA-estimated models have a slightly lower log-likelihood value and percent correct prediction than the models estimated using LIMDEP. This may happen in the case where the genetic algorithm has not converged to the final solution. Since the log-likelihood function for the MNL is concave, it has unique solution. GA may reach that solution if we increase the population size and extend the stopping criteria, although it will take more computational time. Hence, LIMDEP was used for MNL model estimation for further study.

4.0 VTTS AND VPCS FOR DIFFERENT GROUPS OF TRAVELERS

As described in section 3.1 the VTTS or value of time can be estimated from the utility equations. VTTS for the models estimated in this study can be defined by Equation 18. The variable TOLLINC was used in place of the variable toll so as to take into account the fact that the VTTS is often proportional to a traveler's income (8,12). The survey did not collect data on the traveler's hourly wage rate so a surrogate measure, annual household income, was used. It was assumed that there are approximately 2000 work hours in year, hence in order to calculate the VTTS in terms of hourly wage rate Equation 19 was used. For a family with a single earning member this equation will give the desired VTTS, but for households with more than one income earning member, this equation will overestimate the VTTS.

$$VTTS = \frac{\partial U / \partial trtime}{\partial U / \partial tollinc} \dots\dots\dots(18)$$

$$= \frac{\beta_{trtime}}{\beta_{tollinc}} \times \frac{\text{Annual Household Income}}{2000} \times \frac{60 \text{ minutes}}{1 \text{ hour}} \dots\dots\dots(19)$$

where,

$$tollinc = \frac{\text{Toll}}{\frac{\text{Annual Household Income}}{2000}}$$

VTTS = the value of travel time savings, as % of hourly wage rate in \$/hour;

U = the utility function;

β_{trtime} = coefficient of the variable trtime; and

$\beta_{tollinc}$ = coefficient of the variable tollinc.

Similarly the VPCS can also be estimated from the utility equations as shown below:

$$\begin{aligned} VPCS_{\text{Peak to Off-Peak}} &= \frac{\partial U_{\text{SOV-GPL-OP}} / \partial apeak}{\partial U_{\text{SOV-GPL-OP}} / \partial tollinc} \\ &= \frac{\beta_{apeak}}{\beta_{tollinc}} \times \frac{\text{Annual Household Income}}{2000} \end{aligned}$$

Where: $VPCS_{\text{Peak to Off-Peak}}$ = the value of penalty for changing travel schedule from the preferred peak period to the off-peak period (dollars);

$U_{\text{SOV-GPL-OP}}$ = the utility function of mode option A (SOV on GPL off-peak);

$U_{\text{SOV-GPL-P}}$ = the utility function of mode option G (SOV on GPL peak);

β_{apeak} = coefficient of the variable *apeak*;

β_{goffpk} = coefficient of the variable *goffpk*; and

β_{tollinc} = coefficient of the variable *tollinc*.

4.1 Using GA for Groups of Travelers

4.1.1 Rejecting the Sample Not Required

For developing a model corresponding to a group (e.g. male travelers) we need to reject the sample which does not belong to that group (e.g. we need to reject sample corresponding gender=female). In order to carry out such operations in Matlab the commands similar to those given in Figure 13 were used in the command window of the Matlab. Alternatively the unnecessary sample can be rejected in LIMDEP by using the reject command (such as REJECT:GENDER=1\$), and the data can be exported. This exported data can be then read into Matlab as usual.

```
z=1;
lnt= length(kf); % Kf is data file k renamed to store data for
                  female%
while (z<= lnt)
    if (kf(z,21)~=0) % if sex not equal to 0 %
        kf(z,:)=[]; %Deletes the zth row %
        z=z-1;      % Keeps the counter at same position if a row is
                    deleted%
        lnt=length(kf); % new length after deletion of a row%
    end;
    z=z+1;
end;
```

FIGURE 13 Typical Commands for Rejecting Unnecessary Sample

4.1.2 Model Specification

In order to specify a model for any group of travelers using a GA one needs to carry out numerous trials to determine which variables should be included in the model. This is a time consuming process as one needs to estimate the model with and without the variable of interest, find the log-likelihood values, and perform the Likelihood Ratio (LR) test to determine if

addition/removal of the variable improves the model. Details of the LR test are given in the section 3.3 of this report.

The following points need to be considered when adding or removing a variable in the Genetic Algorithm M-file. [The M-file in Matlab is a sequence of user written commands, which return value(s) calculated using those commands whenever the name of the M-file is entered in command window. While using gatool we need to enter the name of the M-file in the interface (e.g. @utl_max). This M-file is used to calculate the value of the fitness function for every set of solutions which is returned to gatool for further optimization.]

- If the number of variables specified (on GA Tool interface) is more than that used (in the GA M-file), then it does not make any difference as far as the optimization is concerned. For example if one specifies the number of variables = 4 in the GA interface but uses only two $x(1)$ and $x(3)$ in the M-file, then after running the optimization with GA one gets required values for $x(1)$ and $x(3)$ along with some random values for $x(2)$ and $x(4)$.
- If the specified number of variables on the GA Tool interface is smaller than that used in the GA M-file, then the GA can not run the optimization and returns the corresponding error message.

Hence it should be noted that specifying more variables than required does little harm as long as we keep track of the used variables and their coefficients. The only drawback is that a very large number of variables slows down the speed of optimization.

4.2 Models for Different Groups of Travelers and Corresponding VTTS and VPCS

The travelers were divided into different groups based on gender, trip purpose, income, and number of vehicles owned to account for the fact that the population may not be homogeneous and to assume that every traveler will have same choice functions. Models were estimated for each of these groups, and the market segmentation test was carried out to verify that the groups (e.g. males and females) have different choice functions. One of these tests for groups based on gender is shown in Table 17. The estimated model coefficients corresponding to VTTS and VPCS are given in Table 18 through Table 21.

TABLE 17 Market Segmentation Test for Models of Groups Based on Gender

| | |
|---------------------|--|
| Degrees of Freedom | 29 |
| LR (pooled) | -1682.95 |
| LU | -1588.62 |
| C=-2(lr-lu) | 188.66 |
| Critical Chi square | 42.6 |
| Conclusion | Reject null that all segments have the same choice functions |

TABLE 18 Models for Groups based on Gender

| Male | | | Female | | |
|---|--------|----------|-----------|--------|---------|
| Variables | Coeff. | t-ratio | Variables | Coeff. | t-ratio |
| A_A | 2.63 | 4.84 | A_A | 2.57 | 6.53 |
| TIME | -0.08 | -8.87 | TIME | -0.07 | -6.28 |
| TOLLINC | -12.74 | -7.15 | TOLLINC | -8.98 | -4.78 |
| APEAK | -0.28 | -1.35 | APEAK | -0.42 | -1.50 |
| ANVEHS | 0.35 | 2.20 | AAGESNR | 0.95 | 2.27 |
| AOCPSAL | -1.03 | -2.32 | A_B | 0.31 | 0.57 |
| A_B | 1.31 | 2.29 | BTRLTH | 0.02 | 3.71 |
| BNVEHS | 0.49 | 2.93 | BHTPUA | -2.09 | -3.26 |
| BPAYPAR | 0.42 | 1.91 | A_C | -0.49 | -0.86 |
| A_C | -0.07 | -0.07 | CAGE_Y | 2.26 | 2.86 |
| CAGE_Y | 4.32 | 3.46 | CEDUCV | 1.79 | 2.96 |
| CEDUCCG | -1.43 | -2.16 | A_D | -0.54 | -0.95 |
| BOCPPRO | 1.83 | 1.74 | DTRLTH | 0.02 | 3.04 |
| A_D | 1.55 | 3.15 | DHTPSA | 1.14 | 3.04 |
| DTRLTH | 0.01 | 2.87 | A_E | -3.80 | -3.69 |
| DTPREC | -2.44 | -3.16 | ETRLTH | 0.03 | 2.97 |
| A_E | -1.38 | -1.48 | EAGE_Y | 2.44 | 3.58 |
| EAGE_Y | 3.34 | 5.14 | ENVEHS | 0.60 | 2.81 |
| ETPCOM | 2.07 | 2.44 | ETPREC | 2.02 | 2.74 |
| A_F | 1.67 | 2.43 | A_F | 1.05 | 1.22 |
| FNVEHS | -0.44 | -1.80 | FNVEHS | -1.84 | -2.95 |
| FTPREC | -2.21 | -2.07 | FHTPMC | 3.23 | 3.12 |
| A_G | 2.76 | 6.04 | A_G | 2.61 | 6.66 |
| GOFFPEAK | -0.15 | 0.86 | GOFFPEAK | -0.28 | 1.17 |
| GHTPM | 0.53 | 2.74 | GOCPSAL | -1.64 | -2.18 |
| GTPCOM | 0.95 | 4.17 | A_H | -3.15 | -1.90 |
| A_H | 1.07 | 2.33 | HTPCOM | 2.53 | 2.16 |
| HAGE_Y | 2.88 | 2.58 | HNVEHS | 0.83 | 2.09 |
| HOCPSAL | 4.13 | 3.51 | HTYPEMC | -1.10 | -1.97 |
| Number of Observations | | 1172 | 657 | | |
| Log-likelihood | | -1034.01 | -554.61 | | |
| VOT (% of hourly wage rate) | | 37% | 50% | | |
| VPCS (Peak to Off-peak) =apeak/tollinc (\$) | | 2% | 5% | | |
| VPCS (Off-peak to Peak) =goffpeak/tollinc (\$) | | 1% | 3% | | |

TABLE 19 Models for Groups based on Trip Purpose

| Trip Purpose-Commute | | | Trip Purpose other than Commute | | |
|---|--------|----------|---------------------------------|--------|---------|
| Variables | Coeff. | t-ratio | Variables | Coeff. | t-ratio |
| A_A | 3.72 | 10.09 | A_A | 3.36 | 5.50 |
| TIME | -0.08 | -10.33 | TIME | -0.06 | -3.75 |
| TOLLINC | -12.30 | -8.33 | TOLLINC | -6.76 | -3.90 |
| APEAK | -0.42 | -2.27 | APEAK | -0.26 | -0.71 |
| AHTPMC | -0.32 | -1.88 | AHTPMC | 0.55 | 1.72 |
| OCPPRO | -0.33 | -1.67 | ASEX | -0.98 | -2.87 |
| A_B | 1.95 | 4.71 | A_B | 2.20 | 3.67 |
| BNVEHS | 0.14 | 1.48 | BEDUCPG | -0.89 | -2.44 |
| A_C | 2.51 | 5.25 | A_C | -2.30 | -1.01 |
| CAGEA | -1.60 | -3.77 | CNVEHS | 0.60 | 0.86 |
| CEDUCG | -0.68 | -1.53 | A_D | 2.96 | 3.57 |
| A_D | 0.98 | 2.39 | DAGES | -1.21 | -1.74 |
| DTRLTH | 0.02 | 3.75 | DOCPPRO | -0.95 | -2.14 |
| A_E | 0.26 | 0.44 | DHSIZE | -0.53 | -2.67 |
| EHTPMC | -0.80 | -2.14 | A_E | -4.11 | -2.80 |
| EEDUCG | -1.04 | -2.72 | EHTPM | 3.07 | 2.94 |
| ETRLTH | 0.02 | 2.98 | EEDUCG | 4.31 | 3.52 |
| A_F | 1.23 | 1.46 | A_F | 2.38 | 2.69 |
| FHTPMC | 2.53 | 4.08 | FHTPMC | 2.29 | 2.78 |
| FOCPPRO | 0.84 | 1.87 | FOCPPRO | -3.30 | -4.44 |
| FNVEHS | -1.66 | -3.83 | FNVEHS | -0.93 | -2.72 |
| A_G | 3.43 | 9.80 | A_G | 3.72 | 5.07 |
| GOFFPEAK | 0.00 | -0.03 | GOFFPEAK | -1.33 | 3.64 |
| HGENDER | 0.28 | 1.78 | GSEX | -0.88 | -2.43 |
| A_H | -0.01 | -0.02 | GTNTALL | -0.09 | -3.00 |
| HTNTALL | 0.14 | 2.63 | GAGEA | -1.22 | -2.70 |
| | | | A_H | -2.00 | -1.74 |
| | | | HEDUCG | 2.07 | 1.74 |
| Number of Observations | | 1465 | 364 | | |
| Log-likelihood | | -1280.30 | -302.51 | | |
| VOT (% of hourly wage rate) | | 39% | 56% | | |
| VPCS (Peak to Off-peak) =apeak/tollinc (\$) | | 3% | 4% | | |
| VPCS (Off-peak to Peak) =goffpeak/tollinc (\$) | | 0% | 20% | | |

TABLE 20 Models for Groups based on Number of Vehicles

| Number of Vehicles 0-1 | | | Number of vehicles 2+ | | |
|---|--------|---------|-----------------------|--------|---------|
| Variables | Coeff. | t-ratio | Variables | Coeff. | t-ratio |
| A_A | 5.84 | 4.90 | A_A | 3.87 | 10.35 |
| TIME | -0.12 | -4.58 | TIME | -0.07 | -10.08 |
| TOLLINC | -7.81 | -3.07 | TOLLINC | -11.47 | -7.90 |
| APEAK | 0.68 | 0.91 | APEAK | -0.42 | -2.41 |
| GTNTALL | -0.23 | -2.41 | ATRLGTH | -0.01 | -3.25 |
| AGENDER | -2.57 | -3.56 | A_B | 2.43 | 6.72 |
| A_B | 3.44 | 2.60 | BEDUCPG | 0.47 | 2.51 |
| BTRLGTH | -0.03 | -2.19 | BOCPPRO | -0.29 | -1.56 |
| BAGEA | -1.59 | -1.97 | A_C | 1.08 | 2.15 |
| BHTPMC | 2.25 | 2.31 | CEDUCSV | 1.74 | 4.10 |
| A_C | -3.47 | -1.23 | CAGEA | -1.05 | -2.45 |
| CTSQR | 0.00 | 1.50 | A_D | 1.05 | 2.28 |
| A_D | -2.35 | -2.08 | DAGEA | 0.56 | 2.11 |
| DTNTALL | 0.29 | 3.43 | DSEX | 0.41 | 1.76 |
| A_E | 0.38 | 0.30 | DHTPS | 0.69 | 1.97 |
| EAGEA | -2.73 | -1.77 | A_E | -0.34 | -0.60 |
| A_F | -2.27 | -2.27 | EHTPM | 0.94 | 2.79 |
| FGENDER | 1.59 | 1.45 | ETRLGTH | 0.01 | 1.76 |
| FHTPMC | 2.83 | 2.00 | A_F | 0.23 | 0.38 |
| A_G | 2.21 | 3.06 | FHTPMC | 1.36 | 3.18 |
| GOFFPEAK | -0.68 | 1.44 | FTNTALL | -0.10 | -2.17 |
| GTPCOM | 0.97 | 1.78 | A_G | 3.14 | 8.11 |
| GEDUCSV | -1.34 | -2.37 | GOFFPEAK | -0.23 | 1.59 |
| A_H | -7.58 | -1.28 | GSEX | 0.33 | 1.93 |
| HTRLNGTH | 0.08 | 1.19 | GTNTALL | -0.03 | -1.88 |
| | | | GOCPPRO | 0.41 | 2.67 |
| | | | A_H | 1.85 | 3.52 |
| | | | HTRLNGTH | -0.01 | -1.33 |
| Number of Observations | | 227 | 1602 | | |
| Log-likelihood | | -150.46 | -1456.65 | | |
| VOT (% of hourly wage rate) | | 89% | 39% | | |
| VPCS (Peak to Off-peak) =apeak/tollinc (\$) | | -9% | 4% | | |
| VPCS (Off-peak to Peak) =goffpeak/tollinc (\$) | | 9% | 2% | | |

TABLE 21 Models for Groups based on Income¹

| Income less than 100K | | | Income more than 100K | | |
|---|--------|---------|-----------------------|--------|---------|
| Variables | Coeff. | t-ratio | Variables | Coeff. | t-ratio |
| A_A | 3.02 | 8.69 | A_A | 3.80 | 5.48 |
| TIME | -0.07 | -6.99 | TIME | -0.08 | -7.85 |
| TOLL | -0.69 | -8.59 | TOLL | -0.40 | -6.37 |
| APEAK | -0.69 | -3.01 | APEAK | -0.16 | -0.65 |
| ATPREC | 1.61 | 3.46 | ATPCOMM | -0.68 | -2.34 |
| AGENDER | -0.41 | -2.03 | ATNTALL | 0.05 | 1.92 |
| A_B | 2.54 | 5.24 | A_B | 2.87 | 4.34 |
| BTPCOM | -0.69 | -2.50 | BTPREC | 0.96 | 2.30 |
| BTRLTH | 0.01 | 1.74 | BSEX | 0.52 | 2.02 |
| A_C | -0.07 | -0.14 | A_C | 2.40 | 2.98 |
| CHTPEM | 1.54 | 3.24 | CAGEACT | -3.22 | -2.83 |
| CTPWORK | 1.79 | 2.14 | A_D | 3.45 | 4.30 |
| A_D | 2.51 | 4.04 | DTRLTH | 0.01 | 1.71 |
| DTRLTH | 0.01 | 2.08 | DTNTALL | 0.02 | 2.13 |
| A_E | -0.87 | -1.32 | DTPCOMM | -0.55 | -1.75 |
| EEDCSCV | 0.83 | 1.97 | A_E | 0.94 | 1.29 |
| ENVEHS | 0.72 | 3.71 | EHTPEM | 1.11 | 1.90 |
| A_F | 0.84 | 1.33 | A_F | 1.99 | 1.93 |
| FHTPEMC | 2.03 | 3.82 | FHTPEMC | 1.26 | 2.00 |
| FNVEHS | -1.01 | -3.11 | FNVEHS | -0.89 | -2.37 |
| A_G | 2.49 | 7.69 | A_G | 4.19 | 6.01 |
| GOFFPEAK | -0.41 | 2.16 | GOFFPEAK | 0.04 | -0.19 |
| GAGESNR | 1.18 | 3.94 | GAGEACT | -0.51 | -1.78 |
| A_H | 1.17 | 2.66 | GOCCPPR | 0.52 | 2.03 |
| HHTPEMC | 0.76 | 1.72 | A_H | 0.41 | 0.39 |
| | | | HHTPEMC | -1.63 | -2.99 |
| | | | HHSIZE | 0.57 | 2.41 |
| | | | HEDCCG | 1.27 | 2.41 |
| Number of Observations | | 1015 | 814 | | |
| Log-likelihood | | -855.75 | -712.85 | | |
| VOT (\$/Hr) | | 5.84 | 12.29 | | |
| VPCS (Peak to Off-peak) =apeak/tollinc (\$) | | 1.00 | 0.39 | | |
| VPCS (Off-peak to Peak) =goffpeak/tollinc (\$) | | 0.59 | -0.10 | | |

Although the VTTS values were obtained from coefficients which were statistically significant, the VPCS values obtained were not significant (as most of the times the coefficients apeak and goffpeak were not significant). The estimated values of travel time savings are summarized in Table 22. It can be observed that for the whole sample the estimated VTTS is within the range as obtained in previous research (11). Gender is one of the important factors in travelers'

¹ Note that in models for groups based on income the variable TOLLINC was replaced by the variable TOLL

willingness to pay, and it was confirmed by different values of times obtained for males and females with both values in the expected range. Similarly, the groups based on trip purpose and number of vehicles had different values of time justifying the segmentation in respective groups. The VTTS obtained for groups based on income are comparable to those obtained in previous research (12,13). This is interesting since the survey used to collect the VTTS offered respondents a complex set of mode choices which was much broader than other surveys. The VPCS values for most of the groups were not significant just as VPCS of the total sample. The insignificant values imply that the choice of modes A and G are not affected largely by time of departure for VPCS which may be due to travelers' reluctance to change the travel schedule.

TABLE 22 Estimated VTTS for Katy Freeway Travelers

| Criterion for Segmentation | Group of Travelers | VTTS |
|-----------------------------------|---------------------------|-------------------------|
| None | Whole Sample | 39% of hourly wage rate |
| Gender | Male | 37% of hourly wage rate |
| | Female | 50% of hourly wage rate |
| Trip Purpose | Commute | 39% of hourly wage rate |
| | Non-commute | 56% of hourly wage rate |
| Number of Vehicles per Household | 0-1 Vehicles | 89% of hourly wage rate |
| | 2 or more Vehicles | 39% of hourly wage rate |
| Annual Household Income | ≤ \$100,000 | 5.84 (\$/hr) |
| | >\$100,000 | 12.29 (\$/hr) |

5.0 CONCLUSIONS

The study started with a simple multinomial logit model for the general purpose lane travelers on Katy Freeway. The stated preference surveys used for data collection in this study involved a large number of modes to choose from. The traveler's choice set contained nine travel alternatives, with some of these alternatives being somewhat similar. Hence, attempts were made to check if more complex modeling methods would work better than a multinomial logit model.

To begin, the study attempted to estimate a random parameter logit model for the given data. The estimation results indicated that nothing was gained by forcing parameters to vary across individuals. The standard deviations of the random parameters were not significant to justify the assumption that these parameters vary across individuals. This may be partly due to the nature of the data which contain four stated preference questions per individual; although for modeling they are counted as four separate individuals. Hence the random parameter logit model was not preferred over the MNL model. This study successfully estimated a nested logit model for modeling mode choice of Katy Freeway travelers on general purpose lanes. The nested logit models were found to have nearly similar log-likelihood values as compared to simple MNL model. Also, the percentage of correct mode choice predictions and value of travel time savings obtained by both MNL and NL were similar, hence the simpler MNL model was used for further analysis.

The study then estimated MNL models using genetic algorithm optimization for solving the likelihood function. The results from the genetic algorithm were comparable to MNL model estimation by using LIMDEP. However, with the given optimization settings, the genetic algorithm could not achieve the optima and hence the log-likelihood values of models obtained by using genetic algorithm optimization were slightly lower than those obtained by LIMDEP. Another drawback of using a genetic algorithm for likelihood estimation of MNL model was large computational time. However, the study showed that the genetic algorithm can be successfully used for mode choice modeling. Further, genetic algorithms can prove very useful in case of model specification with complex likelihood function which is difficult to optimize using other techniques. It is also possible to avoid the need for making assumptions such as

independence of irrelevant alternatives by correctly specifying the fitness functions. One of the attempts for such a fitness function was made by maximizing the number of correct predictions.

To calculate the value of travel time savings and penalty for changing travel schedule for various groups of travelers, the study estimated MNL models using LIMDEP. In order to account for heterogeneity among the travelers the sample was divided into groups (segments) based on gender, number of vehicles, trip purpose and income. These groups were found to have different utility functions and hence different VTTS and VPCS values. The VPCS values calculated were not significant hence they were not used. The VTTS values different by group and were generally comparable to those estimated by previous research. For example, the average VTTS was found to be 39% of the wage rate, with high income respondents having an average VTTS of \$12.29 per hour and lower income travelers having a VTTS of \$5.84 per hour. Females were found to have higher VTTS than males, non-commuters had a higher VTTS than commuters, and travelers with few vehicles had higher VTTS than those with more than two vehicles.

REFERENCES

1. Cain, A., Burris, M.W. and Pendyala, R.M. "The Impact of Variable Pricing on the Temporal Distribution of Travel Demand," *Transportation Research Record 1747*, TRB, National Research Council, Washington, D.C. 2001, pp. 36–43.
2. Xu, Lei, *An Estimation of Potential Single-Occupancy Vehicle Demand for the Katy Freeway and Northwest Freeway High-Occupancy Vehicle Lanes*, MS Thesis, Texas A & M University, College Station, TX. 2005.
3. *LIMDEP User's Manual: Version 7.0*, Econometric Software Inc., Bellport. 1995.
4. Train, K. *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, MA. 2003.
5. McFadden, D. and Train, K. "Mixed MNL models of discrete response", *Journal of Applied Econometrics*, 15, 2000, pp. 447-470.
6. Hensher, D. and Greene, W. "The mixed logit model: The state of practice and warnings for the unwary," Working Paper, School of Business, The University of Sydney, 2001.
7. McFadden, D. "Modeling the choice of residential location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull, eds., *Spatial Interaction Theory and Planning Models*, North-Holland, Amsterdam, 1978, pp. 75–96.
8. Bhat, C. R.; Koppelman, F.; Sethi V.; and Williams, B. *A Self Instructing Manual on Discrete Choice Modeling* (Draft Version). Prepared for the Federal Highway Administration, 1995.
9. Roorda, M.; Miller, E.J.; and Kruchten, N. *Incorporating Within-Household Interactions into Mode Choice Model Using Genetic Algorithm for Parameter Estimation*. Presented to the Transportation Research Board Annual Meeting, National Research Council, Washington D.C. 2006.
10. *MATLAB User's Guide 7: Genetic Algorithm and Direct Search Toolbox*, The MathWorks Inc., 2004.
11. Small, K. *Urban Transportation Economics*. Harwood Academic Publishers, Philadelphia, PA. 1992.

12. Calfee, J. and Winston, C. The Value of Automobile Travel Time: Implications for Congestion Policy. *Journal of Public Economics*, No. 69. 1998.
13. Brownstone, D. and Small, K A. Valuing time and reliability: assessing the evidence from road pricing demonstrations. *Transportation Research Part A-Policy and Practice*. 39 (4), 2005, pp. 279-293.